SPRINGER REFERENCE

Tamer Başar
Georges Zaccour
Editors

Handbook of Dynamic Game Theory

Springer

Handbook of Dynamic Game Theory

Tamer Başar • Georges Zaccour
Editors

Handbook of Dynamic Game Theory

With 135 Figures and 22 Tables

Springer

Editors
Tamer Başar
Coordinated Science Laboratory
and Department of Electrical
and Computer Engineering
University of Illinois at Urbana-Champaign
Urbana, IL, USA

Georges Zaccour
Department of Decision Sciences
GERAD, HEC Montréal
Montreal, QC, Canada

ISBN 978-3-319-44373-7 ISBN 978-3-319-44374-4 (eBook)
ISBN 978-3-319-44685-1 (print and electronic bundle)
https://doi.org/10.1007/978-3-319-44374-4

Library of Congress Control Number: 2018937969

Printed on acid-free paper

This Springer imprint is published by the registered company Springer International Publishing AG part of Springer Nature.
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Preface

The *Handbook of Dynamic Game Theory* is a comprehensive two-part two-volume treatise on dynamic/differential games, which is a mature but still growing and expanding field, in both theoretical developments and the range of its applications. The first part (and volume) of the *Handbook* focuses on fundamentals and theory, and the second part (and volume) covers applications in diverse fields, such as economics, management science, engineering, and biology. Each part is broken into chapters dealing with specific topics or sub-areas, all written by experts on these topics. Each chapter itself provides a comprehensive coverage of the corresponding topics, written with a broader audience in mind, but without dilution of the technical content. We provide below brief descriptions of the contents of the two parts (and volumes), and hence of the *Handbook*.

1 Part I: Theory of Dynamic Games

The first part (and volume) includes fourteen chapters. To set the stage for dynamic games covered by other chapters in the Handbook, **Tamer Başar** provides in Chap. 1 a general *Introduction to the Theory of Games*. In particular, he describes what game theory is about and its historical origins, shows how to formulate a (static) game, introduces main solution concepts for both zero-sum and nonzero-sum games, and discusses some classical results. In Chap. 2, **Tamer Başar, Alain Haurie and Georges Zaccour** provide an overview of the theory of *Nonzero-Sum Differential Games*, describing the general framework for their formulation, the role played by information structures and their importance, and noncooperative solution concepts. The chapter illustrates some of the key concepts with simple examples and places particular emphasis on the tractable class of linear-quadratic differential games, which has often been adopted in applications.

In Chap. 3, **Dean Carlson, Alain Haurie and Georges Zaccour** expose a full theory for *Infinite-Horizon Concave Differential Games with Coupled Constraints*. Concave games provide an attractive setting for many applications of differential games in economics, management science, and engineering, and state coupling constraints happen to arise quite naturally in many of these applications. The theory is illustrated with the classical model of Ramsey. In Chap. 4, **Jacek B. Krawczyk and Vladimir Petkov** provide a comprehensive coverage of the theory of *Multistage Games*, that is, state-space dynamic games in discrete time. The objective and the content of the chapter are similar to those in Chap. 2, which is a continuous-time counterpart to this one. The role of information structure in dynamic games is highlighted and different equilibria are discussed. The authors also show how memory-based (non-Markovian) strategies can support Pareto-efficient outcomes in a dynamic game.

Chapters 5 and 6 deal with zero-sum stochastic games and nonzero-sum stochastic games, respectively. In Chap. 5, **Anna Jaśkiewicz and Andrzej S. Nowak** review all basic streams of research in *Zero-Sum Stochastic Games*, including the existence of value and uniform value, algorithms, vector payoffs, incomplete information, and imperfect state observation. Some models related to continuous-time games are briefly discussed. Chapter 6 describes a number of results obtained in the last 60 years on the theory of *Nonzero-Sum Stochastic Games*. In particular, **Anna Jaśkiewicz and Andrzej S. Nowak** provide an overview of most important results related to the existence of stationary Nash and correlated equilibria in models on countable and general state spaces, the existence of subgame-perfect equilibria, algorithms, stopping games, and the existence of uniform equilibria. The survey also incorporates several examples of games studied in operations research and economics.

In Chap. 7, **Peter E. Caines, Minyi Huang and Roland P. Malhamé** deliver a comprehensive current account of the theory of *Mean Field Games* (MFGs). As the theory and methodology of MFGs has rapidly developed since its (relatively recent) inception and is still advancing, the objective of this chapter is to present the fundamental conceptual framework of MFGs in the continuous time setting and the main techniques that are currently available. In a nutshell, MFG theory studies the existence of Nash equilibria, together with the individual strategies which generate them, in games involving a large number of asymptotically negligible agents modeled by controlled stochastic dynamical systems. Chapter 8 is devoted to two-player, *Zero-Sum Differential Games*, with a special emphasis on the existence of a value and its characterization in terms of a partial differential equation, the Hamilton-Jacobi-Isaacs equation. **Pierre Cardaliaguet and Catherine Rainer** discuss different classes of finite horizon, infinite horizon, and pursuit-evasion games. They also analyze differential games in which the players do not have a full information on the structure of the game or cannot completely observe the state.

In Chap. 9, **Pierre Bernhard** presents the theory of *Robust Control and Dynamic Games*. The chapter's main objective is to describe a series of problems of robust control that can be approached using game theoretical tools. Chapter 10 is about *Evolutionary Game Theory*. **Ross Cressman and Joe Apaloo** first summarize features of matrix games before showing how the theory changes when the two-player game has a continuum of traits or interactions become asymmetric. Its focus is on the connection between static game-theoretic solution concepts (e.g., evolutionarily stable, continuously stable strategies) and stable evolutionary outcomes for deterministic evolutionary game dynamics (e.g., the replicator equation, adaptive dynamics). The chapter provides a series of examples to illustrate some of the main results of this theory. In Chap. 11, **Jason R. Marden and Jeff S. Shamma** provide an overview of the *Game Theoretic Learning in Distributed Control*. In distributed architecture control problems, there is a collection of interconnected decision-making components that seek to realize desirable collective behaviors through local interactions and by processing local information. One approach to control such architectures is to view the components as players in a game. The chapter covers special game classes, measures of distributed efficiency, utility design, and online learning rules, all with the interpretation of using game theory as a prescriptive paradigm for distributed control design. In Chap. 12, **S. Rasoul Etesami and Tamer Başar** provide a general overview of the topic of *Network Games*, its application in a number of areas, and recent advances, by focusing on four major types of games, namely, congestion games, resource allocation games, diffusion games, and network formation games. Several algorithmic aspects and methodologies for analyzing such games are discussed, and connections between network games and other relevant topical areas are identified.

The last two chapters of the first part of the *Handbook* concern cooperative differential games. In many instances, players find it individually and collectively rational to sign a long-term cooperative agreement. A major concern in such a setting is how to ensure that each player will abide by her commitment as time goes by. The players will stick to the agreement if each one of them still finds it individually rational at any intermediate instant of time to continue to implement her cooperative control rather than switch to a noncooperative control. If this condition is satisfied for all players, then we say that the agreement is time consistent. In *Cooperative Differential Games with Transferable Payoffs* (Chap. 13), **Leon A. Petrosyan and Georges Zaccour** deal with the design of schemes that guarantee time consistency in deterministic differential games with transferable payoffs. In *Nontransferable Utility Cooperative Dynamic Games* (Chap. 14), **David W.K. Yeung and Leon A. Petrosyan** deal with the same issues, but assuming away the possibility of side payments between players.

2 Part II: Applications of Dynamic Games

The second part (and volume) of the *Handbook* also has fourteen chapters, with four of them being in economics, two in management science, and five in engineering. The remaining chapters cover the classical area of pursuit-evasion games, social networks, and evolutionary games in biology.

In *Resource Economics* (Chap. 15), **Ngo van Long** provides a selective review of dynamic game models of exploitation of both exhaustible and renewable natural resources. To avoid much overlap with earlier surveys, the chapter focuses on recent literature, while emphasizing economic intuition behind the models and the results. In Chap. 16, *Dynamic Games of International Pollution Control: A Selective Review*, **Aart de Zeeuw** focuses on dynamic games of climate change with one global stock of pollutants. The chapter has two parts. In the first part, the author derives

Nash equilibria under different information structures and compares their outcomes to those that the players would achieve if they were to cooperate. In the second part, the stability of international environmental agreements (or partial-cooperation Nash equilibria) is investigated from different angles and representative examples are discussed.

In *Dynamic Games in Macroeconomics* (Chap. 17), **Łukasz Balbus, Kevin Reffett and Łukasz Woźny** survey how the methods of dynamic and stochastic games have been applied in recent work in macroeconomics. Among other topics, the authors discuss the strategic dynamic programming method with states, which is useful for proving the existence of sequential or subgame perfect equilibrium of a dynamic game. The chapter presents some illustrative cases and concludes with alternative methods that are useful for some macroeconomic problems. In Chap. 18, **Luca Colombo and Paola Labrecciosa** provide an overview of applications of *Differential Games in Industrial Organization.* On the menu are classical contributions on adjustment costs, sticky prices, and R&D races, as well as some more recent ones dealing with imperfect competition in the exploitation of renewable productive assets and strategic investments under uncertainty.

The next two chapters are on applications of dynamic games in management science. In Chap. 19, **Michèle Breton** covers *Dynamic Games in Finance*. Finance is a discipline encompassing all the essential ingredients of dynamic games, involving investors, managers, and financial intermediaries as players who have competing interests and interact strategically over time. This chapter presents dynamic game models used in various applications in the broad area of finance, with the objective of illustrating the scope of possibilities in this field. In Chap. 20, **Steffen Jørgensen** provides a survey of dynamic games in *Marketing*. As a functional area within a firm, marketing includes all the activities that the firm has at its disposal to sell products or services to other firms (wholesalers, retailers) or directly to the final consumers. The objective of this chapter is to demonstrate that the theory of differential games has proved to be useful for the study of a variety of problems in marketing, recognizing that most marketing decision problems are dynamic and involve strategic considerations.

In Chap. 21, **Sadegh Bolouki, Angelia Nedić and Tamer Başar** present some applications of game theory in *Social Networks*. The authors first focus on the formation of opinions over time through strategic interactions. In particular, they first determine whether an agreement (consensus) among all individuals is reached, or a clustering of opinions occurs, or none of these happens. Next, they turn their attention to decision-making processes (such as elections) in social networks, where a collective decision (social choice) must be made by multiple individuals (voters) with different preferences over the alternatives (candidates). In Chap. 22, **Valerii Patsko, Sergey Kumkov and Varvara Turova** provide a comprehensive survey of *Pursuit-Evasion Games*, a class of games that was at the origin of the development of differential game theory. The authors focus on time-optimal problems close to Rufus Isaacs' 'homicidal chauffeur' game and to linear differential games of fixed terminal time, with Josef Shinar's space interception problem as the major example. In *Biology and Evolutionary Games* (Chap. 23), **Mark Broom and Vlastimil**

Křivan survey some evolutionary games used in biological sciences. These include the Hawk-Dove game, the Prisoner's Dilemma, Rock-Paper-Scissors, the war of attrition, the Habitat Selection game, predator prey games, and signalling games.

The next five chapters deal with applications of dynamics games in engineering. In Chap. 24, **Joseph Z. Ben-Asher and Jason L. Speyer** discuss *Games in Aerospace: Homing Missile Guidance*. The development of a homing missile guidance law against an intelligent adversary requires the solution of a differential game. First, the authors formulate this problem as a linear dynamic system with an indefinite quadratic performance criterion. Next, they formulate a deterministic game allowing saturation, which is shown to be superior to the LQ guidance law. Finally, they deal with the presence of uncertainties in the measurements and process noise. In *Stackelberg Routing on Parallel Transportation Networks* (Chap. 25), **Walid Krichene, Jack D. Reilly, Saurabh Amin, and Alexandre M. Bayen** present a game theoretic framework for studying Stackelberg routing games on parallel transportation networks. They introduce a new class of latency functions to model congestion, inspired from the fundamental diagram of traffic. For this new class, several results from the classical congestion games literature do not hold, and the authors provide a characterization of Nash equilibria and show, in particular, that there may exist multiple equilibria that have different total costs. A simple polynomial-time algorithm is provided for computing the best Nash equilibrium, i.e., the one which achieves minimal total cost.

In *Communication Networks: Pricing, Congestion Control, Routing and Scheduling* (Chap. 26), **Srinivas Shakkottai and R. Srikant** consider three fundamental problems in the general area of communication networks and their relationship to game theory, namely: (i) allocation of shared bandwidth resources, (ii) routing across shared links, and (iii) scheduling across shared spectrum. The authors present results on each problem and characterize the efficiency loss that results from requesting information from the competing agents to construct a mechanism to allocate resources, instead of finding a globally optimal solution, which is impractical when the number of agents is very large. In *Power System Analysis: Competitive Markets, Demand Management, and Security* (Chap. 27), **Anibal Sanjab and Walid Saad** provide an overview of the application of game theory to various aspects of the power system, including strategic bidding in wholesale electric energy markets, demand side management mechanisms with special focus on demand response and energy management of electric vehicles, energy exchange and coalition formation between microgrids, as well as security of the power system viewed as a cyber-physical system, presenting a general security framework along with applications to the security of state estimation and automatic generation control. The final chapter in this second part of the Handbook, Chap. 28, is by **Debarun Kar, Thanh H. Nguyen, Fei Fang, Matthew Brown, Arunesh Sinha, Milind Tambe, and Albert Xin Jiang.** *Trends and Applications in Stackelberg Security Games* provides an overview of use-inspired research in security games, including algorithms for scaling up security games to real-world sized problems, handling multiple types of uncertainty and dealing with bounded rationality and bounded surveillance of human adversaries.

Each chapter was an invited contribution to the *Handbook* and was evaluated by at least two reviewers. We thank the authors for their contributions and the reviewers for their benevolent work, often carried out with short deadlines.

University of Illinois
Urbana-Champaign, IL, USA
Tamer Başar

GERAD, HEC Montréal
Montréal, QC, Canada
Georges Zaccour

Contents

Volume 1

Volume 2

About the Editors

Tamer Başar
Coordinated Science Laboratory
and Department of Electrical
and Computer Engineering
University of Illinois
at Urbana-Champaign
Urbana, IL, USA

Tamer Başar has been with the University of Illinois at Urbana-Champaign since 1981, where he currently holds the academic positions of Swanlund Endowed Chair; Center for Advanced Study professor of Electrical and Computer Engineering; professor, Coordinated Science Laboratory; professor, Information Trust Institute; and affiliate professor, Mechanical Sciences and Engineering. Since 2014, he also holds the administrative position of director of the Center for Advanced Study, and prior to that he was interim director of the Beckman Institute for Advanced Science and Technology. He received his BSEE degree from Robert College (Istanbul) and M.S., M.Phil., and Ph.D. degrees from Yale University (New Haven). Dr. Başar is a member of the US National Academy of Engineering and the European Academy of Sciences; Fellow of IEEE, IFAC, and SIAM; a past president of the IEEE Control Systems Society (CSS), the founding (and past) president of the International Society of Dynamic Games (ISDG), and a past president of the American Automatic Control Council (AACC). He has received several awards and recognitions over the years, including the highest awards of IEEE CSS, IFAC, AACC, and ISDG; the IEEE Control Systems Technical Field Award; Medal of Science of Turkey; and a number of international honorary doctorates and professorships. He has over 800 publications in systems, control, communications, optimization, networks, and dynamic games, including books on non-cooperative dynamic game theory, robust control, network security, wireless and communication networks, and stochastic networks. He is editor of several book series.

Georges Zaccour
Department of Decision Sciences
GERAD, HEC Montréal
Montreal, QC, Canada

Georges Zaccour holds the Chair in Game Theory and Management and is full professor of Management Science at HEC Montréal. He holds a Ph.D. in Management Science, an M.Sc. in International Business from HEC Montréal, and a licence in mathematics and economics from Université Paris-Dauphine. He served as director of GERAD, an interuniversity research center, and director of the Ph.D. program at HEC Montréal. His research areas are dynamic games, optimal control and operations research and their applications in management science and economics, areas in which he has published more than 170 papers and co-edited thirteen volumes. He co-authors the books *Differential Games in Marketing* and *Games and Dynamic Games*. He is editor-in-chief of Dynamic Games and Applications and fellow of The Royal Society of Canada and was president of the International Society of Dynamic Games (2002–2006).

Contributors

Saurabh Amin MIT, Cambridge, MA, USA

Joe Apaloo Department of Mathematics, Statistics and Computer Science, St Francis Xavier University, Antigonish, NS, Canada

Łukasz Balbus Faculty of Mathematics, Computer Sciences and Econometrics, University of Zielona Góra, Zielona Góra, Poland

Tamer Başar Coordinated Science Laboratory and Department of Electrical and Computer Engineering, University of Illinois at Urbana-Champaign, Urbana, IL, USA

Alexandre M. Bayen University of California at Berkeley, Berkeley, CA, USA

Joseph Z. Ben-Asher Aerospace Engineering, Technion – Israel Institute of Technology, Haifa, Israel

Pierre Bernhard Emeritus Senior Scientist, Biocore Team, Université de la Côte d'Azur-INRIA, Sophia Antipolis Cedex, France

Sadegh Bolouki Coordinated Science Laboratory, University of Illinois, Urbana, IL, USA

Michèle Breton HEC Montréal, Montréal, QC, Canada

Mark Broom Department of Mathematics, City, University London, London, UK

Matthew Brown University of Southern California, Los Angeles, CA, USA

Peter E. Caines Department of Electrical and Computer Engineering, McGill University, Montreal, QC, Canada

Pierre Cardaliaguet Université Paris-Dauphine, Paris, France

Dean Carlson Mathematical Reviews, American Mathematical Society, Ann Arbor, MI, USA

Luca Colombo Department of Economics, Deakin University, Burwood, VIC, Australia

Ross Cressman Department of Mathematics, Wilfrid Laurier University, Waterloo, ON, Canada

S. Rasoul Etesami Department of Industrial and Systems Engineering, University of Illinois at Urbana-Champaign, Urbana, IL, USA

Fei Fang University of Southern California, Los Angeles, CA, USA

Alain Haurie ORDECSYS and University of Geneva, Geneva, Switzerland

GERAD-HEC Montréal PQ, Montreal, QC, Canada

Minyi Huang School of Mathematics and Statistics, Carleton University, Ottawa, ON, Canada

Anna Jaśkiewicz Faculty of Pure and Applied Mathematics, Wrocław University of Science and Technology, Wrocław, Poland

Albert Xin Jiang Trinity University, San Antonio, TX, USA

Steffen Jørgensen Department of Business and Economics, University of Southern Denmark, Odense, Denmark

Debarun Kar University of Southern California, Los Angeles, CA, USA

Jacek B. Krawczyk Victoria University of Wellington, Wellington, New Zealand

Walid Krichene University of California at Berkeley, Berkeley, CA, USA

Vlastimil Křivan Biology Center, Czech Academy of Sciences, České Budějovice, Czech Republic

Faculty of Science, University of South Bohemia, České Budějovice, Czech Republic

Sergey Kumkov N. N. Krasovskii Institute of Mathematics and Mechanics, Ural Branch of Russian Academy of Sciences, Ekaterinburg, Russia

Paola Labrecciosa Department of Economics, Monash University, Clayton, VIC, Australia

Ngo Van Long McGill University, Montreal, QC, Canada

Roland P. Malhamé Department of Electrical Engineering, École Polytechnique de Montréal, Montreal, QC, Canada

Jason R. Marden Department of Electrical and Computer Engineering, University of California, Santa Barbara, CA, USA

Angelia Nedić School of ECEE, Arizona State University, Tempe, AZ, USA

Thanh H. Nguyen University of Southern California, Los Angeles, CA, USA

Andrzej S. Nowak Faculty of Mathematics, Computer Science and Econometrics, University of Zielona Góra, Zielona Góra, Poland

Valerii Patsko N. N. Krasovskii Institute of Mathematics and Mechanics, Ural Branch of Russian Academy of Sciences, Ekaterinburg, Russia

Vladimir Petkov Victoria University of Wellington, Wellington, New Zealand

Leon A. Petrosyan Faculty of Applied Mathematics-Control Processes, St Petersburg State University, St Petersburg, Russia

Catherine Rainer Département, Laboratoire de Mathématiques LMBA, Université de Brest, Brest Cedex, France

Kevin Reffett Department of Economics, Arizona State University, Tempe, AZ, USA

Jack D. Reilly Google Inc. Mountain View, CA, USA

Walid Saad Virginia Tech, Blacksburg, VA, USA

Anibal Sanjab Virginia Tech, Blacksburg, VA, USA

Srinivas Shakkottai Department of Electrical and Computer Engineering, Texas A&M University, College Station, TX, USA

Jeff S. Shamma Computer, Electrical and Mathematical Science and Engineering Division (CEMSE), King Abdullah University of Science and Technology (KAUST), Thuwal, Saudi Arabia

Arunesh Sinha University of Southern California, Los Angeles, CA, USA

Jason L. Speyer Mechanical and Aerospace Engineering, University of California, Los Angeles, CA, USA

R. Srikant Department of Electrical and Computer Engineering, University of Illinois at Urbana-Champaign, Urbana, IL, USA

Milind Tambe University of Southern California, Los Angeles, CA, USA

Varvara Turova Department of Mathematics, Technical University of Munich, Garching, Germany

Łukasz Woźny Department of Quantitative Economics, Warsaw School of Economics, Warsaw, Poland

David W. K. Yeung SRS Consortium for Advanced Study in Dynamic Games, Hong Kong Shue Yan University, Hong Kong, China

Georges Zaccour Department of Decision Sciences, GERAD, HEC Montréal, Montreal, QC, Canada

Aart de Zeeuw The Beijer Institute of Ecological Economics, The Royal Swedish Academy of Sciences, Stockholm, Sweden

Part I
Theory of Dynamic Games

Introduction to the Theory of Games

1

Tamer Başar

Contents

T. Başar (✉)
Coordinated Science Laboratory and Department of Electrical and Computer Engineering, University of Illinois at Urbana-Champaign, Urbana, IL, USA
e-mail: basar1@illinois.edu

T. Başar, G. Zaccour (eds.), *Handbook of Dynamic Game Theory*,
https://doi.org/10.1007/978-3-319-44374-4_1

Abstract

This chapter provides a general introduction to the theory of games, as a prelude to other chapters in this *Handbook of Dynamic Game Theory* which discuss in depth various aspects of dynamic and differential games. The present chapter describes in general terms what game theory is, its historical origins, general formulation (concentrating primarily on static games), various solution concepts, and some key results (again primarily for static games). The conceptual framework laid out here sets the stage for dynamic games covered by other chapters in the *Handbook*.

Keywords

Game theory · Dynamic games · Historical evolution of game theory · Zero-sum games · Nonzero-sum games · Strategic equivalence · Saddle-point equilibrium · Nash equilibrium · Correlated equilibrium · Stackelberg equilibrium · Computational methods · Linear-quadratic games

1 Game Theory, Its Origins, and Classifications

1.1 What Is Game Theory?

Game theory deals with strategic interactions among multiple decision-makers, called *players* (and in some context *agents*), with each player's preference ordering among multiple alternatives captured in an objective function for that player, which she either tries to maximize (in which case the objective function is a *utility* function or *benefit* function) or minimize (in which case we refer to the objective function as a *cost* function or a *loss* function). For a nontrivial game, the objective function of a player depends on the choices (*actions*, or equivalently *decision variable*) of at least one other player, and generally of all the players, and hence a player cannot simply optimize her own objective function independently of the choices of the other players. This thus brings in a coupling between the actions of the players and binds them together in decision-making even in a noncooperative environment. If the players were able to enter into a cooperative agreement so that the selection of actions or decisions is done collectively and with full trust, so that all players would benefit to the extent possible, and no inefficiency would arise, then we would be in the realm of *cooperative game theory*, where the issues of bargaining, coalition formation, excess utility distribution, etc., are of relevance. Various aspects of cooperative games, in both static and dynamic environments, are covered extensively in other chapters in this *Handbook* and will not be addressed in this introductory chapter. There are also several textbooks that cover the fundamentals of cooperative games, such as Owen (1995), Fudenberg and Tirole (1991), and Vorob'ev et al. (1977); see also the survey article (Saad et al. 2009), which focuses on applications of cooperative game theory to communication systems.

If no cooperation is allowed among the players, then we are in the realm of *noncooperative game theory*, where first one has to introduce a satisfactory solution concept. Leaving aside for the moment the issue of how the players can reach such a satisfactory solution point, let us address the issue of, assuming that the players are at such a solution point, what would be the minimum set of requirements or features one would expect to hold there. To first order, such a solution point should have the property that if all players but one stay put, then the player who has the option of moving away from the solution point should not have any incentive to do so because she cannot improve her payoff. Note that we cannot allow two or more players to move collectively from the solution point, because such a collective move requires cooperation, which is not allowed in a noncooperative game. Such a solution point where none of the players can improve her payoff by a unilateral move is known as a *noncooperative equilibrium* or *Nash equilibrium*, named after John Nash, who introduced it and proved that it exists in finite games (i.e., games where each player has only a finite number of alternatives) (Nash 1950, 1951). We will discuss this result later in this chapter, following some terminology, a classification of noncooperative games according to various attributes, and a mathematical formulation.

Another noncooperative equilibrium solution concept is the *Stackelberg equilibrium*, introduced in von Stackelberg (1934), and predating the Nash equilibrium, where there is a hierarchy in decision-making among the players, with some of the players, designated as *leaders*, having the ability to first announce their actions (and make a commitment to play them), and the remaining players, designated as *followers*, taking these actions as given in the process of computation of their noncooperative (Nash) equilibria (among themselves). Before announcing their actions, the leaders would of course anticipate these responses and determine their actions in a way such that the final outcome will be most favorable to them (in terms of their objective functions).

We say that a noncooperative game is *nonzero-sum* if the sum of the players' objective functions cannot be made zero after appropriate positive scaling and/or translation that do not depend on the players' decision variables. We say that a two-player game is *zero-sum* if the sum of the objective functions of the two players is *zero* or can be made zero by appropriate positive scaling and/or translation that do not depend on the decision variables of the players. If the two players' objective functions add up to a constant (without scaling or translation), then the game is sometimes called *constant sum*, but according to our convention, such games are also zero sum. A game is a *finite game* if each player has only a finite number of alternatives; that is, the players pick their actions out of finite sets (action sets); otherwise, the game is an *infinite game*; finite games are also known as *matrix games*. An infinite game is said to be a *continuous-kernel game* if the action sets of the players are continua and the players' objective functions are continuous with respect to action variables of all players. A game is said to be *deterministic* if the players' actions uniquely determine the outcome, as captured in the objective functions, whereas if the objective function of at least one player depends on an additional variable (*state of nature*) with a known probability distribution, then we have a *stochastic game*. A game is a *complete information* game if the description of the game (that is, the players, the objective functions, and the underlying probability distributions (if stochastic) is common information to all players; otherwise, we have an *incomplete information* game. We say that a game is *static* if players have access to only the a priori information (shared by all), and none of the players has access to information on the actions of any of the other players; otherwise, what we have is a *dynamic game*. A game is a *single-act game* if every player acts only once; otherwise, the game is *multi-act*. Note that it is possible for a single-act game to be dynamic and for a multi-act game to be static. A dynamic game is said to be a *differential game* if the evolution of the decision process (controlled by the players over time) takes place in continuous time, and generally involves a differential equation; if it takes place over a discrete-time horizon, a dynamic game is sometimes called a *discrete-time game*.

In dynamic games, as the game progresses, players acquire information (complete or partial) on past actions of other players and use this information in selecting their own actions (also dictated by the equilibrium solution concept at hand). In finite dynamic games, for example, the progression of a game involves a *tree structure* (also called *extensive form*) where each node is identified with a player along with the time when she acts, and branches emanating from a node show the possible moves of that particular player. A player, at any point in time, could generally be at more than one node, which is a situation that arises when the player does not have perfect information on the past moves of other players, and hence may not know with certainty which particular node she is at at any particular time. This uncertainty leads to a clustering of nodes into what is called *information sets* for that player. What players decide on within the framework of the extensive form is not their actions, but their *strategies*, that is what action they would take at each information set (in other words, correspondences between their information sets and their allowable actions). They then take specific actions (or actions are executed on their behalf), dictated by the strategies chosen as well as the progression of the game (decision) process along the tree. The equilibrium solution is then defined in terms of not actions but strategies.

The notion of a *strategy*, as a mapping from the collection of information sets to action sets, extends readily to infinite dynamic games, and hence in both differential games and dynamic games (also known as difference games, as evolution takes place in discrete time), Nash equilibria and Stackelberg equilibria are defined in terms of strategies (Başar and Olsder 1999). Several chapters in this *Handbook* discuss such equilibria, for both zero-sum and nonzero-sum noncooperative differential and dynamic games, with and without the presence of probabilistic uncertainty.

In the broad scheme of things, game theory, and particularly noncooperative game theory, can be viewed as an extension of two fields: *mathematical programming* and *optimal control theory*. Any problem in game theory collapses to a problem in one of these disciplines if there is only one player. One-player static games are essentially mathematical programming problems (linear programming or nonlinear programming), and one-player difference or differential games can be viewed as optimal control problems. It is therefore quite to be expected for tools of single-player optimization (like mathematical programming (Bertsekas 1999) and optimal control (Bertsekas 2007; Pontryagin et al. 1962)) to be relevant to the analysis of noncooperative games.

1.2 The Past and the Present

Game Theory has enjoyed over 70 years of scientific development, with the publication of the *Theory of Games and Economic Behavior* by John von Neumann and Oskar Morgenstern (1947) generally acknowledged to *kick start* the field. It has experienced incessant growth in both the number of theoretical results and the scope and variety of applications. As a recognition of the vitality of the field, since the 1990s, a total of 10 individuals have received Nobel Prizes in Economic Sciences for work primarily in game theory. The first such recognition was bestowed in 1994 on John Harsanyi, John Nash, and Reinhard Selten "for their pioneering analysis of equilibria in the theory of noncooperative games." The second set of Nobel Prizes in game theory went to Robert Aumann and Thomas Schelling in 2005, "for having enhanced our understanding of conflict and cooperation through game-theory analysis." The third one was in 2007, recognizing Leonid Hurwicz, Eric Maskin, and Roger Myerson, "for having laid the foundations of mechanism design theory." And the most recent one was in 2012, recognizing Alvin Roth and Lloyd Shapley "for the theory of stable allocations and the practice of market design." Also to be added to this list of highest-level awards in game theory is the Crafoord Prize in 1999 (which is the highest prize in biological sciences), which went to John Maynard Smith (along with Ernst Mayr and G. Williams) "for developing the concept of evolutionary biology," where Smith's recognized contributions had a strong game-theoretic underpinning, through his work on evolutionary games and evolutionary stable equilibrium (Smith 1974, 1982; Smith and Price 1973).

Even though von Neumann and Morgenstern's 1944 book is taken as the starting point of the scientific approach to game theory, game-theoretic notions and some isolated key results date back to much earlier years. Sixteen years earlier, in 1928, John von Neumann himself had resolved completely an open fundamental problem in zero-sum games, that *every finite two-player zero-sum game admits a saddle point in mixed strategies*, which is known as the *minimax theorem* (von Neumann 1928) – a result which Emile Borel had conjectured to be false 8 years before. Some early traces of game-theoretic thinking can be seen in the 1802 work (*Considérations sur la théorie mathématique du jeu*) of André-Marie Ampère (1775–1836), who was influenced by the 1777 writings (Essai d'Arithmétique Morale) of Georges Louis Buffon (1707–1788).

Which event or writing has really started game-theoretic thinking or approach to decision-making (in law, politics, economics, operations research, engineering, etc.) may be a topic of debate, but what is indisputable is that the second half of the twentieth century was a golden era of game theory, and the twenty-first century has started with a *big bang* and is destined to be a *platinum* era with the proliferation of textbooks, monographs, and journals covering the theory and applications (to an ever-growing breadth) of static and dynamic games.

Another indisputable fact regarding the origins (as it pertains to dynamic games – the main focus of the *Handbook*) is that in (zero-sum) differential games, the starting point was the work of Rufus Isaacs in the RAND Corporation in the early 1950s, which remained classified for at least a decade, before being made accessible to a broad readership in 1965 (Isaacs 1975); see also the review (Ho 1965), which was the first journal article to introduce the book to a broader community. Several chapters in this *Handbook* deal with Isaacs' theory, its extensions, applications, and computational aspects; another chapter discusses the impact the zero-sum differential game framework has made on robust control design, as introduced in Başar and Bernhard (1995). Extension of the game-theoretic framework to nonzero-sum differential games with Nash equilibrium as the solution concept was initiated in Starr and Ho (1969), and with Stackelberg equilibrium as the solution concept in Simaan and Cruz (1973). A systematic study of the role information structures play in the existence of such equilibria and their uniqueness or nonuniqueness (termed *informational nonuniqueness*) was carried out in Başar (1974, 1976, 1977).

2 Noncooperative Games and Equilibria

2.1 Main Elements and Equilibrium Solution Concepts

For a precise formulation of a noncooperative game, we have to specify (i) the number of players; (ii) the possible actions available to each player, and any constraints that may be imposed on them; (iii) the objective function of each player, which she attempts to optimize (minimize or maximize, as the case may be); (iv) any time ordering of the execution of the actions if the players are allowed to act more than once; (v) any information acquisition that takes place and how the information available to a player at each point in time depends on the past actions of other players; and (vi) whether there is a player (*nature*) whose action is the outcome of a probabilistic event with a fixed (known) distribution. Here we will first consider formulation of games where only items (i)–(iii) above are relevant, that is, players act only once, the game is static so that players do not acquire information on other players' actions, and there is no nature player. Subsequently, in the context of finite games, we will consider more general formulations, particularly dynamic games, that will incorporate all the ingredients listed above.

Accordingly, we consider an N-player game, with $\mathcal{N} := \{1, \ldots, N\}$ denoting the Players set.[1] The decision or action variable of Player i is denoted by $x_i \in X_i$,

[1] It is also possible to define games where the Players set is not finite. This chapter will not discuss such games, known generally as *mean field games* when some structure is imposed on the way other players' decision variables enter into the objective function of a particular (generic) player. Another chapter in this *Handbook* will address primarily such games.

where X_i is the action set of Player i. The action set could be a finite set (so that the player has only a finite number of possible actions), an infinite but finite-dimensional set (such as the unit interval, [0, 1]), or an infinite-dimensional set (such as the space of all continuous functions on the interval [0, 1]). We let x denote the N-tuple of action variables of all players, $x := (x_1, \ldots, x_N)$. Allowing for possibly coupled constraints, we let $\Omega \subset X$ be the constraint set for the game, where X is the N-product of $X_1, \ldots, X_N$; hence for an N-tuple of action variables to be feasible, we need $x \in \Omega$ (e.g., with $N = 2$, we could have a coupled constraint set described by: $0 \le x_1, x_2 \le 1$, $x_1 + x_2 \le 1$, which would arise in a resource allocation game with a hard constraint on the total amount of resource available to the two players).

If we consider the players to be minimizers, the objective function (loss function or cost function) of Player i will be denoted by $L_i(x_i, x_{-i})$, where x_{-i} stands for the action variables of all players except the i'th one. If the players are maximizers, then the objective function (utility function) of Player i will be denoted by $V_i(x_i, x_{-i})$. Note that a game where all players are minimizers, with cost functions L_i's, can be seen as one where all players are maximizers, with utility functions $V_i \equiv -L_i$, $i \in \mathcal{N}$.

Now, an N-tuple of action variables $x^* \in \Omega$ constitutes a *Nash equilibrium* (or, *noncooperative equilibrium*) (NE) if, for all $i \in \mathcal{N}$,

$$L_i(x_i^*, x_{-i}^*) \le L_i(x_i, x_{-i}^*), \quad \forall\, x_i \in X_i, \text{ such that } (x_i, x_{-i}^*) \in \Omega, \tag{1.1}$$

or, if the players are maximizers,

$$V_i(x_i^*, x_{-i}^*) \ge V_i(x_i, x_{-i}^*), \quad \forall\, x_i \in X_i, \text{ such that } (x_i, x_{-i}^*) \in \Omega. \tag{1.2}$$

If $N = 2$ and $L_1 \equiv -L_2 =: L$, then we have a two-player zero-sum game (ZSG), with Player 1 minimizing L and Player 2 maximizing the same quantity. In this case, the Nash equilibrium becomes the *saddle-point equilibrium* (SPE), which is formally defined as follows, where we leave out the coupling constraint set Ω (or simply assume it to be equal to the product set $X := X_1 \times X_2$): A pair of actions $(x_1^*, x_2^*) \in X$ is in *saddle-point equilibrium* (SPE) for a game with cost function L, if

$$L(x_1^*, x_2) \le L(x_1^*, x_2^*) \le L(x_1, x_2^*), \quad \forall (x_1, x_2) \in X. \tag{1.3}$$

This also implies that the order in which minimization and maximization are carried out is inconsequential, that is

$$\min_{x_1 \in X_1} \max_{x_2 \in X_2} L(x_1, x_2) = \max_{x_2 \in X_2} \min_{x_1 \in X_1} L(x_1, x_2) = L(x_1^*, x_2^*) =: L^*,$$

where the first expression on the left is known as the *upper value* of the game, the second expression is the *lower value* of the game, and L^* is known as the *value* of the game.[2] Note that we generally have

$$\min_{x_1 \in X_1} \max_{x_2 \in X_2} L(x_1, x_2) \geq \max_{x_2 \in X_2} \min_{x_1 \in X_1} L(x_1, x_2),$$

or more precisely

$$\inf_{x_1 \in X_1} \sup_{x_2 \in X_2} L(x_1, x_2) \geq \sup_{x_2 \in X_2} \inf_{x_1 \in X_1} L(x_1, x_2),$$

which follows directly from the obvious inequality

$$\sup_{x_2 \in X_2} L(x_1, x_2) \geq \inf_{x_1 \in X_1} L(x_1, x_2),$$

since the LHS expression is only a function of x_1 and the RHS expression only a function of x_2.

Next, note that the value of a game, whenever it exists (which certainly does if there exists a saddle point), is *unique*. Hence, if there exists another saddle-point solution, say $(\hat{x}_1, \hat{x}_2)$, then $L(\hat{x}_1, \hat{x}_2) = L^*$. Moreover, these multiple saddle points are *orderly interchangeable*, that is, the pairs $(x_1^*, \hat{x}_2)$ and $(\hat{x}_1, x_2^*)$ are also in saddle-point equilibrium. This property that saddle-point equilibria enjoy do not extend to multiple Nash equilibria (for nonzero-sum games): multiple Nash equilibria are generally not interchangeable, and furthermore, they do not lead to the same values for the players' cost functions, the implication being that when players switch from one equilibrium to another, some players may benefit from that switch (in terms of reduction in cost), while others may see an increase in their costs. Further, if the players pick their actions randomly from the set of multiple Nash equilibria of the game, then the resulting N-tuple of actions may not be in Nash equilibrium.

Now coming back to the zero-sum game, if there is no value, which essentially means that the upper and lower values are not equal, in which case the former is strictly higher than the latter,

$$\min_{x_1 \in X_1} \max_{x_2 \in X_2} L(x_1, x_2) > \max_{x_2 \in X_2} \min_{x_1 \in X_1} L(x_1, x_2),$$

then a saddle point does not exist. We then say in this case that *the zero-sum game does not have a saddle point in pure strategies*. This opens the door for looking for a *mixed-strategy* equilibrium. A *mixed strategy* for a player is a probability distribution over her action set, which we denote by p_i for Player i. This argument

[2]Upper and lower values are defined in more general terms using infimum (inf) and supremum (sup) replacing minimum and maximum, respectively, to account for the facts that minima and maxima may not exist. When the action sets are finite, however, the latter always exist.

also extends to the general N-player game, which may not have a Nash equilibrium in pure strategies (actions, in this case). In search of a mixed-strategy equilibrium, L_i is replaced by its expected value taken with respect to the mixed strategy choices of the players, which we denote for Player i by $J_i(p_1, \ldots, p_N)$. Nash equilibrium over mixed strategies is then introduced as before, with just J_i's replacing L_i's, and p_i's replacing x_i's, and $p_i \in \mathcal{P}_i$, where $\mathcal{P}_i$ is the set of all probability distributions on X_i (we do not bring Ω into the picture here, since we take the constraint sets to be rectangular). If X_i is finite, then p_i will be a probability vector, taking values in the probability simplex determined by X_i. In either case, the N-tuple $(p_1^*, \ldots, p_N^*)$ is in (mixed-strategy) Nash equilibrium (MSNE) if

$$J_i(p_i^*, p_{-i}^*) \leq J_i(p_i, p_{-i}^*), \quad \forall\, p_i \in \mathcal{P}_i. \tag{1.4}$$

This readily leads, in the case of zero-sum games, as a special case, to the following definition of a saddle point in mixed strategies: A pair (p_1^*, p_2^*) constitutes a *saddle point in mixed strategies* (or a *mixed-strategy saddle-point equilibrium*) (MSSPE), if

$$J(p_1^*, p_2) \leq J(p_1^*, p_2^*) \leq J(p_1, p_2^*), \quad \forall\, (p_1, p_2) \in \mathcal{P}.$$

where $J(p_1, p_2) = E_{p_1,p_2}[L(x_1, x_2)]$ and $\mathcal{P} := \mathcal{P}_1 \times \mathcal{P}_2$. Here $J^* = J(p_1^*, p_2^*)$ is the value of the zero-sum game in mixed strategies.

2.2 Security Strategies

If there is no Nash equilibrium in pure strategies, and the players do not necessarily want to adopt mixed strategies, an alternative approach is for each player to pick that pure strategy that will safeguard her losses under worst scenarios. This will entail each player essentially playing a zero-sum game, minimizing her cost function against collective maximization of all other players. A strategy (or an action, in this case) that provides a loss ceiling for a player is known as a *security strategy* for that player. Assuming again rectangular action product sets, security strategy $x_i^s \in X_i$ for Player i is defined through the relationship

$$\sup_{x_{-i} \in X_{-i}} L_i(x_i^s, x_{-i}) = \inf_{x_i \in X_i} \sup_{x_{-i} \in X_{-i}} L_i(x_i, x_{-i}) =: \bar{L}_i$$

where the "sup" could be replaced with "max" if the action sets are finite. Note that, the RHS value, $\bar{L}_i$, is the upper value of the zero-sum game played by Player i. Also note that even if the security strategies of the players, say $x^s := \{x_i^s,\ i \in \mathcal{N}\}$, are unique, then this N-tuple would not necessarily constitute an equilibrium in any sense. In the actual play, the players will actually end up doing better than just safeguarding their losses, since $L^i(x^s) \leq \bar{L}^i$ for all $i \in \mathcal{N}$.

The notion of a security strategy could naturally also be extended to mixed strategies. Using the earlier notation, $p_i^s \in \mathcal{P}_i$ would be a *mixed security strategy* for Player i if

$$\sup_{p_{-i}\in\mathcal{P}_{-i}} J_i(p_i^s, p_{-i}) = \inf_{p_i\in\mathcal{P}_i} \sup_{p_{-i}\in\mathcal{P}_{-i}} J_i(p_i, p_{-i}) =: \bar{J}_i$$

Remark 1. If the original game is a two-player zero-sum game, and the upper and lower values are equal, then security strategies for the players will have to be in SPE. If the upper and lower values are not equal in pure strategies, but are in mixed strategies, then mixed security strategies for the players will have to be in MSPE. ⋄

2.3 Strategic Equivalence

Strategic equivalence is a useful property that facilitates study of noncooperative equilibria of nonzero-sum games (NZSGs). Let us now make the simple observation that given an N-player NZSG of the type introduced in this section, if two operations are applied to the loss function of a player, *positive scaling* and *translation*, that do not depend on the action variable of that player, this being so for every player, then the set of NE of the resulting NZSG is identical to the set of NE of the original game. In view of this property, we say that the two games are *strategically equivalent*. In mathematical terms, if $\tilde{L}_i$'s are the cost functions of the players in the transformed game, then we have, for some functions, $\alpha_i(x_{-i}) > 0,\ \beta_i(x_{-i}),\ i \in \mathcal{N}$,[3]

$$\tilde{L}_i(x_i, x_{-i}) = \alpha_i(x_{-i})\, L_i(x_i, x_{-i}) + \beta_i(x_{-i}),\ i \in \mathcal{N}.$$

Now note that, if for a given NZSG, there exist α_i's and β_i's of the types above, such that $\tilde{L}_i$ is independent of i, that is, the transformed NZSG features the same cost function, say $\tilde{L}$, for all players, then we have a single objective game, or equivalently a *team* problem. Any NE of this transformed game (which is a team) is a *person-by-person* optimal solution of the team problem. That is, if $x_i^*,\ i \in \mathcal{N}$ is one such solution, we have

$$\tilde{L}(x_i,^* x_{-i}^*) = \min_{x_i\in X_i} \tilde{L}(x_i, x_{-i}^*),\ \forall\, i \in \mathcal{N},$$

which is not as strong as the globally minimizing solution for $\tilde{L}$:

$$\tilde{L}(x_i,^* x_{-i}^*) = \min_{x_i\in X_i} \min_{x_{-i}\in X_{-i}} \tilde{L}(x_i, x_{-i}).$$

[3]Here the positivity requirement on each α_i is uniform for all x_{-i}, that is, there exists a constant $\epsilon > 0$ such that $\alpha_i(x_{-i}) > \epsilon\ \forall\, x_{-i} \in X_{-i},\ i \in \mathcal{N}$.

Clearly, the latter implies the former, but not vice versa. Consider, for example, the two-player game where each player has two possible actions, for which $\tilde{L}$ admits the matrix representation

$$\tilde{L} = \begin{pmatrix} 0 & 1 \\ 1 & 1 \end{pmatrix},$$

where Player 1 is the row player and Player 2 the column player (and both are minimizers). The south-east entry (row 2, column 2) is clearly a person-by-person optimal solution (NE), but is not the globally minimum one, which is the north-west entry (row 1, column 1) (which is of course also a person-by-person optimal solution). Of course, if the players were to cooperate, they would unquestionably pick the latter, but since this is a noncooperative game, they are not allowed to correlate their choices. With the entries as above, however, the chances of them ending up at the global minimum are very high, because neither one would end up worse than the inferior NE if they stick to the first row and first column (even if one player inadvertently deviates). But this is not the whole story, because it would be misleading to make the *mutual benefit* argument by working on the transformed game. Consider now the following two-player, two-action NZSG, where again Player 1 is the row player and Player 2 the column player:

$$L_1 = \begin{pmatrix} 99 & 1 \\ 100 & 1 \end{pmatrix}, \quad L_2 = \begin{pmatrix} 0 & 1 \\ 1 & 1 \end{pmatrix}.$$

This game has two pure-strategy NE, (row 1, column 1) and (row 2, column 2), the same as the game $\tilde{L}$. In fact, it is easy to see that the two games are strategically equivalent (subtract 99 from the first column of L_1). But now, Player 1 would prefer the south-west entry (that is what was inferior in the transformed game), which shows that there are perils in jumping to conclusions based on a transformed game.

When this all comes handy, however, is when the transformed game as a team problem can be shown to have a unique person-by-person optimal solution, which is also the globally optimal team solution. Then, there would be no ambiguity in the selection of the unique NE.

For a given NZSG, if there exists a strategically equivalent team problem, then we say that the original game is *team like*. There could also be situations when a game is strategically equivalent to a zero-sum team problem, that is, there exists a proper subset of $\mathcal{N}$, say $\mathcal{N}_1$, such that for $i \in \mathcal{N}_1$, $\tilde{L}_i$ is independent of i, say $\tilde{L}$, **and** for $j \notin \mathcal{N}_1$, $\tilde{L}_j \equiv -\tilde{L}$. This means that there exists a strategically equivalent game where players in $\mathcal{N}_1$ form a team, playing against another team comprised of all players outside $\mathcal{N}_1$. In particular, if $N = 2$, we have every NE of the original game equal to the SPE of the transformed strategically equivalent ZSG.

3 Finite Games, and Existence and Computation of NE

3.1 Zero-Sum Finite Games and the Minimax Theorem

Let us first consider two-player zero-sum finite games, or equivalently matrix games. For any such game, we have to specify the cardinality of action sets X_1 and X_2 (card (X_1) and card (X_2)) and define the objective function $L(x_1, x_2)$ on the product of these finite sets. As per our earlier convention, Player 1 is the minimizer and Player 2 the maximizer. Let card $(X_1) = m$ and card $(X_2) = n$, that is, the minimizer has m choices and the maximizer has n choices, and let the elements of X_1 and X_2 be ordered according to some (could be arbitrary) convention. We can equivalently associate an $m \times n$ matrix A with this game, whose entries are the values of $L(x_1, x_2)$, following the same ordering as that of the elements of the action sets, that is, ij'th entry of A is the value of $L(x_1, x_2)$ when x_1 is the i'th element of X_1 and x_2 is the j'th element of X_2. Player 1's choices are then the rows of the matrix A and Player 2's are its columns.

It is easy to come of with example matrix games where a saddle point does not exist in pure strategies, with perhaps the simplest one being the game known as *matching pennies*, where

$$A = \begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix},$$

and each entry is cost to Player 1 (minimizer) and payoff to Player 2 (maximizer). Here there is no row–column combination at which the players would not have an incentive to deviate and improve their returns.

The next question is whether there exists a saddle point in mixed strategies. Assume that Player 1 now picks row 1 and row 2 with equal probability $\frac{1}{2}$. Then, regardless of whether Player 2 picks column 1 or column 2, she will face the same expected cost of 0. Hence, in response to this equal probability choice of Player 1, Player 2 is indifferent between the two actions available to her; she could pick column 1, or column 2, or any probability mix between the two. Likewise, if Player 2 picks column 1 and column 2 with equal probability $\frac{1}{2}$, this time Player 1 faces an expected cost of 0 regardless of her choice. In view of this, the mixed strategy pair $\left(p_1^* = \left(\frac{1}{2}, \frac{1}{2}\right), p_2^* = \left(\frac{1}{2}, \frac{1}{2}\right)\right)$ is a MSSPE, and in fact is the *unique* one. The SP value in mixed strategies is 0.

To formalize the above, let A be an $m \times n$ matrix representing the finite ZSG, and as before, let p_1 and p_2 be the probability vectors for Players 1 and 2, respectively (both column vectors, and note that in this case p_1 is of dimension m and p_2 is of dimension n, and components of each are nonnegative and add up to 1). We can then rewrite the expected cost function as

$$J(p_1, p_2) = p_1' A p_2.$$

By the *minimax theorem*, due to *John von Neumann* (1928), J indeed admits a saddle point, which means that the matrix game A has a saddle point in mixed strategies, that is, there exists a pair (p_1^*, p_2^*) such that for all other probability vectors p_1 and p_2, of dimensions m and n, respectively, the following pair of saddle-point inequalities hold:

$$p_1^{*\prime} A p_2 \leq p_1^{*\prime} A p_2^* \leq p_1' A p_2^*. \tag{1.5}$$

The quantity $p_1^{*\prime} A p_2^*$ is the *value* of the game in mixed strategies. This result is now captured in the following *minimax theorem*. Its proof uses the alternating hypotheses lemma in matrix theory, which says that given the matrix A as above, either there exists $y \in \mathbb{R}^m$, $y \geq 0$, such that $y'A \leq 0$, or there exists $z \in \mathbb{R}^n$, $z \geq 0$, such that $Az \geq 0$. Details can be found in Başar and Olsder (1999, p. 26).

Theorem 1. *Every finite two-person zero-sum game has a saddle point in mixed strategies.*

3.2 Neutralization and Domination

A mixed strategy that assigns positive probability to every action of a player is known as an *inner mixed strategy*. A MSSPE where both strategies are inner mixed is known as an *inner MSSPE*, or a completely mixed MSSPE. Note that if (p_1^*, p_2^*) is an inner MSSPE, then $p_1' A p_2^*$ is independent of p_1 on the m-dimensional probability simplex, and $p_2' A' p_1^*$ is independent of p_2 on the n-dimensional probability simplex. The implication is that in an inner MSSPE, all the players do is to *neutralize* each other, and the solution would be the same if their roles were reversed (i.e., Player 1 the maximizer and Player 2 the minimizer). This suggests an obvious computational scheme for solving for the MSSPE, which involves solving linear algebraic equations for p_1 and p_2, of course provided that MSSPE is inner.

Now, if MSSPE is not inner but is proper mixed, that is, it is not a pure-strategy SPE, then a similar neutralization will hold in a lower dimension. For example, if (p_1^*, p_2^*) is a MSSPE where some components of p_2^* are zero, then p_1^* will neutralize only the actions of Player 2 corresponding to the remaining components of p_2^* (which are positive), with the expected payoff for Player 2 (which is minus the cost) corresponding to the non-neutralized actions being no smaller than the neutralized ones. In this case, whether a player is a minimizer or a maximizer does make a difference. The following game, which is an expanded version of *matching pennies*, where Player 2 has a third possible action, illustrates this point:

$$A = \begin{pmatrix} 1 & -1 & 0 \\ -1 & 1 & -1 \end{pmatrix}.$$

Here, the MSSPE is $\left(p_1^* = \left(\frac{1}{2}, \frac{1}{2}\right), p_2^* = \left(\frac{1}{2}, \frac{1}{2}, 0\right)\right)$, where Player 1 neutralizes only the first two actions of Player 2, with the expected cost of the third action being $-\frac{1}{2}$, lower than 0, and hence Player 2, being the maximizer, would not put any positive weight on it. Note that in this game, it does make a difference whether a player is a minimizer or a maximizer, because if we reverse the roles (now Player 1 is the maximizer, and Player 2 the minimizer), the SP value in mixed strategies is no longer 0, but is $-\frac{1}{3}$, with the MSSPE being $\left(p_1^* = \left(\frac{2}{3}, \frac{1}{3}\right), p_2^* = \left(0, \frac{1}{3}, \frac{2}{3}\right)\right)$. Player 2 ends up not putting positive probability to the first column, which is *dominated* by the third column. *Domination* can actually be used to eliminate columns and/or rows which will not affect the MSSPE, and this will lead to reduction in the size of the game (and hence make computation of MSSPE more manageable). MSSPE of a reduced ZS matrix game (reduced through domination) is also an MSSPE of the original ZSG (with appropriate lifting to the higher dimension, by assigning eliminated columns or rows zero probability), but in this process, some mixed SP strategies may also be eliminated. This, however, is not a major issue because the MSSPE value is unique. If only strictly dominated rows and columns are eliminated,[4] then all mixed SP strategies are preserved (see, Başar and Olsder 1999; Vorob'ev et al. 1977).

3.3 Offline Computation of MSSPE

We have seen in the previous subsection that inner MSSPE can be computed using the idea of neutralization and solving linear algebraic equations. The same method can in principle be applied to MSSPE that are not inner, but then one has to carry out an enumeration by setting some components of the probability vectors to zero, and looking for neutralization in a reduced dimension – a process that converges because MSSPE exists by the minimax theorem. In this process, domination can be used (as discussed above) to eliminate some rows or columns, which would sometimes lead to a (reduced) game with an inner MSSPE.

Yet another approach to computation of MSSPE is a graphical one, which however is practical only when one of the players has only two possible actions (Başar and Olsder 1999, pp. 29–31). And yet another offline computational method is to use the powerful tool of linear programming (LP). One can actually show that there is a complete equivalence between a matrix game and an LP. The following proposition captures this result, a proof of which can be found in Başar and Olsder (1999).

[4]In ZS matrix games, under the convention adopted in this chapter regarding the roles of the players, we say that *a row strictly dominates another row* if the difference between the two vectors (first one minus the second one) has all negative entries. Likewise, *a column strictly dominates another column* if the difference has all positive entries.

Proposition 1. *Given a ZS matrix game described by the $m \times n$ matrix A, let B be another matrix game (strategically equivalent to A), obtained from A by adding an appropriate positive constant to make all its entries positive. Let $V_m(B)$ denote the SP value of B in mixed strategies. Introduce the two LPs*[5]*:*

Primal LP: $\max y'\mathbf{1}_m$ *such that* $B'y \leq \mathbf{1}_n,\ y \geq 0$
Dual LP: $\min z'\mathbf{1}_n$ *such that* $Bz \geq \mathbf{1}_m,\ z \geq 0,$

with their optimal values (if they exist) denoted by V_p and V_d, respectively. Then:

(i) *Both LPs admit solutions, and $V_p = V_d = 1/V_m(B)$.*
(ii) *If (y^*, z^*) solves matrix game, B, $y^*/V_m(B)$ solves the primal LP, and $z^*/V_m(B)$ solves the dual LP.*
(iii) *If $\tilde{y}^*$ solves the primal LP, and $\tilde{z}^*$ solves the dual LP, the pair $(\tilde{y}^*/V_p, \tilde{z}^*/V_d)$ constitutes a MSSPE for the matrix game B, and hence for A, and $V_m(B) = 1/V_p$.*

3.4 Nonzero-Sum Finite Games and Nash's Theorem

We now move on to N-player NZS finite games, and study the Nash equilibrium (NE), introduced earlier. As in the case of ZSGs, it is easy to come up with examples of games that do not admit NE in pure strategies. The question then is whether there is a counterpart of the minimax theorem in this case, which guarantees the existence of NE in mixed strategies. This is indeed the case – a result established by *John Nash* (1951) and captured in the following theorem.

Theorem 2. *Every finite N-player nonzero-sum game has a Nash equilibrium in mixed strategies.*

Note that clearly the minimax theorem follows from this one since ZSGs are special cases of NZSGs. The main difference between the two, however, is that in ZSGs, the *value* is unique (even though there may be multiple saddle-point solutions), whereas in genuine NZSGs, the expected cost N-tuple to the players under multiple Nash equilibria need not be the same. In ZSGs, multiple equilibria have the ordered interchangeability property, whereas in NZSGs they do not, as we have discussed earlier.

The notions of inner mixed equilibria, neutralization, and domination introduced earlier in the context of SPE and MSSPE equally apply here, and particularly the inner MSNE also has the neutralization property and can be solved using algebraic equations. These equations, however, will not be linear unless $N = 2$, that is, the NZSG is a bi-matrix game. In two-player NZSGs, a counterpart of the LP

[5] The notation $\mathbf{1}_m$ below stands for the m-dimensional column vector whose entries are all 1s.

equivalence exists, but this time it is a bilinear program, as captured in the following proposition; for a proof, see Başar and Olsder (1999, pp. 96–97).

Proposition 2. *For a bi-matrix game* (A, B)*, where players are minimizers, a pair* (y^*, z^*) *constitutes a MSNE if, and only if, there exists a pair of real numbers* (p^*, q^*) *such that the quadruple* (y^*, z^*, p^*, q^*) *solves the bilinear program:*

$$\min_{y,z,p,q} \left[y'AZ + y'Bz + p + q\right]$$

such that

$$Az \geq -p\mathbf{1}_m, \quad B'y \geq -q\mathbf{1}_n, \quad y \geq 0,\ z \geq 0, \quad y'\mathbf{1}_m = 1,\ z'\mathbf{1}_n = 1.$$

3.5 Online Computation of MSSPE and MSNE: Fictitious Play

In the discussion of the computation of MSNE, as well as MSSPE, we have so far focused on offline methods, where the assumption was that the players have access to the entire game parameters (including other players' payoff or cost matrices). This, however, may not always be possible, which then begs the question of whether it would be possible for the players to end up at a MSSPE or MSNE by following a process where each one observes others' actions in a repetition of the game and builds probabilistic beliefs (empirical probabilities) on other players' moves. Such a process is known as a *fictitious play* (FP). We say that the process converges in beliefs to equilibrium (MSSPE or MSNE, as the case may be) if the sequence of beliefs converges to an equilibrium. We further say that a game has the *fictitious play property* (FPP) if every fictitious play process converges in beliefs to equilibrium.

The fictitious play process was first suggested in 1949 by Brown (1951) as a mechanism to compute MSNE of a finite NZSG. Robinson (1951) then proved in 1950 that every two-player ZSG has the FPP. Miyasawa (1961) proved in 1961 (using a particular tie-breaking rule) that every 2×2 bi-matrix game has the FPP. Shapley (1964) constructed in 1964 an example of a 3×3 bi-matrix game which does not have the FPP. Last 20 years have seen renewed interest and activity on FP, with some representative papers being Monderer and Shapley (1996), Shamma and Arslan (2004), Shamma and Arslan (2005), Nguyen et al. (2010a,b), as well as Alpcan and Başar (2011), which discusses applications to security games. Some details on fictitious play within the context of learning in games can be found in another chapter of the *Handbook*.

3.6 Correlated Equilibrium

One undesirable property of NE is that it is generally *inefficient*, meaning that if the players had somehow correlated their choices of their actions, or better had collaborated in their selections, at least some of them would be able to do better in terms of the outcome (than any of the NE), and the remaining ones would be doing no worse. In mathematical terms, NE is generally *not Pareto efficient*, that is, if $x^* \in X$ is a NE, it would be possible to find another N-tuple $\tilde{x} \in X$ such that $L_i(\tilde{x}) \leq L_i(x^*)$ for all $i \in \mathcal{N}$, with strict inequality for at least one i.

The question then arises as how to improve the costs (or payoffs) to players while still preserving the noncooperative nature of the decision process. One way of doing this is through incentive strategies, or mechanism design, which will require the introduction of another player who would dictate the decision-making process (or equip one of the players with this hierarchical role), having also access to at least partial information on the actions of the other players; this would then definitely make the underlying game a dynamic game (Başar and Olsder 1999). Another way is to correlate the choices of the players through some signaling mechanisms, which leads to the notion of *correlated equilibrium* (CE), introduced by Aumann (1974, 1987), which is briefly discussed below.

Starting with an example scenario, consider the situation faced by two drivers when they meet at an intersection (simultaneously). If both proceed, then that will lead to collision, and hence result in extreme cost to both drivers. If both yield, then they lose some time, which entails some cost, whereas if one yields and the other one proceeds, then the one that yields incurs some cost and the one that proceeds receives positive payoff. This can be modeled as a two-player 2×2 bi-matrix game, of the type below (where the first row and first column correspond to *Cross* (C), and the second row and the second column correspond to *Yield* (Y), and both players are minimizers):

$$\text{Intersection Game:} \quad \begin{pmatrix} (10, 10) & (-5, 0) \\ (0, -5) & (1, 1) \end{pmatrix}$$

The game admits two pure-strategy NE, (C, Y) and (Y, C), and one MSNE, $\big((\frac{3}{8}, \frac{5}{8}), (\frac{3}{8}, \frac{5}{8})\big)$. The costs to the players (drivers) under the two pure-strategy NE are $(-5, 0)$ and $(0, -5)$, respectively, and under the MSNE (expected cost) $(\frac{5}{8}, \frac{5}{8})$. Note that both pure-strategy NE are uniformly better than the MSNE for both players, and therefore so is any convex combination of the two: $(-5\lambda, -5(1-\lambda))$, $\lambda \in [0, 1]$. Any pair in this convex combination can be achieved through correlated randomization, but the question is how such outcomes (or even better ones) can be attained through noncooperative play. How can a randomization device be installed without any enforcement?

Of course, an obvious answer is to install a *traffic light*, which would function as a randomization device which, with a certain probability, would tell the players whether to cross or yield. Note that such a *signal* would help the players to correlate their actions. For example, if the traffic light tells with probability 0.55 cross to Player 1 (green light), and yield to Player 2 (red light); with probability 0.4 the other way around; and with the remaining probability (0.05) yield to both players, then the resulting expected cost pair is $(-2.7, -1.95)$. Note that these expected costs add up to −4.65, which is somewhat worse than any convex combination of the two pure-strategy NE (where the sum is −5), but it is a *safe* outcome and can only be achieved through correlation. Another noteworthy point is that actually the players do not have to obey the traffic light, but once it is there it is to their advantage to use it as a signal to correlate their moves; in that sense, what this yields is an *equilibrium*, which is called a *correlated equilibrium* (CE). We now proceed with a precise definition of CE for bi-matrix games.

Consider a bi-matrix game (A, B), where the matrices are $m \times n$. Consider a randomization device which with probability p_{ij} signals Player 1 to use row i and Player 2 to use column j. This generates an $m \times n$ probability matrix

$$P = \{p_{ij}\}, \ p_{ij} \geq 0, \quad \sum_i \sum_j p_{ij} = 1,$$

which we call a *correlated mixed strategy* (CMS). Such a strategy is in equilibrium if, whenever the signal dictates Player 1 to use row i, his expected cost cannot be lower by using some other action, i.e.,

$$\sum_{j=1}^{n} \left[a_{ij} p_{ij} / \sum_{\ell} p_{i\ell} \right] \leq \sum_{j=1}^{n} \left[a_{kj} p_{ij} / \sum_{\ell} p_{i\ell} \right] \quad \forall k \neq i,$$

which can equivalently be written as

$$\sum_{j=1}^{n} \left(a_{ij} - a_{kj} \right) p_{ij} \leq 0 \ \ \forall k \neq i. \tag{1.6}$$

Likewise for Player 2, if j is the signal,

$$\sum_{i=1}^{m} \left(b_{ij} - b_{i\ell} \right) p_{ij} \leq 0 \ \ \forall \ell \neq j. \tag{1.7}$$

Definition 1. A *correlated equilibrium* (CE) for the bi-matrix game (A, B) is a correlated mixed strategy P that satisfies (1.6) for all $i = 1, \ldots, m$, and (1.7) for all $j = 1, \ldots, n$.

Remark 2. If x is a mixed strategy for Player 1 and y is a mixed strategy for Player 2, then $P = xy'$ is a correlated mixed strategy for the bi-matrix game. Note that in this case $p_{ij} = x_i y_j$. But this is only a one-direction relationship, because not all correlated mixed strategies can be written this way. Hence, the set of all correlated mixed strategies for the bi-matrix game is larger than the set of all mixed strategy pairs. Furthermore, if (x^*, y^*) is a MSNE, then $P^* = x^* y^{*\prime}$ is a CE, which then implies that CE always exists. ◇

4 Games in Extensive Form

If players act in a game more than once, and at least one player has information (complete or partial) on past actions of other players, then we are in the realm of *dynamic games* (as mentioned earlier), for which a complete description (in finite games) involves a tree structure where each node is identified with a player along with the time when she acts, and branches emanating from a node show the possible moves of that particular player. A player, at any point in time, could generally be at more than one node – which is a situation that arises when the player does not have complete information on the past moves of other players, and hence may not know with certainty which particular node she is at at any particular time. This uncertainty leads to a clustering of nodes into what is called *information sets* for that player. A precise definition of extensive form of a dynamic game now follows.

Definition 2. *Extensive form* of an N-person nonzero-sum finite game without chance moves is a tree structure with

(i) a specific vertex indicating the starting point of the game,
(ii) N *cost functions*, each one assigning a real number to each terminal vertex of the tree, where the ith cost function determines the loss to be incurred to Player i,
(iii) a partition of the nodes of the tree into N *player sets*,
(iv) a subpartition of each player set into *information sets* $\{\eta^i_j\}$, such that the same number of branches emanate from every node belonging to the same information set and no node follows another node in the same information set. ◇

What players decide on within the framework of the extensive form is not their actions but their *strategies*, that is, what action they should take at each information set. They then take specific actions (or actions are executed on their behalf), dictated by the strategies chosen as well as the progression of the game (decision) process along the tree. A precise definition now follows.

Definition 3. Let N^i denote the class of all information sets of Player i, with a typical element designated as η^i. Let $U^i_{\eta^i}$ denote the set of alternatives of Player i at the nodes belonging to the information set η^i. Define $U^i = \cup U^i_{\eta^i}$, where the union is over $\eta^i \in N^i$. Then, a *strategy* γ^i for Player i is a mapping from N^i into U^i, assigning one element in U^i for each set in N^i, and with the further property that $\gamma^i(\eta^i) \in \mathbf{U}^i_{\eta^i}$ for each $\eta^i \in \mathcal{N}^i$. The set of all strategies of Player i is called his *strategy set* (*space*), and it is denoted by Γ^i. ◇

Let $J^i(\gamma^1, \ldots, \gamma^N)$ denote the loss incurred to Player i when the strategies $\gamma^1 \in \Gamma^1, \ldots, \gamma^N \in \Gamma^N$ are adopted by the players. This construction leads to what is known as the *normal form* of the dynamic game, which in a sense is no different from the matrix forms we have seen in the earlier sections. In particular, for a finite game with a finite duration (i.e., players act only a finite number of times), the number of elements in each Γ^i is finite, and hence the game can be viewed as a matrix game, of the type considered earlier. In this normal form, the concept of Nash equilibrium (NE) is introduced in exactly the same way as in static games, with now the action variables replaced by strategies. Hence, we have[6]:

Definition 4. An N-tuple of strategies $\gamma^* := \{\gamma^{1*}, \gamma^{2*}, \ldots, \gamma^{N*}\}$ with $\gamma^{i*} \in \Gamma^i$, $i \in \mathcal{N}$ constitutes a *noncooperative (Nash) equilibrium solution* for an N-person nonzero-sum finite game in extensive form, if the following N inequalities are satisfied for all $\gamma^i \in \Gamma^i$, $i \in \mathcal{N}$[7]:

$$J^{1*} := J^i(\gamma^{i*}, \gamma^{-i*}) \leq J^i(\gamma^i, \gamma^{-i*}).$$

The N-tuple of quantities $\{J^{1*}, \ldots, J^{N*}\}$ is known as **a** *Nash equilibrium outcome* of the nonzero-sum finite game in extensive form. ◇

Note that the word **a** is emphasized in the last sentence of the preceding definition, since NE solution could possibly be nonunique, with the corresponding set of NE values being different. This then leads to a partial ordering in the set of all NE solutions.

As in the case of static (matrix) games, pure-strategy NE may not exist in dynamic games also. This leads to the introduction of mixed strategies, which are defined (quite analogously to the earlier definition) as probability distributions on Γ^i's, that is, for each player as a probability distribution on the set of all her pure strategies; denote such a collection for Player i by $\bar{\Gamma}^i$. A MSNE is then defined in exactly the same way as before. Again, since in normal form a finite dynamic game with a finite duration (to be referred to henceforth as *finite-duration multi-act finite games*) can be viewed as a matrix game, there will always exist a MSNE by Nash's theorem:

Proposition 3. *Every N-person nonzero-sum finite-duration multi-act finite game in extensive form admits a Nash equilibrium solution in mixed strategies (MSNE).*

A MSNE may not be desirable in a multi-act game, because it allows for a player to correlate her choices across different information sets. A *behavioral strategy*,

---

[6]Even though the discussion in this section uses the framework of N-player noncooperative games with NE as the solution concept, it applies as a special case to two-player zero-sum games, by taking $J^1 = -J^2$ and noting that in this case NE becomes SPE.

[7]Using the earlier convention, the notation γ^{-i} stands for the collection of all players' strategies, except the i'th one.

on the other hand, allows a player to assign independent probabilities to the set of actions at each information set (that is independent across different information sets); it is an appropriate mapping whose domain of definition is the class of all the information sets of the player. By denoting the behavioral strategy set of Player i by $\hat{\Gamma}^i$, and the average loss incurred to Player i as a result of adoption of the behavioral strategy N-tuple $\{\hat{\gamma}^1 \in \hat{\Gamma}^1, \ldots, \hat{\gamma}^N \in \hat{\Gamma}^N\}$ by $\hat{J}(\hat{\gamma}^1, \ldots, \hat{\gamma}^N)$, the definition of a Nash equilibrium solution in behavioral strategies (BSNE) may be obtained directly from Definition 4 by replacing γ^i, Γ^i, and J^i with $\hat{\gamma}^i$, $\hat{\Gamma}^i$, and $\hat{J}^i$, respectively. A question of interest now is whether a BSNE is necessarily also a MSNE. The following proposition settles that Başar and Olsder (1999).

Proposition 4. *Every BSNE of an N-person nonzero-sum multi-act game also constitutes a Nash equilibrium in the larger class of mixed strategies (i.e., a MSNE).*

Even though MSNE exists in all finite-duration multi-act finite games, there is no guarantee that BSNE will exist. One can in fact construct games where a BSNE will not exist, but it is also possible to impose structures on a game so that BSNE will exist; for details, see Başar and Olsder (1999, p. 127).

Given multi-act games which are identical in all respects except in the construction of the information sets, one can introduce a partial ordering among them depending on the *relative richness* of their strategy sets (induced by the information sets). One such ordering is introduced below, followed by a specific result that it leads to; for a proof, see Başar and Olsder (1999).

Definition 5. Let I and II be two N-person multi-act nonzero-sum games with fixed orders of play, and with the property that at the time of her act each player has perfect information concerning the current level of play, that is, no information set contains nodes of the tree belonging to different levels of play. Further let Γ^i_{I} and Γ^i_{II} denote the strategy sets of Player i in I and II, respectively. Then, I is *informationally inferior* to II if $\Gamma^i_{\mathrm{I}} \subseteq \Gamma^i_{\mathrm{II}}$ for all $i \in \mathcal{N}$, with strict inclusion for at least one i. $\diamond$

Proposition 5. *Let I and II be two N-person multi-act nonzero-sum games as introduced in Definition 5, so that I is informationally inferior to II. Then,*

(i) *any NE for I is also a NE for II,*
(ii) *if $\{\gamma^1, \ldots, \gamma^N\}$ is a NE for II so that $\gamma^i \in \Gamma^i_I$ for all $i \in \mathcal{N}$, then it is also a NE for I.*

An important conclusion to be drawn from the result above is that dynamic games will generally admit a plethora of NE, because for a given game, the NE of all inferior games will also constitute NE of the original game, and these are generally not even partially orderable – which arises due to informational richness. We call such occurrence of multiple NE *informational nonuniqueness*.

5 Refinements on Nash Equilibrium

As we have seen in the previous sections, finite NZSGs will generally have multiple NE, in both pure and mixed strategies, and these equilibria are generally not interchangeable, with each one leading to a different set of equilibrium cost values or payoff values to the players, and they are not strictly ordered. In dynamic games, the presence of multiple NE is more a rule rather than an exception, with the multiplicity arising in that case because of the *informational richness* of the underlying decision problem (in addition to the structure of the players' cost matrices). As a means of shrinking the set of Nash equilibria in a rational way, refinement schemes have been introduced in the literature; we discuss in this section some of those relevant to finite games. Refinement schemes for infinite games are discussed in a different chapter of the *Handbook*.

To motivate the discussion, let us start with a two-player matrix game (A, B) where the players are minimizers and have identical cost matrices (which is what we called a *team problem* earlier).

$$A = B = \begin{array}{c} \\ U \\ D \\ \\ \end{array}\begin{array}{c} \textbf{P2} \\ \begin{array}{|c|c|} \hline 0 & 1 \\ \hline 1 & 1 \\ \hline \end{array} \\ L \;\; R \end{array} \;\textbf{P1} \tag{1.8}$$

The game admits two pure-strategy Nash equilibria: (U, L) and (D, R). Note, however, that if we perturb the entries of the two matrices slightly, and independently

$$A + \Delta A = \begin{array}{c} \textbf{P2} \\ \begin{array}{|c|c|} \hline \epsilon_{11}^1 & 1+\epsilon_{12}^1 \\ \hline 1+\epsilon_{21}^1 & 1+\epsilon_{22}^1 \\ \hline \end{array} \end{array} \textbf{P1}; \quad B + \Delta B = \begin{array}{c} \textbf{P2} \\ \begin{array}{|c|c|} \hline \epsilon_{11}^2 & 1+\epsilon_{12}^2 \\ \hline 1+\epsilon_{21}^2 & 1+\epsilon_{22}^2 \\ \hline \end{array} \end{array} \textbf{P1};$$

where $\epsilon_{ij}^k, i, j, k = 1, 2$, are infinitesimally small (positive or negative) numbers, then (U, L) will still retain its equilibrium property (as long as $|\epsilon_{ij}^k| < 1/2$), but (D, R) will not. More precisely, there will exist infinitely many perturbed versions of the original game for which (D, R) will not constitute a Nash equilibrium. Hence, in addition to admissibility,[8] (U, L) can be singled out in this case as the Nash solution that is *robust* to infinitesimal perturbations in the entries of the cost matrices.

Can such perturbations be induced naturally by some behavioral assumptions imposed on the players? The answer is yes, as discussed next. Consider the scenario where a player who intends to play a particular pure strategy (out of a set of n possible alternatives) errs and plays with some small probability one of the other $n - 1$ alternatives. In the matrix game (1.8), for example, if both players err with

[8] An NE is said to be *admissible* if there is no other NE which yields better outcome for all players.

equal (independent) probability $\epsilon > 0$, the resulting matrix game is (A_ϵ, B_ϵ), where

$$A_\epsilon = B_\epsilon = \begin{array}{c|c|c|} & \multicolumn{2}{c}{\mathbf{P2}} \\ \cline{2-3} U & \epsilon(2-\epsilon) & 1-\epsilon+\epsilon^2 \\ \cline{2-3} D & 1-\epsilon+\epsilon^2 & 1-\epsilon^2 \\ \cline{2-3} & L & R \end{array} \quad \mathbf{P1}$$

Note that for all $\epsilon \in (0, 1/2)$, this matrix game admits the unique Nash equilibrium (U, L), with a cost pair of $(\epsilon(2-\epsilon), \epsilon(2-\epsilon))$, which converges to $(0, 0)$ as $\epsilon \downarrow 0$, thus recovering one of the NE cost pairs of the original game. An NE solution that can be recovered this way is known as a *perfect equilibrium*, which was first introduced in precise terms by Selten (1975), in the context of N-player games in extensive form.[9] Given a game of perfect recall,[10] denoted $\mathcal{G}$, the idea is to generate a sequence of games, $\mathcal{G}_1, \mathcal{G}_2, \ldots, \mathcal{G}_k, \ldots$, a limiting equilibrium solution of which (in behavioral strategies, and as $k \to \infty$)[11] is an equilibrium solution of $\mathcal{G}$. If $\mathcal{G}_k$ is obtained from $\mathcal{G}$ by forcing the players at each information set to choose every possible alternative with positive probability (albeit small, for those alternatives that are not optimal), then the equilibrium solution(s) of $\mathcal{G}$ that are recovered as a result of the limiting procedure above is (are) called *perfect equilibrium* (*equilibria*).[12] Selten has shown in Selten (1975) that every finite game in extensive form with perfect recall (and as a special case in normal form) admits at least one perfect equilibrium, thus making this refinement scheme a legitimate one.

The procedure discussed above, which amounts to "completely" perturbing a game with multiple equilibria, is one way of obtaining perfect equilibria; yet another one, as introduced by Myerson (1978), is to restrict the players to use completely mixed strategies (with some lower positive bound on the probabilities) at each information set. Again referring back to the matrix game (A, B) of (1.8), let the players' mixed strategies be restricted to the class

$$\hat{\gamma}^1 = \begin{cases} U \text{ w.p. } & y \\ D \text{ w.p. } & 1-y \end{cases} \; ; \quad \hat{\gamma}^2 = \begin{cases} L \text{ w.p. } & z \\ R \text{ w.p. } & 1-z \end{cases}$$

---

[9] Selten's construction and approach also apply to static games of the types discussed heretofore, where slight perturbations are made in the entries of the matrices, instead of at information sets.

[10] A game is one with *perfect recall* if all players recall their past moves – a concept that applies to games in extensive form.

[11] As introduced in the previous section, *behavioral strategy* is a mixed strategy for each information set of a player (in a dynamic game in extensive form). When the context is static games, it is identical with mixed strategy.

[12] This is also called "trembling hand equilibrium," as the process of erring at each information set is reminiscent of a "trembling hand" making unintended choices with small probability. Here, as $k \to \infty$, this probability of unintended plays converges to zero.

where $\epsilon \le y \le 1-\epsilon$, $\epsilon \le z \le 1-\epsilon$, for some (sufficiently small) positive ϵ. Over this class of strategies, the average cost functions of the players will be

$$\hat{J}^1 = \hat{J}^2 = -yz + 1,$$

which admits (assuming that $0 \le \epsilon < \frac{1}{2}$) a unique Nash equilibrium:

$$p_\epsilon^{1*} = \hat{\gamma}_\epsilon^{2*} = \begin{cases} L \text{ w.p. } 1-\epsilon \\ R \text{ w.p. } \epsilon \end{cases} ; \quad \hat{J}_\epsilon^{1*} = \hat{J}_\epsilon^{2*} = 1-(1-\epsilon)^2.$$

Such a solution is called an ϵ*-perfect equilibrium* (Myerson and Selten 1978), which in the limit as $\epsilon \downarrow 0$ clearly yields the perfect Nash equilibrium obtained earlier. Myerson in fact proves, for N-person games in normal form, that every perfect equilibrium can be obtained as the limit of an appropriate ϵ-perfect equilibrium (Myerson and Selten 1978), with the converse statement also being true. More precisely, letting y^i denote a mixed strategy for Player i, and Y^i the simplex of probabilities, we have:

Proposition 6. *For an N-person finite game in normal form, an MSNE $\{y^{i^*} \in Y^i, i \in \mathcal{N}\}$ is a perfect equilibrium if, and only if, there exist some sequences $\{\epsilon_k\}_{k=1}^\infty$, $\{y_{\epsilon_k}^i \in \overset{\circ}{Y}{}^i, i \in \mathcal{N}\}_{k=1}^\infty$ such that*

(i) $\epsilon_k > 0$ *and* $\lim_{k\to\infty} \epsilon_k = 0$
(ii) $\{y_{\epsilon_k}^i, i \in \mathcal{N}\}$ *is an ϵ_k-perfect equilibrium*
(iii) $\lim_{k\to\infty} y_{\epsilon_k}^i = y^{i^*}$, $i \in \mathcal{N}$.

Furthermore, a perfect equilibrium necessarily exists, and every perfect equilibrium is an NE. ◇

Even though perfect equilibrium provides a refinement of Nash equilibrium with some appealing properties, it also carries some undesirable features as the following example of an identical cost matrix game (due to *Myerson* (1978) exhibits:

$$A = B = \begin{array}{c} \\ U \\ M \\ D \\ \\ \end{array} \begin{array}{c} \mathbf{P}2 \\ \begin{array}{|c|c|c|} \hline 0 & 1 & 10 \\ \hline 1 & 1 & 8 \\ \hline 10 & 8 & 8 \\ \hline \end{array} \\ L \quad M \quad R \end{array} \; \mathbf{P}1 \tag{1.9}$$

Note that this is a matrix game derived from (1.8) by adding a completely dominated row and a completely dominated column. It now has three Nash equilibria: (U, L), (M, M), (D, R), the first two of which are perfect equilibria, while the last one is not. Hence, inclusion of completely dominated rows and columns could create additional perfect equilibria not present in the original game – a feature that is clearly not desirable. To remove this shortcoming of perfect equilibrium, Myerson introduced in Myerson and Selten (1978) what is called *proper equilibria*, which corresponds to a particular construction of the sequence of strategies used in Proposition 6. *Proper equilibrium* is defined as in Proposition 6, with only the ϵ_k*-perfect equilibrium* in ii) replaced by the notion of ϵ_k*-proper equilibrium* to be introduced next. Toward this end, let $\bar{J}^i(j; y_\epsilon)$ denote the average cost to Player i when she uses her j'th strategy (such as j'th column or row of the matrix) in the game and all the other players use their mixed strategies y_ϵ^k, $k \in \mathcal{N}$, $k \neq i$. Furthermore, let $y_\epsilon^{i,j}$ be the probability attached to her j'th strategy under the mixed strategy y_ϵ^i. Then, the N-tuple $\{y_\epsilon^i, i \in \mathcal{N}\}$ is said to be in ϵ*-proper equilibrium* if the strict inequality

$$\bar{J}^i(j; y_\epsilon) > \bar{J}^i(k; y_\epsilon)$$

implies that $y_\epsilon^{i,j} \leq \epsilon\, y_\epsilon^{i,k}$, this being so for every $j, k \in \mathbf{M}_i$,[13] and every $i \in \mathcal{N}$. In other words, an ϵ-proper equilibrium is one in which every player is giving his better responses much more probability weight than this worse responses (by a factor $1/\epsilon$), regardless of whether those "better" responses are "best" or not. Myerson proves in Myerson and Selten (1978) that such an equilibrium necessarily exists, that is:

Proposition 7. *Every finite N-player game in normal form admits at least one proper equilibrium. Furthermore, every proper equilibrium is a perfect equilibrium (but not vice versa).* ◇

Remark 3. Note that in the matrix game (1.9), there is only one proper equilibrium, which is (U, L), the perfect equilibrium of (1.8). ◇

Another undesirable feature of a perfect equilibrium is that it is very much dependent on whether the game is in extensive or normal form (whereas the Nash equilibrium property is form independent). As it has been first observed by Selten (1975), and further elaborated on by van Damme (1984), a perfect equilibrium of the extensive form of a game need not be perfect in the normal form, and conversely a perfect equilibrium of the normal form need not be perfect in the extensive form. To remove this undesirable feature, van Damme introduced the concept of *quasi-perfect* equilibria for games in extensive form, and has shown that a proper equilibrium of a normal form game induces a quasi-perfect equilibrium in every extensive form game having this normal form van Damme (1984, 1987). Quasi-perfect equilibrium is defined as a behavioral strategy combination which prescribes at every information set a choice that is optimal against mistakes ("trembling hands") of the other players; its difference from perfect equilibrium is that here

[13] $\mathbf{M}_i$ is the set of all pure strategies of Player 1, with corresponding labeling of positive integers.

in the construction of perturbed matrices, each player ascribes "trembling hand" behavior to all other players (with positive probability) but not to himself.

Other types of refinement have also been proposed in the literature, such as *sequential equilibria* (Kreps and Wilson 1982) and *strategic equilibria* (Kohlberg and Mertens 1986), which are not further discussed here. None of these, however, are uniformly powerful, in the sense of shrinking the set of Nash equilibria to the smallest possible set. This topic of "refinement on Nash equilibria" has further been discussed in another chapter of the *Handbook* in the context of infinite dynamic games and with emphasis placed on the issue of time consistency. In the context of infinite dynamic games, *Başar* introduced in Başar (1976) stochastic perturbations in the system dynamics ("trembling dynamics") to eliminate multiplicity of Nash equilibria.

6 Hierarchical Finite Games and Stackelberg Equilibria

The Nash equilibrium solution concept heretofore discussed in this chapter provides a reasonable noncooperative equilibrium solution for nonzero-sum games when the roles of the players are symmetric, that is to say, when no single player dominates the decision process. However, there are yet other types of noncooperative decision problems wherein one of the players has the ability to enforce her strategy on the other player(s), and for such decision problems, one has to introduce a hierarchical equilibrium solution concept. Following the original work of *H. von Stackelberg* (1934), the player who holds the powerful position in such a decision problem is called the *leader*, and the other players who react (rationally) to the leader's decision (strategy) are called the *followers*. There are, of course, cases of multiple levels of hierarchy in decision-making, with many leaders and followers; but for purposes of brevity and clarity in exposition, the discussion is confined here to hierarchical decision problems which incorporate only two players – one leader and one follower.

6.1 Stackelberg Equilibria in Pure Strategies

To set the stage to introduce the hierarchical (Stackelberg) equilibrium solution concept, let us first consider the bi-matrix game (A, B) displayed (under our standard convention) as

$$A = \begin{array}{c|c|c|c|} & \multicolumn{3}{c}{\mathbf{P}2} \\ \cline{2-4} L & 0 & 2 & 3/2 \\ \cline{2-4} M & 1 & 1 & 3 \\ \cline{2-4} R & -1 & 2 & 2 \\ \cline{2-4} \multicolumn{1}{c}{} & \multicolumn{1}{c}{L} & \multicolumn{1}{c}{M} & \multicolumn{1}{c}{R} \end{array} \ \mathbf{P}1, \qquad B = \begin{array}{c|c|c|c|} & \multicolumn{3}{c}{\mathbf{P}2} \\ \cline{2-4} L & -1 & 1 & -2/3 \\ \cline{2-4} M & 2 & 0 & 1 \\ \cline{2-4} R & 0 & 1 & -1/2 \\ \cline{2-4} \multicolumn{1}{c}{} & \multicolumn{1}{c}{L} & \multicolumn{1}{c}{M} & \multicolumn{1}{c}{R} \end{array} \ \mathbf{P}1 \qquad (1.10)$$

This bi-matrix game clearly admits a unique NE in pure strategies, which is $\{M, M\}$, with the corresponding outcome being $(1, 0)$. Let us now stipulate that the roles of the players are not symmetric and **P**1 (Player 1) can enforce her strategy on **P**2 (Player 2).[14] Then, before she announces her strategy, Player 1 has to take into account possible responses of Player 2 (the follower), and in view of this, she has to decide on the strategy that is most favorable to her. For the decision problem whose possible cost pairs are given as entries of A and B, above, let us now work out the reasoning that Player 1 (the leader) will have to go through. If Player 1 chooses L, then Player 2 has a unique response (that minimizes his cost) which is L, thereby yielding a cost of 0 to Player 1. If the leader chooses M, Player 2's response is again unique (which is M), with the corresponding cost incurred to Player 1 being 1. Finally, if she picks R, Player 2's unique response is also R, and the cost to Player 1 is 2. Since the lowest of these costs is the first one, it readily follows that L is the most reasonable choice for the leader (**P**1) in this hierarchical decision problem. We then say that L is the Stackelberg strategy of the leader (**P**1) in this game, and the pair $\{L, L\}$ is the Stackelberg solution with Player 1 as the leader. Furthermore, the cost pair $(0, -1)$ is the Stackelberg (equilibrium) outcome of the game with Player 1 as the leader. It should be noted that this outcome is actually more favorable for both players than the unique Nash outcome – this latter feature, however, is not a rule in such games. If, for example, Player 2 is the leader in the bi-matrix game (1.10), then the unique Stackelberg solution is $\{L, R\}$ with the corresponding outcome being $(3/2, -2/3)$ which is clearly not favorable for Player 1 (the follower) when compared with her unique NE cost. For Player 2 (the leader), however, the Stackelberg outcome is again better than his NE outcome.

The Stackelberg equilibrium (SE) solution concept introduced above within the context of the bi-matrix game (1.10) is applicable to all two-person finite games in normal form, provided that they exhibit one feature which was inherent to the bi-matrix game (1.10) and was used implicitly in the derivation: *the follower's response to every strategy of the leader should be unique*. If this requirement is not satisfied, then there is ambiguity in the possible responses of the follower and thereby in the possible attainable cost levels of the leader. As an explicit example to demonstrate such a decision situation, consider the bi-matrix game

$$A = \overset{\mathbf{P}2}{\begin{array}{c|c|c|c|} & L & M & R \\ \hline L & 0 & 1 & 3 \\ \hline R & 2 & 2 & -1 \\ \hline \end{array}}\ \mathbf{P}1, \qquad B = \overset{\mathbf{P}2}{\begin{array}{c|c|c|c|} & L & M & R \\ \hline L & 0 & 0 & 1 \\ \hline R & -1 & 0 & -1 \\ \hline \end{array}}\ \mathbf{P}1, \tag{1.11}$$

and with **P**1 (Player 1) acting as the leader. Here, if **P**1 chooses (and announces) L, **P**2 has two optimal responses L and M, whereas if **P**1 picks R, **P**2 again has two optimal responses, L and R. Since this multiplicity of optimal responses for

[14] In this asymmetric decision-making setting, we will refer to Player 1 as "she" and Player 2 as "he."

the follower results in a multiplicity of cost levels for the leader for each one of her strategies, the Stackelberg solution concept introduced earlier cannot directly be applied here. However, this ambiguity in the attainable cost levels of the leader can be resolved if we stipulate that the leader's attitude is toward securing her possible losses against the choices of the follower within the class of his optimal responses, rather than toward taking risks. Then, under such a mode of play, **P**1's secured cost level corresponding to her strategy L would be 1, and the one corresponding to R would be 2. Hence, we declare $\gamma^{1*} = L$ as the unique Stackelberg strategy of **P**1 in the bi-matrix game of (1.11), when she acts as the leader.[15] The corresponding Stackelberg cost for **P**1 (the leader) is $J^{1*} = 1$. It should be noted that, in the actual play of the game, **P**1 could actually end up with a lower cost level, depending on whether the follower chooses his optimal response $\gamma^2 = L$ or the optimal response $\gamma^2 = M$. Consequently, the outcome of the game could be either $(1, 0)$ or $(0, 0)$, and hence we cannot talk about a unique Stackelberg equilibrium outcome of the bi-matrix game (1.11) with **P**1 acting as the leader. Before concluding the discussion on this example, we note that the admissible NE outcome of the bi-matrix game (1.11) is $(-1, -1)$ which is more favorable for both players than the possible Stackelberg outcomes given above.

We now provide a precise definition for the Stackelberg solution concept introduced above within the context of two bi-matrix games, so as to encompass all two-person finite games of the single-act and multi-act type which do not incorporate any chance moves. For such a game, let Γ^1 and Γ^2 again denote the pure-strategy spaces of Player 1 and Player 2, respectively, and $J^i(\gamma^1, \gamma^2)$ denote the cost incurred to Player i corresponding to a strategy pair $\{\gamma^1 \in \Gamma^1, \gamma^2 \in \Gamma^2\}$. Then, we have

Definition 6. In a two-person finite game, the set $R^2(\gamma^1) \subset \Gamma^2$ defined for each $\gamma^1 \in \Gamma^1$ by

$$R^2(\gamma^1) = \left\{\xi \in \Gamma^2 : J^2(\gamma^1, \xi) \le J^2(\gamma^1, \gamma^2), \quad \forall \gamma^2 \in \Gamma^2\right\} \tag{1.12}$$

is the *optimal response (rational reaction) set* of Player 2 to the strategy $\gamma^1 \in \Gamma^1$ of Player 1. ◇

Definition 7. In a two-person finite game with Player 1 as the leader, a strategy $\gamma^1 \in \Gamma^1$ is called a *Stackelberg equilibrium strategy* for the leader, if

$$\max_{\gamma^2 \in R^2(\gamma^{1*})} J^1(\gamma^{1*}, \gamma^2) = \min_{\gamma^1 \in \Gamma^1} \max_{\gamma^2 \in R^2(\gamma^1)} J^1(\gamma^1, \gamma^2) \triangleq J^{1*}. \tag{1.13}$$

[15] Of course, the "strategy" here could also be viewed as an "action" if what we have is a static game, but since we are dealing with normal forms here (which could have an underlying extensive form description) we will use the term "strategy" throughout, to be denoted by γ^i for **P**i, and the cost to **P**i will be denoted by J^i.

The quantity J^{1*} is the *Stackelberg cost* of the leader. If, instead, Player 2 is the leader, the same definition applies with only the superscripts 1 and 2 interchanged. $\diamond$

Theorem 3. *Every two-person finite game admits a Stackelberg strategy for the leader.*

Remark 4. The result above follows directly from (1.13), since the strategy spaces Γ^1 and Γ^2 are finite and $R^2(\gamma^1)$ is a subset of Γ^2 for each $\gamma^1 \in \Gamma^1$. Note that the Stackelberg strategy for the leader does not necessarily have to be unique, but nonuniqueness of the equilibrium strategy does not create any problem here (as it did in the case of NE), since the Stackelberg cost for the leader is unique. $\diamond$

Remark 5. If $R^2(\gamma^1)$ is a singleton for each $\gamma^1 \in \Gamma^1$, then there exists a mapping $T^2 : \Gamma^1 \to \Gamma^2$ such that $\gamma^2 \in R^2(\gamma^1)$ implies $\gamma^2 = T^2\gamma^1$. This corresponds to the case when the optimal response of the follower (which is T^2) is unique for every strategy of the leader, and it leads to the following simplified version of (1.13) in Definition 7:

$$J^1(\gamma^{1*}, T^2\gamma^{1*}) = \min_{\gamma^1 \in \Gamma^1} J^1(\gamma^1, T^2\gamma^1) \triangleq J^{1*}. \tag{1.14}$$

Here J^{1*} is no longer only a secured equilibrium cost level for the leader (**P**1), but it is the cost level that is actually attained. $\diamond$

From the follower's point of view, the equilibrium strategy in a Stackelberg game is any optimal response to the announced Stackelberg strategy of the leader. More precisely,

Definition 8. Let $\gamma^{1*} \in \Gamma^1$ be a Stackelberg strategy for the leader (**P**1). Then, any element $\gamma^{2*} \in R^2(\gamma^{1*})$ is an *optimal strategy* for the follower (**P**2) that is *in equilibrium* with γ^{1*}. The pair $\{\gamma^{1*}, \gamma^{2*}\}$ is a *Stackelberg solution* for the game with Player 1 as the leader and the cost pair $(J^1(\gamma^{1*}, \gamma^{2*}), J^2(\gamma^{1*}, \gamma^{2*}))$ is the corresponding *Stackelberg equilibrium outcome*. $\diamond$

Remark 6. In the preceding definition, the cost level $J^1(\gamma^{1*}, \gamma^{2*})$ could in fact be lower than the Stackelberg cost J^{1*} – a feature that has already been observed within the context of the bi-matrix game (1.11). However, if $R^2(\gamma^{1*})$ is a singleton, then these two cost levels have to coincide. $\diamond$

For a given two-person finite game, let J^{1*} again denote the Stackelberg cost of the leader (**P**1) and J^1_N denote any Nash equilibrium cost for the same player. We have already seen within the context of the bi-matrix game (1.11) that J^{1*} is not necessarily lower than J^1_N, in particular, when the optimal response of the follower is not unique. The following proposition now provides one sufficient condition under which the leader never does worse in a "Stackelberg game" than in a "Nash game"; for a proof, see Başar and Olsder (1999).

Proposition 8. *For a given two-person finite game, let J^{1*} and J^1_N be as defined before. If $R^2(\gamma^1)$ is a singleton for each $\gamma^1 \in \Gamma^1$, then*

$$J^{1*} \leq J^1_N .$$

Remark 7. One might be tempted to think that if a nonzero-sum game admits a unique Nash equilibrium solution and a unique Stackelberg strategy (γ^{1*}) for the leader, and further if $R^2(\gamma^{1*})$ is a singleton, then the inequality of Proposition 8 still should hold. This, however, is not true as the following bi-matrix game demonstrates

$$A = \begin{array}{c} \quad\ \mathbf{P}2 \\ \begin{array}{c|c|c|} \cline{2-3} L & 0 & 1 \\ \cline{2-3} R & -1 & 2 \\ \cline{2-3} \end{array} \\ \quad\ L \quad R \end{array} \ \mathbf{P}1, \qquad B = \begin{array}{c} \quad\ \mathbf{P}2 \\ \begin{array}{c|c|c|} \cline{2-3} L & 0 & 2 \\ \cline{2-3} R & 1 & 1 \\ \cline{2-3} \end{array} \\ \quad\ L \quad R \end{array} \ \mathbf{P}1$$

Here, there exists a unique Nash equilibrium solution, as indicated, and a unique Stackelberg strategy $\gamma^{1*} = L$ for the leader (**P**1). Furthermore, the follower's optimal response to $\gamma^{1*} = L$ is unique (which is $\gamma^2 = L$). However, $0 = J^{1*} > J^1_N = -1$. This counterexample indicates that the sufficient condition of Proposition 8 cannot be relaxed any further in any satisfactory way. ◇

6.2 Stackelberg Equilibria in Mixed and Behavioral Strategies

The motivation behind introducing mixed strategies in the investigation of saddle-point equilibria and Nash equilibria was that such equilibria do not always exist in pure strategies, whereas within the enlarged class of mixed strategies, one can ensure existence of noncooperative equilibria. In the case of the Stackelberg solution of two-person finite games, however, an equilibrium always exists (cf. Theorem 3), and thus, at the outset, there seems to be no need to introduce mixed strategies. Besides, since the leader dictates her strategy on the follower, in a Stackelberg game, it might at first seem to be unreasonable to imagine that the leader would ever employ a mixed strategy. Such an argument, however, is not always valid, and there are cases when the leader can actually do better (in the average sense) with a proper mixed strategy than the best she can do within the class of pure strategies. As an illustration of such a possibility, consider the bi-matrix game (A, B) displayed below:

$$A = \begin{array}{c} \textbf{P2} \\ \begin{array}{c|c|c|} & & \\ L & 1 & 0 \\ R & 0 & 1 \\ & L & R \end{array} \end{array} \textbf{P1}, \qquad B = \begin{array}{c} \textbf{P2} \\ \begin{array}{c|c|c|} & & \\ L & 1/2 & 1 \\ R & 1 & 1/2 \\ & L & R \end{array} \end{array} \textbf{P1}. \tag{1.15}$$

If **P1** acts as the leader, then the game admits two pure-strategy Stackelberg equilibrium solutions, which are $\{L, L\}$ and $\{R, R\}$, the Stackelberg outcome in each case being $(1, 1/2)$. However, if the leader (**P1**) adopts the mixed strategy which is to pick L and R with equal probability $1/2$, then the average cost incurred to **P1** will be equal to $1/2$, quite independent of the follower's (pure or mixed) strategy. This value $\bar{J}^1 = 1/2$ is clearly lower than the leader's Stackelberg cost in pure strategies, which can further be shown to be the unique Stackelberg cost of the leader in mixed strategies, since any deviation from $(1/2, 1/2)$ for the leader results in higher values for $\bar{J}^1$, by taking into account the optimal responses of the follower.

The preceding result then establishes the significance of mixed strategies in the investigation of Stackelberg equilibria of two-person nonzero-sum games and demonstrates the possibility that a proper mixed-strategy Stackelberg solution could lead to a lower cost level for the leader than the Stackelberg cost level in pure strategies. To introduce the concept of mixed-strategy Stackelberg equilibrium in mathematical terms, we take the two-person nonzero-sum finite game to be in normal form (without any loss of generality) and associate with it a bi-matrix game (A, B). Abiding by the earlier notation and terminology, let Y and Z denote the mixed-strategy spaces of Player 1 and Player 2, respectively, with their typical elements denoted by y and z. Then, we have:

Definition 9. For a bi-matrix game (A, B), the set

$$\bar{R}^2(y) = \{z^\circ \in Z : y'Bz^\circ \leq y'Bz, \forall z \in Z\} \tag{1.16}$$

is the *optimal response* (*rational reaction*) *set* of Player 2 *in mixed strategies* to the mixed strategy $y \in Y$ of Player 1. ◇

Definition 10. In a bi-matrix game (A, B) with Player 1 acting as the leader, a mixed strategy $y^* \in Y$ is called a *mixed Stackelberg equilibrium strategy* for the leader if

$$\max_{z \in \bar{R}^2(y^*)} y^{*\prime} A z = \inf_{y \in Y} \max_{z \in \bar{R}^2(y)} y' A z \triangleq \bar{J}^{1*}. \tag{1.17}$$

The quantity $\bar{J}^{1*}$ is the *Stackelberg cost* of the leader in mixed strategies. ◇

It should be noted that the "maximum" in (1.17) always exists since, for each $y \in Y$, $y'Az$ is continuous in z, and $R^2(y)$ is a closed and bounded subset of Z (which is a finite dimensional simplex). Hence, $\bar{J}^{1*}$ is a well-defined quantity.

The "infimum" in (1.17), however, cannot always be replaced by a "minimum," unless the problem admits a mixed Stackelberg equilibrium strategy for the leader. The following example, taken from Başar and Olsder (1999), now demonstrates the possibility that a two-person finite game might not admit a mixed-strategy Stackelberg strategy even though $\bar{J}^{1*} < J^{1*}$.

Example. Consider the following modified version of the bi-matrix game of (1.15):

$$A = \begin{array}{c|c|c|}\multicolumn{3}{c}{\mathbf{P}2}\\ \hline L & 1 & 0 \\ \hline R & 0 & 1 \\ \hline \multicolumn{1}{c}{} & \multicolumn{1}{c}{L} & \multicolumn{1}{c}{R}\end{array}\ \mathbf{P}1, \qquad B = \begin{array}{c|c|c|}\multicolumn{3}{c}{\mathbf{P}2}\\ \hline L & 1/2 & 1 \\ \hline R & 1 & 1/3 \\ \hline \multicolumn{1}{c}{} & \multicolumn{1}{c}{L} & \multicolumn{1}{c}{R}\end{array}\ \mathbf{P}1\,.$$

With **P**1 as the leader, this bi-matrix game also admits two pure-strategy Stackelberg equilibria, which are $\{L, L\}$ and $\{R, R\}$, the Stackelberg cost for the leader being $J^{1*} = 1$. Now, let the leader adopt the mixed strategy $y = (y_1, (1 - y_1))'$, under which $\bar{J}^2$ is

$$\bar{J}^2(y, z) = y'Bz = \left(-\frac{7}{6}y_1 + \frac{2}{3}\right) z_1 + \frac{2}{3}y_1 + \frac{1}{3},$$

where $z = (z_1, (1-z_1))'$ denotes any mixed strategy of **P**2. Then, the mixed-strategy optimal response set of **P**2 can readily be determined as

$$\bar{R}^2(y) = \begin{cases} \{z = (1,0)\} & \text{if } y_1 > 4/7 \\ \{z = (0,1)\} & \text{if } y_1 < 4/7 \\ Z & \text{if } y_1 = 4/7. \end{cases}$$

Hence, for $y_1 > 4/7$, the follower chooses "column 1" with probability 1, and this leads to an average cost of $\bar{J}^1 = y_1$ for **P**1. For $y_1 < 4/7$, on the other hand, **P**2 chooses "column 2" with probability 1, which leads to an average cost level of $\bar{J}^1 = (1-y_1)$ for **P**1. Then, clearly, the leader will prefer to stay in this latter region; in fact, if he employs the mixed strategy $y = (4/7 - \epsilon, 3/7 + \epsilon)'$ where $\epsilon > 0$ is sufficiently small, his realized average cost will be $\bar{J}^1 = 3/7 + \epsilon$, since then **P**2 will respond with the unique pure-strategy $\gamma^2 = R$. Since $\epsilon > 0$ can be taken as small as possible, we arrive at the conclusion that $\bar{J}^{1*} = \frac{3}{7} < 1 = J^{1*}$. In spite of this fact, the leader does not have a mixed Stackelberg strategy since for the only candidate $y^\circ = (4/7, 3/7)$, $\bar{R}^2(y^\circ) = Z$, and therefore $\max_{z \in \bar{R}^2(y^\circ)} y^{\circ\prime} Az = 4/7$, which is higher than $\bar{J}^{1*}$. ◇

The preceding example thus substantiates the possibility that a mixed Stackelberg strategy might not exist for the leader, but she can still do better than her pure Stackelberg cost J^{1*} by employing some suboptimal mixed strategy (such as the one $y = (4/7 - \epsilon, 3/7 + \epsilon)'$ in the example, for sufficiently small $\epsilon > 0$). In fact, whenever $\bar{J}^{1*} < J^{1*}$, there will always exist such an approximating mixed strategy for the leader. If $\bar{J}^{1*} = J^{1*}$, however, it is, of course, reasonable to employ the pure

Stackelberg strategy which always exists by Theorem 3. The following proposition now verifies that $\bar{J}^{1*} < J^{1*}$ and $\bar{J}^{1*} = J^{1*}$ are the only two possible relations we can have between $\bar{J}^{1*}$ and J^{1*}; in other words, the inequality $\bar{J}^{1*} > J^{1*}$ never holds (Başar and Olsder 1999, p. 142).

Proposition 9. *For every two-person finite game, we have*

$$\bar{J}^{1*} \leq J^{1*}. \tag{1.18}$$

Computation of a mixed-strategy Stackelberg equilibrium (whenever it exists) is not as straightforward as in the case of pure-strategy equilibria, since the spaces Y and Z are not finite. The standard technique is first to determine the minimizing solution(s) of

$$\min_{z \in Z} y' B z$$

as functions of $y \in Y$. This will lead to a decomposition of Y into subsets (regions), on each of which a reaction set for the follower is defined. (Note that in the analysis of the previous example, Y has been decomposed into three regions.) Then, one has to minimize $y' A z$ over $y \in Y$, subject to the constraints imposed by these reaction sets, and under the stipulation that the same quantity is maximized on these reaction sets whenever they are not singletons. This brute-force approach also provides approximating strategies for the leader, whenever a mixed Stackelberg solution does not exist, together with the value of $\bar{J}^{1*}$.

If the two-person finite game under consideration is a dynamic game in extensive form, then it is more reasonable to restrict attention to behavioral strategies. Stackelberg equilibrium within the class of behavioral strategies can be introduced as in Definitions 9 and 10, by replacing the mixed strategy sets with the behavioral strategy sets. Hence, using the earlier terminology and notation, we have the following counterparts of Definitions 9 and 10, in behavioral strategies:

Definition 11. Given a two-person finite dynamic game with behavioral-strategy sets $(\hat{\Gamma}^1, \hat{\Gamma}^2)$ and average cost functions $(\hat{J}^1, \hat{J}^2)$, the set

$$\widehat{R}^2(\hat{\gamma}^1) = \left\{\hat{\gamma}^{2\circ} \in \hat{\Gamma}^2 : \hat{J}^2(\hat{\gamma}^1, \hat{\gamma}^{2\circ}) \leq \hat{J}^2(\hat{\gamma}^1, \hat{\gamma}^2), \forall \hat{\gamma}^2 \in \hat{\Gamma}^2\right\}, \tag{1.19}$$

is the *optimal response (rational reaction) set* of **P**2 in *behavioral strategies* to the behavioral strategy $\hat{\gamma}^1 \in \hat{\Gamma}^1$ of **P**1. ◇

Definition 12. In a two-person finite dynamic game with Player 1 acting as the leader, a behavioral strategy $\hat{\gamma}^{1*} \in \hat{\Gamma}^1$ is called a *behavioral Stackelberg equilibrium* strategy for the leader if

$$\sup_{\hat{\gamma}^2 \in \bar{R}^2(\hat{\gamma}^{1*})} \hat{J}^1(\hat{\gamma}^{1*}, \hat{\gamma}^2) = \inf_{\hat{\gamma}^1 \in \hat{\Gamma}^1} \sup_{\hat{\gamma}^2 \in \widehat{R}^2(\hat{\gamma}^1)} \hat{J}^1\left(\hat{\gamma}^1, \hat{\gamma}^2\right) \triangleq \hat{J}^{1*}. \tag{1.20}$$

The quantity $\hat{J}^{1*}$ is the *Stackelberg cost* of the leader in *behavioral strategies*. ◇

7 Nash Equilibria of Infinite/Continuous-Kernel Games

7.1 Formulation, Existence, and Uniqueness

We now go back to the general class of N-player games introduced through (1.1), with X_i being a finite-dimensional space (e.g., m_i-dimensional Euclidean space, $\mathbb{R}^{m_i}$), for $i \in \mathcal{N}$; L_i a continuous function on the product space X, which of course is also finite-dimensional (e.g., if $X_i = \mathbb{R}^{m_i}$, X can be viewed as $\mathbb{R}^m$, where $m := \sum_{i \in \mathcal{N}} m_i$); and the constraint set Ω a subset of X. This class of games is known as *continuous-kernel games with coupled constraints*, and of course if the constraints are not coupled, for example, with each player having a separate constraint set $\Omega_i \subset X_i$, this would also be covered as a special case. Now, further assume that Ω is closed, bounded, and convex, and for each $i \in \mathcal{N}$, $L_i(x_i, x_{-i})$ is convex in $x_i \in X_i$ for every $x_{-i} \in \times_{j \neq i} X_j$. Then, the basic result for such games is that they admit Nash equilibria in pure strategies (but the equilibria need not be unique), as stated in the theorem below, due to *Rosen* (1965); see also Başar and Olsder (1999, pp. 176–177).

Theorem 4. *For the N-player nonzero-sum continuous-kernel game formulated above, with the constraint set Ω a closed, bounded, and convex subset of $\mathbb{R}^m$, and with $L_i(x_i, x_{-i})$ convex in x_i for each x_{-i} and each $i \in \mathcal{N}$, there exists a Nash equilibrium in pure strategies.*

Remark 8. The proof of the result above uses Kakutani's fixed-point theorem.[16] If the constraint sets are decoupled, and $L_i(x_i, x_{-i})$ is strictly convex in $x_i \in \Omega_i$, then there is an alternative proof for Theorem 4, which uses Brouwer's fixed-point theorem.[17] Under the given hypotheses, it follows from Weirstrass theorem and strict convexity that the minimization problem

$$\min_{x_i \in \Omega_i} L_i(x_i, x_{-i})$$

[16]This fixed point theorem says that if S is a compact subset of $\mathbb{R}^n$, and f is an upper semicontinuous function which assigns to each $x \in S$ a closed and convex subset of S, then there exists $x \in S$ such that $x \in f(x)$.

[17] Brouwer's theorem says that a continuous mapping, f, of a closed, bounded, convex subset, S, of a finite-dimensional space into itself has a fixed point.

admits a unique solution for each x_{-i}, this being so for each $i \in \mathcal{N}$, that is, there exists a unique map $T_i : \Omega_i \to \Omega_{-i}$,[18] such that the solution to the minimization problem is

$$x_i = T_i(x_{-i}), \quad i \in \mathcal{N} \tag{1.21}$$

Furthermore, T_i is continuous on Ω_{-i}. Clearly, every pure-strategy NE has to provide a solution to (1.21), and vice versa. Stacking these maps, there exists a corresponding continuous map $T : \Omega \to \Omega$, whose components are the T_i's, and (1.21) is equivalent to $x = T(x)$, which is a fixed-point equation. Since T is a continuous mapping of Ω into itself, and Ω is a closed and bounded subset of a finite-dimensional space (and thus compact), by Brouwer's fixed-point theorem, T has a fixed point, and hence an NE exists. ◇

For the special class of 2-person ZSGs structured the same way as the NZSG of Theorem 4, a similar result clearly holds (as a special case), implying the existence of a SPE (in pure strategies). Note that in this case, the single objective function ($L \equiv L_1 \equiv -L_2$) to be minimized by Player 1 and maximized by Player 2 is convex in x_1 and concave in x_2, in view of which such zero-sum games are known as *convex-concave games*. Even though convex-concave games could admit multiple saddle-point solutions, they are ordered interchangeable, and the values of the games are unique (which is not the case for multiple Nash equilibria in genuine NZSGs, as we have also seen earlier). Now, if the convexity-concavity is replaced by strict convexity-concavity (for ZSGs), then the result can be sharpened as below,[19] which however has no a counterpart for Nash equilibria in genuine NZSGs.

Theorem 5. *For a two-person zero-sum game on closed, bounded, and convex finite-dimensional action sets $\Omega_1 \times \Omega_2$, defined by the continuous kernel $L(x_1, x_2)$, let $L(x_1, x_2)$ be strictly convex in x_1 for each $x_2 \in \Omega_2$ and strictly concave in x_2 for each $x_1 \in \Omega_1$. Then, the game admits a unique pure-strategy SPE.*

If the structural assumptions of Theorem 4 do not hold, then a pure-strategy Nash equilibrium may not exist, but there may exist one in mixed strategies. Mixed strategy (MS) for a player (say, Player i) is a probability distribution on that player's action set, which we take to be a closed and bounded subset, Ω_i, of $X_i = \mathbb{R}^{m_i}$, and denote an MS of Player i by p_i, and the set of all probability distributions on Ω_i by $\mathcal{P}_i$. NE, then, is defined by the N-tuple of inequalities (1.4), using the expected values of L_i's given mixed strategies of all the players, which we denote by J_i as

[18] T_i is known as the reaction function (or response function) of Player i to other players' actions.

[19] Here, existence of SPE is a direct consequence of Theorem 4. By strict convexity and strict concavity, there can be no SPE outside the class of pure strategies, and uniqueness follows from the ordered interchangeability property of multiple SPs, in view of strict convexity/concavity (Başar and Olsder 1999).

before. The following theorem, whose proof can be found in Owen (1974), now states the basic result on existence of MSNE in continuous-kernel games.[20]

Theorem 6. *For the N-player continuous-kernel NZSG formulated above, with the constrained action set Ω_i for Player i a closed and bounded subset of $\mathbb{R}^{m_i}$, and with $L_i(x_i, x_{-i})$ continuous on $\Omega = \Omega_1 \times \cdots \times \Omega_N$, for each $i \in \mathcal{N}$, there exists an MSNE, $(p_1^*, \ldots, p_N^*)$, satisfying (1.4).*

Remark 9. As in the case of finite (matrix) games, the existence of a pure-strategy NE does not preclude the existence of also a genuine MSNE,[21] and all such (multiple) NE are generally noninterchangeable, unless the game is a ZSG or is strategically equivalent to one. ◇

As a special case of Theorem 6, we now have:

Corollary 1. *Every continuous-kernel 2-player ZSG with compact action spaces has an MSSPE.* ◇

7.2 Stability and Computation

We have seen in the previous subsection that when the cost functions of the players are strictly convex in a continuous-kernel NZSG, then the NE is completely characterized by the solution of a fixed-point equation, namely, (1.21). Since solutions of fixed-point equations can be obtained recursively (under some condition), this brings up the possibility of computing the NE recursively, using the iteration

$$x_i(k+1) = T_i(x_{-i}(k)), \quad k = 0, 1, \ldots, \quad i \in \mathcal{N}, \tag{1.22}$$

where k stands for times of updates by the players. Note that this admits an *online computation* interpretation for the underlying game, where each player needs to know only the most recent actions of the other players (and not their cost functions) and her own reaction function T_i (for which only the individual cost function of the player is needed). Hence, this recursion entails a distributed computation with little information on the parameters of the game. Lumping all players' actions together and writing (1.22) as

[20]The underlying idea of the proof is to make the kernels L_i discrete so as to obtain an N-person matrix game that suitably approximates the original game in the sense that an MSNE of the latter (which exists by Nash's theorem) is arbitrarily close to a mixed equilibrium solution of the former. Compactness of the action spaces ensures that a limit to the sequence of solutions obtained for approximating finite matrix games exists.

[21]The qualifier *genuine* is used here to stress the point that mixed strategies in this statement are not pure strategies (even though pure strategies are indeed special types of mixed strategies, with all probability weight concentrated on one point).

$$x(k+1) = T(x(k)), \quad k = 0, 1, \ldots,$$

we note that the sequence generated converges for all possible initial choices, $x(0) = x_0$, if T is a contraction from X into itself.[22] As an immediate by-product, we also have *uniqueness* of the NE.

The recursion above is not the only way one can generate a sequence converging to its fixed point. But before discussing other possibilities, it is worth to make a digression and introduce a classification of NE based on such recursions, provided by the notion of "stability" of the solution(s) of the fixed-point equation. This discussion will then immediately lead to other possible recursions (for $N > 2$). For the sake of simplicity in the initial discussion, let us consider the two-player case (because in this case there is only one type of recursion for the fixed-point equation, as will be clear later). Given an NE (and assuming that the players are at the NE point), consider the following sequence of moves: (i) One of the players (say Player 1) deviates from his corresponding equilibrium strategy, (ii) Player 2 observes this and minimizes her cost function in view of the new strategy of Player 1, (iii) Player 1 now optimally reacts to that (by minimizing his cost function), (iv) Player 2 optimally reacts to that optimum reaction, etc. Now, if this infinite sequence of moves converges back to the original NE solution, and this being so regardless of the nature of the initial deviation of Player 1, we say that the NE is *stable*. If convergence is valid only under small initial deviations, then we say that the NE is *locally stable*. Otherwise, the NE is said to be *unstable*. An NZSG can of course admit more than one locally stable equilibrium solution, but a stable NE solution has to be unique.

The notion of *stability*, as introduced above for two-person games, brings in a refinement to the concept of NE, which finds natural extensions to the N-player case. Essentially, we have to require that the equilibrium be "restorable" under any rational readjustment scheme when there is a deviation from it by any player. For $N > 2$, this will depend on the specific scheme adopted, which brings us to the following formal definition of a stable Nash equilibrium.

Definition 13. An NE $x_i^*, i \in \mathcal{N}$, is (*globally*) *stable* with respect to an adjustment scheme $\mathcal{S}$ if it can be obtained as the limit of the iteration:

$$x_i^* = \lim_{k \to \infty} x_i^{(k)}, \tag{1.23}$$

$$x_i^{(k+1)} = \arg\min_{x_i \in \Omega_i} L_i \left(x_{-i}^{(\mathcal{S}_k)}, x_i \right), \quad x_i^{(0)} \in \Omega_i, \quad i \in \mathcal{N}, \tag{1.24}$$

[22]This follows from Banach's contraction mapping theorem. If T maps a normed space X into itself, it is a contraction if there exists $\alpha \in [0,1)$ such that $\|T(x) - T(y)\| \leq \alpha \|x - y\|$, $\forall x, y \in X$.

where the superscript $\mathcal{S}_k$ indicates that the precise choice of $x_{-i}^{(\mathcal{S}_k)}$ depends on the readjustment scheme selected. ◇

One possibility for the scheme above is $x_{-i}^{(\mathcal{S}_k)} = x_{-i}^{(k)}$, which corresponds to the situation where the players update (readjust) their actions simultaneously, in response to the most recently determined actions of the other players. Yet another possibility is

$$x_{-i}^{(\mathcal{S}_k)} = \left(x_1{}^{(k+1)}, \ldots, x_{i-1}{}^{(k+1)}, x_{i+1}{}^{(k)}, \ldots, x_N{}^{(k)}\right),$$

where the players update in an predetermined (in this case numerical) order. A third possibility is

$$x_{-i}^{(\mathcal{S}_k)} = \left(x_1{}^{m^i_{1,k}}, \ldots, x_{i-1}{}^{(m^i_{i-1,k})}, x_{i+1}{}^{(m^i_{i+1,k})}, \ldots, x_N{}^{(m^i_{N,k})}\right),$$

where $m^i_{j,k}$ is an integer-valued random variable, satisfying the bounds:

$$\max(0, k-d) \le m^i_{j,k} \le k+1, \quad j \ne i, j \in \mathcal{N}, i \in \mathcal{N},$$

which corresponds to a situation where Player i receives action update information from Player j at random times, with the delay not exceeding d time units.

Clearly, if the iteration of Definition 13 converges under any one of the readjustment schemes above (or any other readjustment scheme where a player receives update information from every other player infinitely often), then the NE is *unique*. Every unique NE, however, is not necessarily *stable*, nor is an NE that is stable with respect to a particular readjustment scheme is necessarily stable with respect to some other scheme. Hence, *stability* is generally given with some qualification (such as "stable with respect to scheme $\mathcal{S}$" or "with respect to a given class of schemes"), except when $N = 2$, in which case all schemes (with at most a finite delay in the transmission of update information) lead to the same condition of stability, as one then has the simplified recursions

$$x_i{}^{(r_{k+1,i})} = \tilde{T}_i\left(x_i{}^{(r_{k,i})}\right), \quad k = 0, 1, \ldots; \quad i = 1, 2,$$

where $r_{1,i}, r_{2,i}, r_{3,i}, \ldots$ denote the time instants when Player i receives new action update information from Player j, $j \ne i$, $i, j = 1, 2$.

8 Continuous-Kernel Games and Stackelberg Equilibria

This section is on the Stackelberg solution of static NZSGs when the number of alternatives available to each player is not a finite set and the cost functions are described by continuous kernels. For the sake of simplicity and clarity in exposition, the focus will be on two-person static games. A variety of possible extensions of the Stackelberg solution concept to N-person static games with different levels of hierarchy can be found in Başar and Olsder (1999).

Here, the notation will be slightly different from the one in the previous section, with $u^i \in U^i$ denoting the action variable of Player i (instead of $x_i \in \Omega_i$), where her action set U^i is assumed to be a subset of an appropriate metric space (such as X_i). The cost function J^i of Player i is defined as a continuous function on the product space $U^1 \times U^2$. Then, we can state the following general definition of a Stackelberg equilibrium solution (SES), which is the counterpart of Definition 7 for infinite games.

Definition 14. In a two-person game, with Player 1 as the leader, a strategy $u^{1^*} \in U^1$ is called a *Stackelberg equilibrium strategy* for the leader if

$$J^{1^*} \triangleq \sup_{u^2 \in R^2(u^{1^*})} J^1\left(u^{1^*}, u^2\right) \leq \sup_{u^2 \in R^2(u^1)} J^1\left(u^1, u^2\right), \tag{1.25}$$

for all $u^1 \in U^1$. Here, $R^2(u^1)$ is the rational reaction set of the follower as introduced in (1.12). ◇

Remark 10. If $R^2(u^1)$ is a singleton for each $u^1 \in U^1$, in other words, if it is described completely by a reaction curve $T_2 : U^1 \to U^2$, then inequality (1.25) in the above definition can be replaced by

$$J^{1^*} \triangleq J^1\left(u^{1^*}, T_2(u^{1^*})\right) \leq J^1\left(u^1, T_2(u^1)\right), \tag{1.26}$$

for all $u^1 \in U^1$. ◇

If a SES exists for the leader, then the LHS of inequality (1.25) is known as the *leader's Stackelberg cost* and is denoted by J^{1^*}. A more general definition for J^{1^*} is, in fact,

$$J^{1^*} = \inf_{u^1 \in U^1} \sup_{u^2 \in R^2(u^1)} J^1\left(u^1, u^2\right), \tag{1.27}$$

which also covers the case when a Stackelberg equilibrium strategy does not exist. It follows from this definition that the Stackelberg cost of the leader is a well-defined quantity, and that there will always exist a sequence of strategies for the leader which will insure her a cost value arbitrarily close to J^{1^*}. This observation brings us to the following definition of ϵ Stackelberg strategies.

Definition 15. Let $\epsilon > 0$ be a given number. Then, a strategy $u_\epsilon^{1^*} \in U^1$ is called an ϵ *Stackelberg strategy* for the leader (**P**1) if

$$\sup_{u^2 \in R^2(u_\epsilon^{1^*})} J^1\left(u_\epsilon^{1^*}, u^2\right) \leq J^{1^*} + \epsilon.$$

◇

The next two properties of ϵ Stackelberg strategies now readily follow.

Property 1. In a two-person game, let J^{1^*} be a finite number. Then, given an arbitrary $\epsilon > 0$, an ϵ Stackelberg strategy necessarily exists. ◇

Property 2. Let $\{u_{\epsilon_i}^{1^*}\}$ be a given sequence of ϵ Stackelberg strategies in U^1, with $\epsilon_i > \epsilon_j$ for $i < j$ and $\lim_{j\to\infty} \epsilon_j = 0$. Then, if there exists a convergent subsequence $\{u_{\epsilon_{ik}}^{1^*}\}$ in U^1 with limit u^{1^*}, and further if $\sup_{u^2 \in R^2(u^1)} J^1(u^1, u^2)$ is a continuous function of u^1 in an open neighborhood of $u^{1^*} \in U^1$, u^{1^*} is a Stackelberg strategy for Player 1. ◇

The equilibrium strategy of the follower, in a Stackelberg game, would be any strategy that constitutes an optimal response to the one adopted (and announced) by the leader. Mathematically speaking, if u^{1^*} (respectively, $u_\epsilon^{1^*}$) is adopted by the leader, then any $u^2 \in R^2(u^1)$ (respectively, $u^2 \in R^2(u_\epsilon^{1^*})$ will be referred to as an *optimal strategy* for the follower, in *equilibrium* with the Stackelberg (respectively, ϵ Stackelberg) strategy of the leader. This pair is referred to as a *Stackelberg* (respectively, ϵ Stackelberg) solution of the two-person game with Player 1 as the leader (see Definition 8). The following theorem now provides a set of sufficient conditions for two-person NZSGs to admit a SES; see Başar and Olsder (1999).

Theorem 7. *Let U^1 and U^2 be compact metric spaces and J^i be continuous on $U^1 \times U^2$, $i = 1, 2$. Further, let there exist a finite family of continuous mappings $l^{(i)} : U^1 \to U^2$, indexed by a parameter $i \in I \triangleq \{1, \ldots, M\}$, so that $R^2(u^1) = \{u^2 \in U^2 : u^2 = l^{(i)}(u^1), i \in I\}$. Then, the two-person nonzero-sum static game admits a Stackelberg equilibrium solution.*

Remark 11. The assumption of Theorem 7 concerning the structure of $R^2(\cdot)$ imposes some severe restrictions on J^2; but such an assumption is inevitable as the following example demonstrates. Take $U^1 = U^2 = [0, 1]$, $J^1 = -u^1u^2$, and $J^2 = (u^1 - \frac{1}{2})u^2$. Here, $R^2(\cdot)$ is determined by a mapping $l(\cdot)$ which is continuous on the half-open intervals $[0, \frac{1}{2})$, $(\frac{1}{2}, 1]$, but is multivalued at $u^1 = \frac{1}{2}$. The Stackelberg cost of the leader is clearly $J^{1^*} = -\frac{1}{2}$, but a Stackelberg strategy does not exist because of the "infinitely multivalued" nature of l. ◇

When $R^2(u^1)$ is a singleton for every $u^1 \in U^1$, the hypothesis of Theorem 7 can definitely be made less restrictive. One such set of conditions is provided in the following corollary to Theorem 7 under which there exists a unique l which is continuous.

Corollary 2. *Let U^1 and U^2 be compact metric spaces and J^i be continuous on $U^1 \times U^2$, $i = 1, 2$. Further let $J^2(u^1, \cdot)$ be strictly convex for all $u^1 \in U^1$ and Player 1 act as the leader. Then, the game admits a SES.* ◇

It should be noted that the SES for a two-person game exists under a set of sufficiency conditions which are much weaker than those required for existence of Nash equilibria. It should further be noted, however, that the statement of Theorem 7 does not also rule out the existence of a mixed-strategy Stackelberg solution which might provide the leader with a lower average cost. We have already observed occurrence of such a phenomenon within the context of matrix games earlier in the chapter, and we now investigate to what extent such a result could remain valid in continuous-kernel games.

If mixed strategies are also allowed, then permissible strategies for Player i will be probability measures μ^i on the space U^i. Let us denote the collection of all such probability measures for Player i by M^i. Then, the quantity replacing J^i will be the average cost function

$$\bar{J}^i(\mu^1, \mu^2) = \int_{U^1} \int_{U^2} J^i(u^1, u^2)\, \mathrm{d}\mu^1(u^1)\, \mathrm{d}\mu^2(u^2), \tag{1.28}$$

and the reaction set R^2 will be replaced by

$$\bar{R}^2(\mu^1) \triangleq \left\{ \widehat{\mu}^2 \in M^2 : \bar{J}^2(\mu^1, \widehat{\mu}^2) \le \bar{J}^2(\mu^1, \mu^2), \forall \mu^2 \in M^2 \right\}. \tag{1.29}$$

Hence, we have:

Definition 16. In a two-person game with Player 1 as the leader, a mixed strategy $\mu^{1^*} \in M^1$ is called a *mixed Stackelberg equilibrium strategy* for the leader if

$$\bar{J}^{1^*} \triangleq \sup_{\mu^2 \in \bar{R}^2(\mu^{1^*})} \bar{J}^1(\mu^{1^*}, \mu^2) \le \sup_{\mu^2 \in \bar{R}^2(\mu^1)} \bar{J}^1(\mu^1, \mu^2)$$

for all $\mu^1 \in M^1$, where $\bar{J}^{1^*}$ is known as the *average Stackelberg cost* of the leader in *mixed strategies*. ◇

The following result says that under mixed strategies, the average Stackelberg cost for the leader cannot be higher than under pure strategies Başar and Olsder (1999).

Proposition 10.

$$\bar{J}^{1^*} \le J^{1^*} \tag{1.30}$$

We now show, by a counterexample, that, even under the hypothesis of Theorem 7, it is possible to have strict inequality in (1.30).

(Counter-) Example. Consider a two-person continuous-kernel game with $U^1 = U^2 = [0, 1]$, and with cost functions

$$J^1 = \epsilon(u^1)^2 + u^1\sqrt{u^2} - u^2; \quad J^2 = (u^2 - (u^1)^2)^2,$$

where $\epsilon > 0$ is a sufficiently small parameter. The unique Stackelberg solution of this game, in pure strategies, is $u^{1^*} = 0$, $u^{2*} = (u^1)^2$, and the Stackelberg cost for the leader is $J^{1^*} = 0$. We now show that the leader can actually do better by employing a mixed strategy.

First note that the follower's unique reaction to a mixed strategy of the leader is $u^2 = E[(u^1)^2]$ which, when substituted into $\bar{J}^1$, yields the expression

$$\bar{J}^1 = \epsilon E[(u^1)^2] + E[u^1]\sqrt{\{E[(u^1)^2]\}} - E[(u^1)^2].$$

Now, if the leader uses the uniform probability distribution on [0, 1], his average cost becomes

$$\bar{J}^1 = \frac{\epsilon - 1}{3} + \frac{1}{2}\sqrt{\frac{1}{3}},$$

which clearly indicates that, for ϵ sufficiently small, $\bar{J}^{1^*} < 0 = J^{1^*}$. ◇

The preceding example has demonstrated the fact that even in Stackelberg games with *strictly convex* cost functionals, there may exist mixed-strategy SE in addition to pure-strategy one(s), and in fact the former could lead to a better performance for the leader than any of the latter.[23] However, if we further restrict the cost structure to be quadratic, then only pure-strategy Stackelberg equilibria will exist (Başar and Olsder 1999, pp. 183–184).

Proposition 11. *Consider the two-person nonzero-sum game with* $U^1 = \mathbb{R}^{m_1}$, $U^2 = \mathbb{R}^{m_2}$, *and*

$$J^i = \frac{1}{2}u^{i\prime}R^i_{ii}u^i + u^{1\prime}R^i_{ij}u^j + \frac{1}{2}u^{j}.{}'R^i_{jj}u^j + u^{i\prime}r^i_i + u^{j\prime}r^i_j;\ i, j = 1, 2, i \neq j,$$

where $R^i_{ii} > 0$, R^i_{ii}, R^i_{ij}, R^i_{jj} *are appropriate dimensional matrices and* r^i_i, r^i_j *are appropriate dimensional vectors. This "quadratic" game can only admit a pure-strategy Stackelberg solution, with either Player* 1 *or Player* 2 *as the leader.*

[23] In retrospect, this should not be surprising since for the special case of ZSGs (without pure-strategy saddle points), we have already seen that the minimizer could further decrease her guaranteed expected cost by playing a mixed strategy; here, however, it holds even if $J^1 \not\equiv -J^2$.

9 Quadratic Games: Deterministic and Stochastic

This section presents explicit expressions for the Nash, saddle-point, and Stackelberg equilibrium solutions of static nonzero-sum games in which the cost functions of the players are quadratic in the decision variables – the so-called *quadratic games*. The action (strategy) spaces will be taken as appropriate dimensional Euclidean spaces, but the results are also equally valid (under the right interpretation) when the strategy spaces are taken as infinite-dimensional Hilbert spaces. In that case, the Euclidean inner products will have to be replaced by the inner product of the underlying Hilbert space, and the positive definiteness requirements on some of the matrices will have to be replaced by *strong positive definiteness* of the corresponding self-adjoint operators. The section also includes some discussion on iterative algorithms for the computation of Nash equilibria in the quadratic case, as well as some results on Nash equilibria of static stochastic games with quadratic costs where the players have access to the state of nature through (possibly independent) noisy channels.

9.1 Deterministic Games

A general quadratic cost function for Player i, which is strictly convex in her action variable, can be written as

$$J^i = \frac{1}{2}\sum_{j=1}^{N}\sum_{k=1}^{N} u^{j\prime} R^i_{jk} u^k + \sum_{j=1}^{N} r^{i\prime}_j u^j + c^i, \tag{1.31}$$

where $u^j \in U^j = \mathbb{R}^{m_j}$ is the m_j-dimensional action variable of Player j, R^i_{jk} is an $(m_j \times m_k)$-dimensional matrix with $R^i_{ii} > 0$, r^i_j is an m_j-dimensional vector, and c^i is a constant. Without loss of generality, we may assume that, for $j \neq k$, $R^i_{jk} = R^{i\,\prime}_{kj}$, since if this were not the case, the corresponding two quadratic terms could be written as

$$u^{j\prime} R^i_{jk} u^k + u^{k\prime} R^i_{kj} u^j = u^{j\prime}\left(\frac{R^i_{jk}+R^{i\,\prime}_{kj}}{2}\right) u^k + u^{k\prime}\left(\frac{R^i_{jk}+R^{i\,\prime}_{kj}}{2}\right) u^j, \tag{1.32}$$

and redefining R^i_{jk} as $(R^i_{jk} + R^{i\,\prime}_{jk})/2$, a symmetric matrix could be obtained. By an analogous argument, we may take R^i_{jj} to be symmetric, without any loss of generality.

Quadratic cost functions are of particular interest in game theory, first because they constitute second-order approximation to other types of nonlinear cost functions, and second because games with quadratic cost or payoff functions are analytically tractable, admitting, in general, closed-form equilibrium solutions which provide insight into the properties and features of the equilibrium solution concept under consideration.

To determine the NE solution in strictly convex quadratic games, we differentiate J^i with respect to u^i $(i \in \mathcal{N})$, set the resulting expressions equal to zero, and solve the set of equations thus obtained. This set of equations, which also provides a sufficient condition because of strict convexity, is

$$R^i_{ii}u^i + \sum_{j \neq i} R^i_{ij}u^j + r^i_i = 0 \quad (i \in \mathcal{N}), \tag{1.33}$$

which can be written in compact form as

$$Ru = -r \tag{1.34}$$

where

$$R \triangleq \begin{bmatrix} R^1_{11} & R^1_{12} & \cdots & R^1_{1N} \\ R^2_{12} & R^2_{22} & \cdots & R^2_{2N} \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ R^N_{1N} & R^N_{2N} & \cdots & R^N_{NN} \end{bmatrix} \tag{1.35}$$

$$u' \triangleq (u^1, u^2, \ldots, u^N), \quad r' \triangleq (r^1_1, r^2_2, \ldots, r^N_N). \tag{1.36}$$

This then leads to the following result.

Proposition 12. *The quadratic N-player nonzero-sum static game defined by the cost functions (1.31) and with $R^i_{ii} > 0$ admits a Nash equilibrium (NE) solution if, and only if, (1.34) admits a solution, say u^*; this NE solution is unique if the matrix R defined by (1.35) is invertible, in which case it is given by*

$$u^* = -R^{-1}r. \tag{1.37}$$

◇

Remark 12. Since each player's cost function is strictly convex and continuous in his action variable, quadratic nonzero-sum games of the type discussed above cannot admit a NE solution in mixed strategies. Hence, in strictly convex quadratic games, the equilibrium analysis can be confined to the class of pure strategies. ◇

We now investigate the stability properties of the unique NE solution of quadratic games, where the notion of stability was introduced earlier. Taking $N = 2$, and directly specializing recursion (1.22) to the quadratic case (with the obvious change in notation, and in a sequential update mode), we arrive at the following iteration:

$$u^{1(k+1)} = C_1 u^{2(k)} + d_1, \quad u^{2(k+1)} = C_2 u^{1(k+1)} + d_2, \quad k = 0, 1, \ldots, \tag{1.38}$$

with an arbitrary starting choice $u^{2(0)}$, where

$$C_i = -(R^i_{ii})^{-1} R^i_{ij}, \quad d_i = -(R^i_{ii})^{-1} r^i_i, \quad j \neq i, \ i, j = 1, 2.$$

This iteration corresponds to the sequential (Gauss-Seidel) update scheme where Player 1 responds to the most recent past action of Player 2, whereas Player 2 responds to the current action of Player 1. The alternative to this is the parallel (Jacobi) update scheme where (1.38) is replaced by[24]

$$u^{1(k+1)} = C_1 u^{2(k)} + d_1, \quad u^{2(k+1)} = C_2 u^{1(k)} + d_2, \quad k = 0, 1, \ldots, \tag{1.39}$$

starting with arbitrary initial choices $(u^{1(0)}, u^{2(0)})$. Then, the question of stability of the NE solution (1.37), with $N = 2$, reduces to the question of stability of the fixed point of either (1.38) or (1.39). Note that, apart from a relabeling of indices, stability of these two iterations is equivalent to the stability of the single iteration:

$$u^{1(k+1)} = C_1 C_2 u^{1(k)} + C_1 d_2 + d_1.$$

Since this is a linear difference equation, a necessary and sufficient condition for it to converge (to the actual NE strategy of Player 1) is that the eigenvalues of the matrix $C_1 C_2$ or equivalently those of $C_2 C_1$ should be in the unit circle, *i.e.*

$$\rho(C_1 C_2) \equiv \rho(C_2 C_1) < 1 \tag{1.40}$$

where $\rho(A)$ is the spectral radius of the matrix A.

Note that the condition of stability is considerably more stringent than the condition of existence of a unique Nash equilibrium, which is

$$\det(I - C_1 C_2) \neq 0. \tag{1.41}$$

The question we address now is whether, in the framework of Gauss-Seidel or Jacobi iterations, this gap between (1.40) and could be (1.41) shrunk or even totally eliminated, by allowing players to incorporate memory into the iterations. While doing this, it would be desirable for the players to need to know as little as possible regarding the reaction functions of each other (note that no such information is necessary in the Gauss-Seidel or Jacobi iterations given above).

[24] This one corresponds to (1.22).

To study this issue, consider the Gauss-Seidel iteration (1.38), but with a one-step memory for (only) Player 1. Then, the "relaxed" algorithm will be (using the simpler notation $u^{1^{(k)}} = u_k$, $u^{2^{(k)}} = v_k$):

$$\left.\begin{aligned} u_{k+1} &= C_1 v_k + d_1 + A(u_k - C_1 v_k - d_1) \\ v_{k+1} &= C_2 u_{k+1} + d_2 \end{aligned}\right\}, \tag{1.42}$$

where A is a gain matrix, yet to be chosen. Substituting the second (for v_k) into the first, we obtain the single iteration

$$u_{k+1} = [C + A(I - C)]u_k + (I - A)[d_1 + C_1 d_2],$$

where

$$C \triangleq C_1 C_2.$$

By choosing

$$A = -C(I - C)^{-1}, \tag{1.43}$$

where the required inverse exists because of (1.41), we obtain a finite-step convergence, assuming that the true value of C_2 is known to Player 1. If the true value of C_2 is not known, but a nominal value is given in a neighborhood of which the true value lies, the scheme (1.42) along with the choice (1.43) for the nominal value still leads to convergence (but not in a finite number of steps) provided that the neighborhood is sufficiently small (Başar 1987).

Now, if the original scheme is instead the parallel (Jacobi) scheme, then a one-step memory for Player 1 will not be sufficient to obtain a finite-step convergence result as above. In this case, we replace (1.42) by

$$\left.\begin{aligned} u_{k+1} &= C_1 v_k + d_1 + B(u_{k-1} - C_1 v_k - d_2) \\ v_{k+1} &= C_2 u_k + d_2 \end{aligned}\right\}, \tag{1.44}$$

where B is another gain matrix. Note that here Player 1 uses, in the computation of u_{k+1}, not u_k but rather u_{k-1}. Now, substituting for v_k from the second into the first equation of (1.44), we arrive at the iteration

$$u_{k+1} = [C + B(I - C)]u_{k-1} + (I - B)[d_1 + C_1 d_2],$$

which again shows finite-step convergence, with B chosen as

$$B = -C(I - C)^{-1}. \tag{1.45}$$

Again, there is a certain neighborhood of nominal C_2 or equivalently of the nominal C, where the iteration (1.44) is convergent.

In general, however, the precise scheme according to which Player 2 responds to Player 1's policy choices may not be common information, and hence one would like to develop relaxation-type algorithms for Player 1 which would converge to the true equilibrium solution regardless of what particular scheme Player 2 adopts (e.g., Gauss-Seidel or Jacobi). Consider, for example, the scheme where Player 2's responses for different k are modeled by

$$v_{k+1} = C_2 u_{k+1-i_k} + d_2, \tag{1.46}$$

where $i_k \geq 0$ is an integer denoting the delay in the receipt of current policy information by Player 2 from Player 1. The choice $i_k = 0$ for all k would correspond to the Gauss-Seidel iteration, and the choice $i_k = 1$ for all k to the Jacobi iteration – assuming that u_{k+1} is still determined according to (1.38). An extreme case would be the totally asynchronous communication where $\{i_k\}_{k \geq 0}$ could be any sequence of positive integers. Under the assumptions that Player 1 communicates new policy choices to Player 2 *infinitely often* and he uses the simple ("nonrelaxed") iteration

$$u_{k+1} = C_1 v_k + d_1, \tag{1.47}$$

it is known from the work of *Chazan and Miranker* (1969) that such a scheme converges if, and only if,

$$\rho(|C|) < 1, \tag{1.48}$$

where $|C|$ is the matrix derived from C by multiplying all its negative entries by -1.

This condition can be improved upon, however, by incorporating relaxation terms in (1.47), such as

$$u_{k+1} = \alpha u_k + (1-\alpha) C_1 v_k + (1-\alpha) d_1, \tag{1.49}$$

where α is some scalar. The condition for convergence of any asynchronously implemented version of (1.46) and (1.49) in this case is

$$\rho(\bar{A}(\alpha)) < 1, \tag{1.50}$$

where

$$\bar{A}(\alpha) := \begin{pmatrix} |\alpha| I & |(1-\alpha) C_1| \\ |C_2| & 0 \end{pmatrix}. \tag{1.51}$$

Clearly, there is a value of $\alpha \neq 0$ for which (1.50) requires a less stringent condition (on C_1 and C_2) than (1.48). For example, if C_1 and C_2 are scalars, and $\alpha = \frac{1}{2}$, (1.50) dictates

$$C_1 C_2 < 4,$$

while (1.48) requires that $C_1 C_2 < 1$.

From a game theoretic point of view, each of the iteration schemes discussed above corresponds to a game with a sufficiently large number of stages and with a particular mode of play among the players. Moreover, the objective of each player is to minimize a kind of an average long horizon cost, with costs at each stage contributing to this average cost. Viewing this problem overall as a multi-act NZSG, we observe that the behavior of each player at each stage of the game is rather "myopic," since at each stage the players minimize their cost functions only under past information, and quite in ignorance of the possibility of any future moves. If the possibility of future moves is also taken into account, then the rational behavior of each player at a particular stage could be quite different. Such myopic decision-making could make sense, however, if the players have absolutely no idea as to how many stages the game comprises, in which case there is the possibility that at any stage a particular player could be the last one to act in the game. In such a situation, risk-averse players would definitely adopt "myopic" behavior, minimizing their current cost functions under only the past information, whenever given the opportunity to act.

9.1.1 Two-Person Zero-Sum Games

Since ZSGs are special types of two-person NZSGs with $J_1 = -J_2$ (Player 1 minimizing and Player 2 maximizing), in which case the NE solution concept coincides with the concept of SPE, a special version of Proposition 12 will be valid for quadratic zero-sum games. To this end, we first note that the relation $J_1 = -J_2$ imposes in (1.31) the restrictions

$$R_{12}^{1\,\prime} = -R_{21}^2, R_{11}^2 = -R_{11}^1, R_{22}^1 = -R_{22}^2, r_1^2 = -r_1^1, r_2^1 = -r_2^2, c_1 = -c_2,$$

under which matrix R defined by (1.35) can be written as

$$R = \begin{pmatrix} R_{11}^1 & R_{12}^1 \\ -R_{12}^{1\,\prime} & R_{22}^2 \end{pmatrix}$$

which has to be nonsingular for existence of a saddle point. Since R can also be written as the sum of two matrices

$$R = \begin{pmatrix} R_{11}^1 & 0 \\ 0 & R_{22}^2 \end{pmatrix} + \begin{pmatrix} 0 & R_{12}^1 \\ -R_{12}^{1\,\prime} & 0 \end{pmatrix}$$

the first one being positive definite and the second one skew symmetric, and since eigenvalues of the latter are always imaginary, it readily follows that R is a nonsingular matrix. Hence, we arrive at the conclusion that quadratic strictly convex-concave zero-sum games admit unique saddle-point equilibrium in pure strategies.

Corollary 3. *The strictly convex-concave quadratic zero-sum game*

$$J = \frac{1}{2}u^{1'}R^1_{11}u^1 + u^{1'}R^1_{12}u^2 - \frac{1}{2}u^{2'}R^2_{22}u^2 + u^{1'}r^1_1 + u^{2'}r^1_2 + c^1;$$
$$R^1_{11} > 0, R^2_{22} > 0,$$

admits a unique saddle-point equilibrium in pure strategies, which is given by

$$u^{1^*} = -[R^1_{11} + R^1_{12}(R^2_{22})^{-1}R^{1\,\prime}_{12}]^{-1}[r^1_1 + R^1_{12}(R^2_{22})^{-1}r^1_2],$$
$$u^{2*} = [R^2_{22} + R^{1\,\prime}_{12}(R^1_{11})^{-1}R^1_{12}]^{-1}[r^1_2 + R^{1\,\prime}_{12}(R^1_{11})^{-1}r^1_1].$$

◇

Remark 13. The positive-definiteness requirements on R^1_{11} and R^2_{22} in Corollary 3 are necessary and sufficient for the game kernel to be strictly convex-strictly concave, but this structure is clearly not necessary for the game to admit a saddle point. If the game is simply convex-concave (i.e., if the matrices above are nonnegative definite, with a possibility of zero eigenvalues), then an SPE will still exist provided that the upper and lower values are bounded.[25] If the quadratic game is not convex-concave, however, then either the upper or the lower value (or both) will be unbounded, implying that a saddle point will not exist. ◇

9.1.2 Team Problems

Yet another special class of NZSGs are the team problems in which the players (or equivalently, members of the team) share a common objective. Within the general framework, this corresponds to the case $J_1 \equiv J^2 \equiv \cdots \equiv J^N \triangleq J$, and the objective is to minimize this cost function over all $u^i \in U^i$, $i = 1, \ldots, N$. The resulting solution N-tuple $(u^{1^*}, u^{2^*}, \ldots, u^{N^*})$ is known as the *team-optimal solution*. The NE solution, however, corresponds to a weaker solution concept in team problems (as we have already seen), the so-called *person-by-person (pbp) optimality*. In a two-member team problem, for example, a pbp optimal solution (u^{1^*}, u^{2*}) dictates satisfaction of the pair of inequalities

$$J(u^{1^*}, u^{2^*}) \le J(u^1, u^{2^*}), \quad \forall u^1 \in U^1,$$
$$J(u^{1^*}, u^{2^*}) \le J(u^{1^*}, u^2), \quad \forall u^2 \in U^2,$$

whereas a team-optimal solution (u^{1^*}, u^{2^*}) requires satisfaction of a single inequality

[25] For a convex-concave quadratic game, the upper value will not be bounded if, and only if, there exists a $v \in \mathbb{R}^{m_2}$ such that $v'R^2_{22}v = 0$ while $v'r^1_2 \neq 0$. A similar result also applies to the lower value.

$$J(u^{1^*}, u^{2^*}) \leq J(u^1, u^2), \quad \forall u^1 \in U^1, u^2 \in U^2.$$

A team-optimal solution always implies pbp optimality, but not vice versa. Of course, if J is quadratic and strictly convex on the product space $U^1 \times \cdots \times U^N$, then a unique pbp optimal solution exists, and it is also team-optimal.[26] However, for a cost function that is strictly convex only on individual spaces U^i, but not on the product space, this latter property may not be true. Consider, for example, the quadratic cost function

$$J = (u^1)^2 + (u^2)^2 + 10u^1u^2 + 2u^1 + 3u^2,$$

which is strictly convex in u^1 and u^2, separately. The matrix corresponding to R defined by (1.35) is

$$\begin{pmatrix} 2 & 10 \\ 10 & 2 \end{pmatrix},$$

which is clearly nonsingular. Hence, a unique pbp optimal solution will exist. However, a team-optimal solution does not exist since the said matrix (which is also the Hessian of J) has one positive and one negative eigenvalue. By cooperating along the direction of the eigenvector corresponding to the negative eigenvalue, the members of the team can make the value of J as small as possible. In particular, taking $u^2 = -\frac{2}{3}u^1$ and letting $u^1 \to +\infty$ drives J to $-\infty$.

9.1.3 The Stackelberg Solution

We now elaborate on the SESs of quadratic games of type (1.31) but with $N = 2$ and Player 1 acting as the leader. We first note that since the quadratic cost function J^i is strictly convex in u^i, by Proposition 11, we can confine our investigation of an equilibrium solution to the class of pure strategies. Then, to every announced strategy u^1 of Player 1, the follower's unique response will be as given by (1.33) with $N = 2, i = 2$:

$$u^2 = -(R^2_{22})^{-1}[R^2_{21}u^1 + r^2_2]. \tag{1.52}$$

Now, to determine the Stackelberg strategy of the leader, we have to minimize J^1 over U^1 and subject to the constraint imposed by the reaction of the follower. Since the reaction curve gives u^2 uniquely in terms of u^1, this constraint can best be handled by substitution of (1.52) in J^1 and by minimization of the resulting functional (to be denoted by $\tilde{J}^1$) over U^1. To this end, we first determine $\tilde{J}^1$:

[26] This result may fail to hold true for team problems with strictly convex but nondifferentiable kernels.

$$\begin{aligned}\tilde{J}^1(u^1) = \frac{1}{2}u^{1'}R^1_{11}u^1 &+ \frac{1}{2}[R^2_{21}u^1 + r^2_2]'(R^2_{22})^{-1}R^1_{22}(R^2_{22})^{-1}[R^2_{21}u^1 + r^2_2] \\ &- u^{1'}R^1_{21}(R^2_{22})^{-1}[R^2_{21}u^1 + r^2_2] + u^{1'}r^1_1 \\ &- [R^2_{21}u^1 + r^2_2]'(R^2_{22})^{-1}r^1_2 + c^1.\end{aligned}$$

For the minimum of $\tilde{J}^1$ over U^1 to be unique, we have to impose a strict convexity condition on $\tilde{J}^1$. Because of the quadratic structure of $\tilde{J}^1$, this condition amounts to having the coefficient matrix of the quadratic term in u^1 positive definite, which is

$$\begin{aligned}R^1_{11} + R^2_{21}{}'(R^2_{22})^{-1}R^1_{22}(R^2_{22})^{-1}R^2_{21} - R^1_{21}(R^2_{22})^{-1}R^2_{21}& \\ - R^2_{21}{}'(R^2_{22})^{-1}R^1_{21}{}' > 0.&\end{aligned} \tag{1.53}$$

Under this condition, the unique minimizing solution can be obtained by setting the gradient of $\tilde{J}^1$ equal to zero, which yields

$$\begin{aligned}u^{1^*} = &-[R^1_{11} + R^2_{21}(R^2_{22})^{-1}R^1_{22}(R^2_{22})^{-1}R^2_{21} - R^1_{21}(R^2_{22})^{-1}R^2_{21} \\ &- R^2_{21}{}'(R^2_{22})^{-1}R^1_{21}{}']^{-1}[R^2_{21}{}'(R^2_{22})^{-1}R^1_{22}R^1_{22}(R^2_{22})^{-1}r^2_2 \\ &- R^1_{21}(R^2_{22})^{-1}r^2_2 + r^1_1 - R^2_{21}{}'(R^2_{22})^{-1}r^1_2].\end{aligned} \tag{1.54}$$

Proposition 13. *Under condition (1.53), the two-person version of the quadratic game (1.31) admits a unique Stackelberg strategy for the leader, which is given by (1.54). The follower's unique response is then given by (1.52).* ⋄

Remark 14. A sufficient condition for condition (1.53) is strict convexity of J^1 on the product space $U^1 \times U^2$. ⋄

9.2 Stochastic Games

In this subsection, we introduce and discuss the equilibria of stochastic static games with quadratic cost functions, for only the case $N = 2$. Stochasticity will enter the game through the cost functions of the players, as weights on the linear terms of the action variables. Accordingly, the quadratic cost functions will be given by (where we differentiate between players using subscripts instead of superscripts)

$$L_1(u_1, u_2; \xi_1, \xi_2) = \frac{1}{2}u_1'R_{11}u_1 + u_1'R_{12}u_2 + u_1'xi_1,$$

$$L_2(u_1, u_2; \xi_1, \xi_2) = \frac{1}{2}u_2'R_{22}u_2 + u_2'R_{21}u_1 + u_2'\xi_2,$$

where R_{ii} are positive definite and ξ_i's are random vectors of appropriate dimensions. Player 1 and Player 2 do not have access to the values of these random vectors, but they measure another pair of random vectors, y_1 (for Player 1) and y_2 (for Player

2), which carry some information on ξ_i's. We assume that all four random variables have bounded first and second moments, and their joint distribution is common information to both players.

Player 1 uses y_1 in the construction of her policy and subsequently action, where we denote her policy variable (strategy) by γ_1, so that $u_1 = \gamma_1(y_1)$. Likewise, we introduce γ_2 as the strategy for Player 2, so that $u_2 = \gamma_2(y_2)$. These policy variables have no restrictions imposed on them other than *measurability* and that u_i's should have bounded first and second moments. Let Γ_1 and Γ_2 be the corresponding spaces where γ_1 and γ_2 belong. Then, for each $\gamma_i \in \Gamma_i$, $i = 1, 2$, using $u_1 = \gamma_1(y_1)$ and $u_2 = \gamma_2(y_2)$ in L_1 and L_2, and taking expectation over the statistics of the four random variables, we arrive at the normal form of the game (in terms of the strategies), captured by the expected costs:

$$J_1(\gamma_1, \gamma_2) = E[L_1(\gamma_1(y_1), \gamma_2(y_2); \xi_1, \xi_2)]$$
$$J_2(\gamma_1, \gamma_2) = E[L_2(\gamma_1(y_1), \gamma_2(y_2); \xi_1, \xi_2)]$$

We are looking for a NE in $\Gamma_1 \times \Gamma_2$, where NE is defined in the usual way.

Using properties of conditional expectation, for fixed $\gamma_2 \in \Gamma_2$, there exists a unique $\gamma_1 \in \Gamma_1$ that minimizes $J_1(\gamma_1, \gamma_2)$ over Γ_1. This unique solution is given by

$$\gamma_1(y_1) = R_{11}^{-1} \left[R_{12} E \left[\gamma_2(y_2)|y_1\right] + E\xi_1|y_1\right]\right] =: T_1(\gamma_2)(y_1),$$

which is the unique response by Player 1 to a strategy of Player 2. Likewise, Player 2's response to Player 1 is unique:

$$\gamma_2(y_2) = R_{22}^{-1} \left[R_{21} E \left[\gamma_1(y_1)|y_2\right] + E\xi_2|y_2\right]\right] =: T_2(\gamma_1)(y_2).$$

Hence, in the policy space, we will be looking for a fixed point of

$$\gamma_1 = T_1(\gamma_2), \quad \gamma_2 = T_2(\gamma_1),$$

and substituting the second one into the first, we have

$$\gamma_1 = (T_1 \circ T_2)(\gamma_1),$$

where $T_1 \circ T_2$ is the composite map. This will admit a unique solution if $T_1 \circ T_2$ is a contraction (note that Γ_1 is a Banach space).

Now, writing out this fixed-point equation

$$\begin{aligned}\gamma_1(y) = {} & R_{11}^{-1} R_{12} R_{22}^{-1} R_{21} E\left[E\left[\gamma_1(y_1)|y_2\right]|y_1\right] + R_{11}^{-1} R_{12} R_{22}^{-1} E[\xi_2|y_1] - R_{11}^{-1} E[\xi_1|y_1] \\ =: {} & \tilde{T}_1(\gamma_1)(y_1) + R_{11}^{-1} R_{12} R_{22}^{-1} E[\xi_2|y_1] - R_{11}^{-1} E[\xi_1|y_1].\end{aligned}$$

Hence, $T_1 \circ T_2$ is a contraction if, and only if, $\tilde{T}_1$ is a contraction, since conditional expectation is a nonexpansive mapping, it follows that the condition for existence of NE (and its stability) is exactly the one obtained in the previous subsection for the deterministic game, that is,

$$\rho(C_1C_2) = \rho(R_{11}^{-1}R_{12}R_{22}^{-1}R_{21}) < 1.$$

In this case, the recursion

$$\gamma_1^{(k+1)} = (T_1 \circ T_2)\left(\gamma_1^{(k)}\right)$$

will converge for all $\gamma_1^{(0)} \in \Gamma_1$. Note that if this sequence converges, so does the one generated by

$$\gamma_1^{(k+1)} = T_1\left(\gamma_2^{(k)}\right), \quad \gamma_2^{(k+1)} = T_2\left(\gamma_1^{(k)}\right),$$

for all $\gamma_i^{(0)} \in \Gamma_i,\ i = 1, 2$. And *the limit is the unique NE.*

If the four random vectors are jointly Gaussian distributed, then the unique NE will be affine in y_1 (for Player 1) and y_2 (for Player 2), which follows from properties of Gaussian random variables, by taking $\gamma_i^{(0)} = 0$. Further results on this class of stochastic games for the Stackelberg equilibrium, for the Nash equilibrium with $N > 2$, and when the players do not agree on a common underlying statistics for the uncertainty can be found in Başar (1978, 1985).

10 Games with Incomplete Information

The game models and solution concepts discussed heretofore in this chapter were all built on the governing assumption that the players all have *complete information* on the elements of the game, particularly on the action spaces of all players and the players' cost functions, and that this is all common information to all players. It could, however, happen, especially in a competitive environment, that some private information available to a player may not be publicly available to other players. In particular, a player may not have complete information on other players' possible actions, strategies, and objective functions. One way of addressing such decision-making problems with incomplete information in precise mathematical terms is to formulate them as *Bayesian games* – a formulation introduced by Harsanyi in the 1960s (Harsanyi 1967). This section provides a brief introduction to such games, within the *static* framework, as extension of complete information static games introduced earlier; more details on such games, as well as Bayesian dynamic games in extensive form, and applications in economics can be found in Fudenberg and Tirole (1991). Applications of Bayesian games in wireless communications and networking have been discussed in Han et al. (2011).

Now, for static noncooperative games with complete information, we had identified three basic elements for a precise formulation: (i) the number of players, N, with players set being $\mathcal{N}$; (ii) the possible actions available to each player, and any constraints that may be imposed on them (we had denoted a generic action for Player i by x_i and her action set by X_i, for $i \in \mathcal{N}$); and (iii) the objective function of each player ($L_i(x_i, x_{-i})$ for Player i) which she attempts to optimize (specifically, minimize). In static Bayesian games (or static games with incomplete information), there is one additional element, which is the *type* of a player, which we denote by $t_i \in T_i$ for Player i, where T_i is the player's type space, assumed to be finite. Then, the objective function of a player depends not only on actions but also on types (possibly of all players), which we write (using the standard convention) as $L_i(a_i, a_{-i}; t_i, t_{-i})$ for Player i. Player i knows her own type, t_i, but has only a belief on other players' types, captured by conditional probability of t_{-i} given her own type t_i, denoted by $p_i(t_{-i}|t_i)$. Events in such a game follow the following sequence, initiated by *nature's* choice of types:

1. Nature chooses the types of all players.
2. Players observe their own types, as drawn by nature. A full description of a particular player's type is known only to that player (*private information*).
3. Players simultaneously choose their actions. In particular, Player i chooses an action based on her belief on the types of the other players, given her own type, this being so for all $i \in \mathcal{N}$.
4. Each player incurs a cost based on the actions and types of all players; that is, for Player i, we have the cost $L_i(a_i, a_{-i}; t_i, t_{-i})$. Note that this quantity is random.

To introduce the proper equilibrium concept for this game (as an extension of Nash equilibrium for complete information games), it is convenient to introduce *strategies* for the players as mappings from their type spaces to their action sets, that is, for Player i, $s_i : T_i \to X_i$, with S_i being the set of all such maps. Let us also introduce the notation $s_{-i}(t_{-i}) := \{s_j(t_j), j \in \mathcal{N}, j \neq i\}$. Then, we say that an N-tuple $\{s_i^* \in S_i,\ i \in \mathcal{N}\}$ constitutes an equilibrium (more precisely, *Bayesian Nash Equilibrium* (BNE)) if $x_i^* = s_i^*(t_i)$ minimizes the conditional expectation of $L_i(a_i, s_{-i}^*(t_i); t_i, t_{-i})$, where expectation is taken over t_{-i} conditioned on t_i. In mathematical terms, as the counterpart of (1.1), we have (for BNE):

$$\sum_{t_{-i} \in T_{-i}} L_i(x_i^*, s_{-i}^*(t_{-i}); t_i, t_{-i}) p_i(t_{-i}|t_i) \leq \sum_{t_{-i} \in T_{-i}} L_i(x_i, s_{-i}^*(t_{-i}); t_i, t_{-i}) p_i(t_{-i}|t_i), \tag{1.55}$$

holding for all $x_i \in X_i,\ i \in \mathcal{N}$.

The BNE, as defined above, can be qualified as a pure-strategy BNE and, as in the case of NE in finite games discussed earlier, such an equilibrium may not exist. Then, we have to extend the definition to encompass mixed strategies, defined as a probability distribution for each player on her action set for each of her types (i.e., a different probability distribution for each type). With such an extension, again the counterpart of Nash's theorem holds, that is, every finite incomplete information game formulated as in this section has a BNE in mixed strategies (Fudenberg and Tirole 1991).

11 Conclusions

This introductory chapter of the *Handbook of Dynamic Game Theory* has provided an exposition to the fundamentals of game theory, by focusing on finite games and static continuous-kernel games, as a prelude to the rest of the *Handbook* that has an in-depth coverage of dynamic and differential games. The focus in this chapter has also been on noncooperative solution concepts, such as saddle point, Nash, correlated, and Stackelberg, as opposed to cooperative ones. The chapter has also provided a historical account of the development and evolution of the discipline of game theory.

References

Alpcan T, Başar T (2011) Network security: a decision and game theoretic approach. Cambridge University Press, Cambridge

Aumann RJ (1974) Subjectivity and correlation in randomized strategy. J Math Econ 1(1):67–96

Aumann RJ (1987) Correlated equilibrium as an expression of Bayesian rationality. Econometrica 55(1):1–18

Başar T (1974) A counter example in linear-quadratic games: existence of non-linear Nash solutions. J Optim Theory Appl 14(4):425–430

Başar T (1976) On the uniqueness of the Nash solution in linear-quadratic differential games. Int J Game Theory 5:65–90

Başar T (1977) Informationally nonunique equilibrium solutions in differential games. SIAM J Control 15(4):636–660

Başar T (1978) Decentralized multicriteria optimization of linear stochastic systems. IEEE Trans Autom Control AC-23(2):233–243

Başar T (1985) An equilibrium theory for multi-person decision making with multiple probabilistic models. IEEE Trans Autom Control AC-30(2):118–132

Başar T (1987) Relaxation techniques and on-line asynchronous algorithms for computation of noncooperative equilibria. J Econ Dyn Control 71:531–549

Başar T, Bernhard P (1995) H^∞ Optimal control and related minimax design problems: a dynamic game approach. Birkhäuser, Boston

Başar T, Olsder GJ (1999) Dynamic noncooperative game theory. Classics in applied mathematics, 2nd edn. SIAM, Philadelphia

Bertsekas DP (1999) Nonlinear programming, 2nd edn. Athena Scientific, Boston

Bertsekas DP (2007) Dynamic programming and optimal control. Athena Scientific, Boston

Brown GW (1951) Iterative solutions of games by fictitious play. In: Koopmans TC (ed) Activity analysis of production and allocation. Wiley, New York, pp 374–376

Chazan D, Miranker W (1969) Chaotic relaxation. Linear Algebra Appl 2:199–222

Fudenberg D, Tirole J (1991) Game theory. MIT Press, Cambridge

Han Z, Niyato D, Saad W, Başar T, Hjorungnes A (2011) Game theory in wireless and communication networks: theory, models, and applications. Cambridge University Press, Cambridge/New York

Harsanyi J (1967) Games with incomplete information played by "Bayesian" players, Parts I–III. Manage Sci 14(3):159–182, Nov 1967; 14(5):320–334, Jan 1968; 14(7):486–502, Mar 1968
Ho Y-C (1965) Review of 'Differential Games' by R. Isaacs. IEEE Trans Autom Control AC-10(4):501–503
Isaacs R (1975) Differential games, 2nd edn. Kruger, New York; First edition: Wiley, New York (1965)
Kohlberg E, Mertens JF (1986) On the strategic stability of equilibria. Econometrica 54:1003–1037
Kreps M, Wilson R (1982) Sequential equilibria. Econometrics 50:863–894
Miyasawa K (1961) On the convergence of learning processes in a 2×2 non-zero-sum two person game. Technical report 33, Economic Research Program, Princeton University, Princeton
Monderer D, Shapley LS (1996) Fictitious play property for games with identical interests. J Econ Theory 68:258–265
Myerson RB, Selten R (1978) Refinement of the Nash equilibrium concept. Int J Game Theory 7:73–80
Nash JF Jr (1950) Equilibrium points in n-person games. Proc Natl Acad Sci 36(1):48–49
Nash JF Jr (1951) Non-cooperative games. Ann Math 54(2):286–295
Nguyen KC, Alpcan T, Başar T (2010a) Security games with decision and observation errors. In: Proceedings 2010 American control conference, Baltimore, pp 510–515
Nguyen KC, Alpcan T, Başar T (2010b) Fictitious play with time-invariant frequency update for network security. In: Proceedings of the IEEE multi-conference on systems and control (MSC) and IEEE CCA, Yokohama, pp 8–10
Owen G (1974) Existence of equilibrium pairs in continuous games. Int J Game Theory 5:97–105
Owen G (1995) Game theory, 3rd edn. Academic Press, San Diego
Pontryagin LS, Boltyanskii VG, Gamkrelidze RV, Mishchenko EF (1962) The mathematical theory of optimal processes. Interscience Publishers, New York
Robinson J (1951) An iterative method for solving a game. Ann Math 54(2):296–301
Rosen JB (1965) Existence and uniqueness of equilibrium points for concave N-person games. Econometrica 33(3):520–534
Saad W, Han Z, Debbah M, Hjorungnes A, Başar T (2009) Coalitional game theory for communication networks [A tutorial]. IEEE Signal Process Mag Spec Issue Game Theory 26(5):77–97
Selten R (1975) Reexamination of the perfectness concept for equilibrium points in extensive games. Int J Game Theory 4:25–55
Shamma JS, Arslan G (2004) Unified convergence proofs of continuous-time fictitious play. IEEE Trans Autom Control 49(7):1137–1142
Shamma JS, Arslan G (2005) Dynamic fictitious play, dynamic gradient play, and distributed convergence to Nash equilibria. IEEE Trans Autom Control 50(3):312–327
Shapley LS (1964) Some topics in two-person games. In: Shapley LS, Dresher M, Tucker AW (eds) Advances in game theory. Princeton University Press, Princeton, pp 1–29
Simaan M, Cruz JB Jr (1973) On the Stackelberg strategy in nonzero sum games. J Optim Theory Appl 11:533–555
Smith JM (1974) The theory of games and the evolution of animal conflicts. J Theor Biol 47: 209–221
Smith JM (1982) Evolution and the theory of games. Cambridge University Press, Cambridge/New York
Smith JM, Price GR (1973) The logic of animal conflict. Nature 246:15–18
Starr AW, Ho YC (1969) Nonzero-sum differential games. J Optim Theory Appl 3:184–206
van Damme E (1984) A relationship between perfect equilibria in extensive games and proper equilibria in normal form. Int J Game Theory 13:1–13
van Damme E (1987) Stability and perfection of equilibria. Springer, New York
von Neumann J (1928) Zur theorie der Gesellschaftspiele. Mathematische Annalen 100:295–320

von Neumann J, Morgenstern O (1947) Theory of games and economic behavior, 2nd edn. Princeton University Press, Princeton; first edition: 1944

von Stackelberg H (1934) Marktform und gleichgewicht. Springer Verlag, Vienna (An English translation appeared in 1952 entitled "The Theory of the Market Economy," published by Oxford University Press, Oxford)

Vorob'ev NH (1977) Game theory. Springer, Berlin

Nonzero-Sum Differential Games

2

Tamer Başar, Alain Haurie, and Georges Zaccour

Contents

T. Başar (✉)
Coordinated Science Laboratory and Department of Electrical and Computer Engineering,
University of Illinois at Urbana-Champaign, Urbana, IL, USA
e-mail: basar1@illinois.edu

A. Haurie
ORDECSYS and University of Geneva, Geneva, Switzerland

GERAD-HEC Montréal PQ, Montreal, QC, Canada
e-mail: ahaurie@gmail.com

G. Zaccour
Department of Decision Sciences, GERAD, HEC Montréal, Montreal, QC, Canada
e-mail: georges.zaccour@gerad.ca

T. Başar, G. Zaccour (eds.), *Handbook of Dynamic Game Theory*,
https://doi.org/10.1007/978-3-319-44374-4_5

Abstract
This chapter provides an overview of the theory of nonzero-sum differential games, describing the general framework for their formulation, the importance of information structures, and noncooperative solution concepts. Several special structures of such games are identified, which lead to closed-form solutions.

Keywords
Closed-loop information structure · Information structures · Linear-quadratic games · Nash equilibrium · Noncooperative differential games · Non-Markovian equilibrium · Open-loop information structure · State-feedback information structure · Stackelberg equilibrium

1 Introduction

Differential games are games played by agents, also called *players*, who jointly control (through their actions over time, as inputs) a dynamical system described by differential state equations. Hence, the game evolves over a continuous-time horizon (with the length of the horizon known to all players, as common knowledge), and over this horizon each player is interested in optimizing a particular objective function (generally different for different players) which depends on the state variable describing the evolution of the game, on the self-player's action variable, and also possibly on other players' action variables. The objective function for each player could be a reward (or payoff, or utility) function, in which case the player is a maximizer, or it could be a cost (or loss) function, in which case the player would be a minimizer. In this chapter we adopt the former, and this clearly brings in no loss of generality, since optimizing the negative of a reward function would make the corresponding player a minimizer. The players determine their actions in a way to optimize their objective functions, by also utilizing the information they acquire on the state and other players' actions as the game evolves, that is, their actions are generated as a result of the *control policies* they design as mappings from their information sets to their action sets. If there are only two players and their objective functions add up to zero, then this captures the scenario of two totally conflicting objectives – what one player wants to minimize the other one wants to maximize. Such differential games are known as *zero-sum differential games*. Otherwise, a differential game is known to be *nonzero-sum*.

The study of differential games (more precisely, zero-sum differential games) was initiated by Rufus Isaacs at the Rand Corporation through a series of memoranda in the 1950s and early 1960s of the last century. His book Isaacs (1965), published in 1965, after a long delay due to classification of the material it covered, is still considered as the starting point of the field. The early books following Isaacs, such as those by Blaquière et al. (1969), Friedman (1971), and Krassovski and Subbotin (1977), all dealt (in most part) with two-player zero-sum differential games. Indeed, initially the focal point of differential games research stayed within the zero-sum domain and was driven by military applications and the presence of antagonistic elements. The topic of two-player zero-sum differential games is covered in some detail in *this chapter (TPZSDG)* of this *Handbook*.

Motivated and driven by applications in management science, operations research, engineering, and economics (see, e.g., Sethi and Thompson 1981), the theory of differential games was then extended to the case of many players controlling a dynamical system while playing a nonzero-sum game. It soon became clear that nonzero-sum differential games present a much richer set of features than zero-sum differential games, particularly with regard to the interplay between information structures and nature of equilibria. Perhaps the very first paper on this topic, by Case, appeared in 1969, followed closely by a two-part paper by Starr and Ho (1969a,b). This was followed by the publication of a number of books on the topic, by Leitmann (1974), and by Başar and Olsder (1999), with the first edition dating back to 1982, Mehlmann (1988), and Dockner et al. (2000), which focuses on applications of differential games in economics and management science. Other selected key book references are the ones by Engwerda (2005), which is specialized to linear-quadratic differential (as well as multistage) games, Jørgensen and Zaccour (2004), which deals with applications of differential games in marketing, and Yeung and Petrosjan (2005), which focuses on cooperative differential games.

This chapter is on noncooperative nonzero-sum differential games, presenting the basics of the theory, illustrated by examples. It is based in most part on material in Chaps. 6 and 7 of Başar and Olsder (1999) and Chap. 7 of Haurie et al. (2012).

2 A General Framework for *m*-Player Differential Games

2.1 A System Controlled by *m* Players

2.1.1 System Dynamics

Consider an n-dimensional dynamical system controlled by a set of m players over a time interval $[t^0, T]$, where $T > t^0$ is a final time that can either be a given data or defined endogenously as the time of reaching a given target, as to be detailed below. For future use, let $M = \{1, \dots, m\}$ denote the players set, that is, the set of all players. This dynamical system has the following elements:

1. A state variable $x \in X \subset \mathbb{R}^n$, and for each player $j \in M$, a control vector $u_j \in U_j \subset \mathbb{R}^{p_j}$, where X and U_j's are open domains.
2. A state equation (which is an n-dimensional ordinary differential equation) and an initial value for the state (at time t^0)

$$\dot{x}(t) = f(x(t), u(t), t), \tag{2.1}$$

$$x(t^0) = x^0, \tag{2.2}$$

where $\{x(t) : t \in [t^0, T]\}$ is the state trajectory and $\{u(t) \triangleq (u_1(t), \dots, u_m(t)) : t \in [0, T]\}$ is the control (or action) schedule (or simply the control) chosen by the m players, with $u_j(\cdot)$ generated by player j as a p_j-dimensional function. Here $\dot{x}(t)$ denotes the time derivative $\frac{d}{dt}x(t)$. The function $f(\cdot,\cdot,\cdot) : \mathbb{R}^n \times \mathbb{R}^{p_1+\cdots+p_m} \times \mathbb{R} \mapsto \mathbb{R}^n$ is assumed to be continuously differentiable (of class C^1) in $x, u,$ and t.

3. If the control vector generated by the m players is a measurable function of t, or more simply a piecewise continuous function, there is a unique state trajectory solution of (2.1) and (2.2), and each player $j \in M$ receives a cumulative reward over the time horizon $[t^0, T]$:

$$J_j(u(\cdot); x^0, t^0) = \int_{t^0}^{T} g_j(x(t), u(t), t)\, dt + S_j(x(T), T), \tag{2.3}$$

where g_j is player j's instantaneous reward rate and S_j is the terminal reward, also called *salvage value function*. The functions $g_j(\cdot,\cdot,\cdot) : \mathbb{R}^n \times \mathbb{R}^{p_1+\cdots+p_m} \times \mathbb{R} \mapsto \mathbb{R}$, $j \in M$, are assumed to be continuously differentiable in x, u, and t, and $S_j(\cdot,\cdot) : \mathbb{R}^n \times \mathbb{R} \mapsto \mathbb{R}$, $j \in M$, are assumed to be continuously differentiable in x and t.

2.1.2 Control Constraints

The choice of a control by player j is subject to a pointwise constraint for each $t \in [t^0, T]$

$$u_j(t) \in U_j, \quad t \in [t^0, T], \tag{2.4}$$

where U_j is referred to as the player's *admissible pointwise control set*. In a more general setting, the admissible control set may depend on time t and state $x(t)$. Then, the choice of a control is subject to a constraint

$$u_j(t) \in U_j(x(t), t)), \quad t \in [t^0, T], \tag{2.5}$$

where the correspondence, or point-to-set mapping $\left\{U_j(\cdot,\cdot) : \mathbb{R}^n \times \mathbb{R} \mapsto 2^{\mathbb{R}^{p_j}}\right\}$ is assumed to be *upper-semicontinuous*. In such a case, player j will of course also have to have access to the current value of the state, which brings in the question of what information a player has to have access to before constructing her control; this is related to the information structure of the differential game, without which the formulation of a differential game would not be complete. Information structures will be introduced shortly, in the next subsection.

2.1.3 Target

The determination of the terminal time T can be either prespecified (as part of the initial data), $T \in \mathbb{R}^+$, or the result of the state trajectory reaching a *target*. The target is defined by a *surface* or *manifold* defined by an equation of the form

$$\Theta(x, t) = 0, \tag{2.6}$$

where $\Theta(\cdot, \cdot) : \mathbb{R}^n \times \mathbb{R} \mapsto \mathbb{R}$ is continuously differentiable. The trajectory ends (reaches the target), and the rewards are computed, at the first time T when the condition $\Theta(x(T), T) = 0$ is satisfied.

2.1.4 Infinite-Horizon Games

In economic and engineering applications, one also considers games where the terminal time T may tend to ∞. The payoff to player j is then defined as

$$J_j(u(\cdot); x^0, t_0) = \int_0^\infty e^{-\rho_j t} g_j(x(t), u(t))dt. \tag{2.7}$$

Note that player j's payoff does not include a terminal reward and the reward rate depends explicitly on the running time t through a discount factor $e^{-\rho_j t}$, where ρ_j is a discount rate satisfying $\rho_j \geq 0$, which could be player dependent. An important issue in an infinite-horizon dynamic optimization problem (one-player version of the problem above) is the fact that when the discount rate ρ_j is set to zero, then the integral payoff (2.7) may not be well defined, as the integral may not converge to a finite value for all feasible control paths $u(\cdot)$, and in some cases for none. In such situations, one has to rely on a different notion of optimality, e.g., **overtaking optimality**, a concept well developed in Carlson et al. (1991). We refer the reader to the next chapter (Chap. 3) for a deeper discussion of this topic.

2.2 Information Structures and Strategies

2.2.1 Open Loop Versus State Feedback

To complete the formulation of a differential game, one has to describe precisely the information available to each player (regarding the state and past actions of other players) when they choose their controls at time t. Let us first focus on two information structures of common use in applications of differential games, namely, *open-loop* and *state-feedback* information structures. Letting $\nu(t)$ denote the information available to a generic player at time t, we say that the information structure is *open loop* if

$$\nu(t) = \{x^0, t\},$$

that is, the available information is the current time and the initial state. An information structure is *state feedback* if

$$\nu(t) = \{x(t), t\},$$

that is, the available information is the current state of the system in addition to the current time. We say that a differential game has open-loop (respectively, state-feedback) information structure if every player in the game has open-loop (respectively, state-feedback) information. It is of course possible for some players to have open-loop information while others have state-feedback information, but we will see later that such a mixed information structure does not lead to a well-defined differential game unless the players who have access to the current value of the state also have access to the initial value of the state, that is,

$$\nu(t) = \{x(t), x^0, t\}.$$

Another more general information structure is the one with memory, known as *closed-loop with memory*, where at any time t a player has access to the current value of the state and also recalls all past values, that is,

$$\nu(t) = \{x(s), s \leq t\}.$$

The first two information structures above (open loop and state feedback) are common in *optimal control theory*, i.e., when the system is controlled by only one player. In *optimal control* of a deterministic system, the two information structures are in a sense equivalent. Typically, an optimal state-feedback control is obtained by "synthesizing" the optimal open-loop controls defined from all possible initial states.[1] It can also be obtained by employing *dynamic programming* or equivalently *Bellman's optimality principle* (Bellman 1957). The situation is, however, totally different for nonzero-sum differential games. The open-loop and state-feedback information structures generally lead to two very different types of differential games, except for the cases of two-player zero-sum differential games (see ▸ Chap. 8, "Zero-sum Differential Games" in this *Handbook* and also our brief discussion later in this chapter) and differential games with identical objective functions for the players (known as dynamic teams, which are equivalent to optimal control problems as we are dealing with deterministic systems) – or differential games that are *strategically equivalent*[2] to zero-sum differential games or dynamic team problems. Now, to understand the source of the difficulty in the nonequivalence of two differential games that differ (only) in their information structures, consider the case when the control sets are state dependent, i.e., $u_j(t) \in U_j(x(t), t)$. In the optimal control case, when the only player who controls the system selects a control schedule, she can compute also the associated unique state trajectory. In fact, selecting a control amounts to selecting a trajectory. So, it may be possible to select jointly the control and the associated trajectory to ensure that at each time t the constraint $u(t) \in U(x(t), t)$ is satisfied; hence, it is possible to envision an open-loop control for such a system. Now, suppose that there is another player involved in controlling the system; let us call them players 1 and 2. When player 1 defines

[1]See, e.g., the classical textbook on optimal control by Lee and Markus (1972) for examples of synthesis of state-feedback control laws.

[2]This property will be discussed later in the chapter; in the context of static games, "strategic equivalence" has been discussed in Chap. 1.

her control schedule, she does not know the control schedule of the other player, unless there has been an exchange of information between the two players and a tacit agreement to coordinate their choices of control. Therefore, player 1, not knowing what player 2 will do, cannot decide in advance if her control at time t will be in the admissible set $U_1(x(t), t)$ or not. Hence, in that case, it is impossible for the players to devise feasible and implementable open-loop controls, whereas this would indeed be possible under the state-feedback information structure. The difference between the two information structures is in fact even more subtle, since even when the admissible control sets are not state dependent, knowing at each instant t what the state $x(t)$ is, or not having access to this information will lead to two different types of *noncooperative games in normal form* as we will see in the coming sections.

2.2.2 Strategies

In game theory one calls *strategy* (or *policy* or *law*) a rule that associates an action to the information available to a player at a position of the game. In a differential game, a strategy γ_j for player j is a function that associates to each possible information $\nu(t)$ at t, a control value $u_j(t)$ in the admissible control set. Hence, for each information structure we have introduced above, we will have a different class of strategies in the corresponding differential game. We make precise below the classes of strategies corresponding to the first two information structures, namely, open loop and state feedback.

Definition 1. Assuming that the admissible control sets U_j are not state dependent, an **open-loop strategy** γ_j for player j $(j \in M)$ selects a control action according to the rule

$$u_j(t) = \gamma_j(x^0, t), \quad \forall x^0, \forall t, j \in M, \tag{2.8}$$

where $\gamma_j(\cdot, \cdot) : \mathbb{R}^n \times \mathbb{R} \mapsto U_j$ is a function measurable (or piecewise continuous) in t, for each fixed x^0. The class of all such strategies for player j is denoted by Γ_j^{OL} or simply by Γ_j.

Definition 2. A **state-feedback strategy** γ_j for player j $(j \in M)$ selects a control action according to a state-feedback rule

$$u_j(t) = \gamma_j(x(t), t), \quad j \in M, \tag{2.9}$$

where $\gamma_j(\cdot, \cdot) : (x, t) \in \mathbb{R}^n \times \mathbb{R} \mapsto U_j(x, t)$ is a given function that must satisfy the required regularity conditions imposed on feedback controls.[3] The class of all such strategies for player j is denoted by Γ_j^{SF} or simply by Γ_j.

[3]These are conditions which ensure that when all players' strategies are implemented, then the differential equation (2.1) describing the evolution of the state admits a unique piecewise continuously differentiable solution for each initial condition x^0; see, e.g., Başar and Olsder (1999).

Remark 1. In the literature on dynamic/differential games, state-feedback strategy is sometimes called "Markovian," in contrast to "open loop," with the argument being that the former implies less "commitment" than the latter. Such an interpretation is misleading on two counts. First, one can actually view both classes of strategies as Markovian, since, at each time t, they exploit only the information received at time t. The strategies do not exploit the **history** of the information received up to time t, which is in fact not available. Second, in both cases, a strategy is a full commitment. Using an open-loop strategy means that the player commits, at the initial time, to a fixed time path for her control, that is, her choice of control at each instant of time is predetermined. When using a state-feedback strategy, a player commits to the use of a well-defined servomechanism to control the system, that is, her reaction to the information concerning the state of the system is predetermined. The main advantages of state-feedback strategies lie elsewhere: (i) state-feedback strategies are essential if one has a stochastic differential game (a differential game where the state dynamics are perturbed by disturbance (or noise) with a stochastic description); in fact, if we view a deterministic differential game as the "limit" of a sequence of stochastic games with vanishing noise, we are left with state-feedback strategies. (ii) State-feedback strategies allow us to introduce the refined equilibrium solution concept of "subgame-perfect Nash equilibrium," which is a concept much appreciated in economic applications, and will be detailed below.

3 Nash Equilibria

Recall the definition of a *Nash equilibrium* for a game in *normal form* (equivalently, *strategic form*).

Definition 3. With the initial state x^0 fixed, consider a differential game in normal form, defined by a set of m players, $M=\{1,\dots,m\}$, and for each player j $(j\in M)$ a strategy set Γ_j and a payoff function

$$\bar{J}_j : \Gamma_1 \times \cdots \times \Gamma_j \times \cdots \times \Gamma_m \mapsto \mathbb{R}, \quad j \in M.$$

Nash equilibrium is a strategy m-tuple $\gamma^* = (\gamma_1^*, \dots, \gamma_m^*)$, such that for each player j the following holds:

$$\bar{J}_j(\gamma^*) \geq \bar{J}_j([\gamma_j, \gamma_{-j}^*]), \quad \forall \gamma_j \in \Gamma_j, \tag{2.10}$$

where $\gamma_{-j}^* := (\gamma_i^* : i \in M \setminus j)$ and $[\gamma_j, \gamma_{-j}^*]$ is the m-tuple obtained when, in γ^*, γ_j^* is replaced by γ_j. In other words, in Nash equilibrium, for each player j, the strategy γ_j^* is the best reply to the $(m-1)$-tuple of strategies γ_{-j}^* chosen by the other players.

Corresponding to the first two information structures we have introduced for differential games, we will now define two different *games in normal form*, leading to two different concepts of Nash equilibrium for nonzero-sum differential games.

3.1 Open-Loop Nash Equilibrium (OLNE)

Assume that the admissible control sets U_j, $j \in M$ are not state dependent. If the players use open-loop strategies (2.8), each γ_j defines a unique control schedule $u_j(\cdot) : [0, T] \mapsto U_j$ for each initial state x^0. The payoff functions for the normal form game are defined by

$$\bar{J}_j(\gamma) = J_j(u(\cdot); x^0, t^0), \quad j \in M, \tag{2.11}$$

where $J_j(\cdot; \cdot, \cdot)$ is the reward function defined in (2.3). Then, we have the following definition:

Definition 4. The control m-tuple $u^*(\cdot) = \left(u_1^*(\cdot), \dots, u_m^*(\cdot)\right)$ is an **open-loop Nash equilibrium** (OLNE) at $(\mathbf{x}^0, t^0)$ if the following holds:

$$J_j(u^*(\cdot); x^0, t^0) \geq J_j([u_j(\cdot), u_{-j}^*(\cdot)]; x^0, t^0), \quad \forall u_j(\cdot), j \in M,$$

where $u_j(\cdot)$ is any admissible control of player j and $[u_j(\cdot), u_{-j}^*(\cdot)]$ is the m-tuple of controls obtained by replacing the j-th block component in $u^*(\cdot)$ by $u_j(\cdot)$.

Note that in the OLNE, for each player j, $u_j^*(\cdot)$ solves the optimal control problem

$$\max_{u_j(\cdot)} \left\{ \int_{t^0}^{T} g_j\left(x(t), [u_j(t), u_{-j}^*(t)], t\right) dt + S_j(x(T)) \right\},$$

subject to the state equation

$$\dot{x}(t) := \frac{d}{dt}x(t) = f\left(x(t), [u_j(t), u_{-j}^*(t)], t\right), \quad x(t^0) = x^0, \tag{2.12}$$

control constraints $u_j(t) \in U_j$, and target $\Theta(\cdot, \cdot)$. Further note that OLNE strategies will in general also depend on the initial state x^0, but this is information available to each player under the open-loop information structure.

3.2 State-Feedback Nash Equilibrium (SFNE)

Now consider a differential game with the state-feedback information structure. The system is then driven by a state-feedback strategy m-tuple $\gamma(\mathbf{x}, t)=(\gamma_j(\mathbf{x}, t): j \in M)$, with $\gamma_j \in \Gamma_j^{\mathrm{SF}}$ for $j \in M$. Its dynamics are thus defined by

$$\dot{x}(t) := \frac{d}{dt}x(t) = f(x(t), \gamma(x(t), t), t), \quad x(t^0) = x^0. \tag{2.13}$$

The normal form of the game, at (x^0, t^0), is now defined by the payoff functions[4]

$$\bar{J}_j(\gamma; x^0, t^0) = \int_{t^0}^{T} g_j(x(t), \gamma(x(t), t), t)dt + S_j(x(T)), \tag{2.14}$$

where, for each fixed x^0, $x(\cdot) : [t^0, T] \mapsto \mathbb{R}^n$ is the state trajectory solution of (2.13).

In line with the convention in the OL case, let us introduce the notation

$$\gamma_{-j}(t, x(t)) \triangleq \left(\gamma_1(x(t), t), \ldots, \gamma_{j-1}(x(t), t), \gamma_{j+1}(x(t), t), \ldots, \gamma_m(x(t), t)\right),$$

for the *strategy* $(m-1)$*-tuple* where the strategy of player j does not appear.

Definition 5. The state-feedback m-tuple $\gamma^* = \left(\gamma_1^*, \ldots, \gamma_m^*\right)$ is a **state-feedback Nash equilibrium** (SFNE) on[5] $X \times \left[t^0, \mathcal{T}\right]$ if for any initial data $(x^0, t^0) \in X \times [0, \mathcal{T}] \subset \mathbb{R}^n \times \mathbb{R}^+$, the following holds:

$$\bar{J}_j(\gamma^*; x^0, t^0) \geq \bar{J}_j([\gamma_j(\cdot), \gamma_{-j}^*(\cdot)]; x^0, t^0), \quad \forall \gamma_j \in \Gamma_j^{SF}, \ j \in M,$$

where $[\gamma_j, \gamma^{-j*}]$ is the m-vector of strategies obtained by replacing the j-th block component in γ^* by γ_j.

In other words, $\{u_j^*(t) \equiv \gamma_j^*(x^*(t), t) : t \in [t^0, T]\}$, where $x^*(\cdot)$ is the equilibrium trajectory generated by γ^* from (x^0, t^0), solves the optimal control problem

$$\max_{u_j(\cdot)} \left\{ \int_{t^0}^{T} g_j\left(x(t), \left[u_j(t), \gamma_{-j}^*(x(t), t)\right], t\right) dt + S_j(x(T)) \right\}, \tag{2.15}$$

[4]With a slight abuse of notation, we have included here also the pair (x^0, t^0) as an argument of $\bar{J}_j$, since under the SF information γ does not have (x^0, t^0) as an argument for $t > t^0$.

[5]We use $\mathcal{T}$ instead of T because, in a general setting, T may be endogenously defined as the time when the target is reached.

subject to the state equation

$$\dot{x}(t) = f(x(t), \left[u_j(t), \gamma^*_{-j}(x(t), t)\right], t), \ x(t^0) = x^0, \tag{2.16}$$

control constraints $u_j(t) \in U_j(x(t), t)$, and target $\Theta(\cdot, \cdot)$. We can also say that γ_j^* is the optimal state-feedback control $u_j^*(\cdot)$ for the problem (2.15) and (2.16). We also note that the single-player optimization problem (2.15) and (2.16) is a standard optimal control problem whose solution can be expressed in a way compatible with the state-feedback information structure, that is, solely as a function of the current value of the state and current time, and not as a function of the initial state and initial time. The remark below further elaborates on this point.

Remark 2. Whereas an open-loop Nash equilibrium is defined only for the given initial data, here the definition of a state-feedback Nash equilibrium asks for the equilibrium property to hold for all initial points, or data, in a region $X \times \left[t^0, \mathcal{T}\right] \subset \mathbb{R}^n \times \mathbb{R}^+$. This is tantamount to asking a state-feedback Nash equilibrium to be **subgame perfect** (Selten 1975), in the parlance of game theory, or **strongly time consistent** (Başar 1989). Indeed, even if the state trajectory is perturbed, either because a player has had a "trembling hand" or an unforeseen small shock happened, holding on to the same state-feedback strategy will still constitute a Nash equilibrium in the limit as the perturbations vanish; this property is more pronounced in the case of linear-quadratic differential games (games where the state dynamics are linear, payoff functions are jointly quadratic in the state and the controls, and the time horizon is fixed), in which case the stochastic perturbations in the state equation do not have to be vanishingly small as long as they have zero mean (Başar 1976, 1977). It should be clear that open-loop Nash equilibrium strategies do not possess such a property.

3.3 Necessary Conditions for a Nash Equilibrium

For the sake of simplicity in the exposition below, we will henceforth restrict the target set to be defined by the simple given of a terminal time, that is, the set $\{(t, x) : t = T\}$. Also the control constraint set U_j, $j \in M$ will be taken to be independent of state and time. As noted earlier, at a Nash equilibrium, each player solves an optimal control problem where the system's dynamics are influenced by the strategic choices of the other players. We can thus write down necessary optimality conditions for each of the m optimal control problems, which will then constitute a set of necessary conditions for a Nash equilibrium. Throughout, we make the assumption that sufficient regularity holds so that all the derivatives that appear in the necessary conditions below exist.

3.3.1 Necessary Conditions for an OLNE

By using the necessary conditions for an open-loop optimal control, obtained, e.g., from the *maximum principle* (see, e.g., Başar and Olsder 1999; Bryson et al. 1975), we arrive at the conditions (2.17), (2.18), (2.19), (2.20), and (2.21) below, which are necessary for an open-loop Nash equilibrium. Let us introduce the individual Hamiltonians, with H_j being the Hamiltonian for player j,[6]

$$H_j(x, u, \lambda_j, t) = g_j(x, u, t) + \lambda_j(t) f(x, u, t), \tag{2.17}$$

where $\lambda_j(\cdot)$ is the *adjoint* (or *costate*) variable, which satisfies the *adjoint variational equation* (2.18), along with the *transversality condition* (2.19):

$$\dot{\lambda}_j(t) = -\frac{\partial}{\partial x} H_j|_{x^*(t),u^*(t),t}, \tag{2.18}$$

$$\lambda_j(T) = \frac{\partial}{\partial x} S_j|_{x^*(T),T.} \tag{2.19}$$

Further, H_j is maximized with respect to u_j, with all other players' controls fixed at NE, that is,

$$u_j^*(t) = \arg\max_{u_j \in U_j} H_j(x^*(t), u_j, u_{-j}^*(t), \lambda_j(t), t). \tag{2.20}$$

If the solution to the maximization problem above is in the interior of U_j, then naturally a necessary condition is for the first derivative to vanish at u_j^* for all t, that is,

$$\frac{\partial}{\partial u_j} H_j|_{x^*(t),u^*(t),t} = 0, \tag{2.21}$$

and for the Hessian matrix of second derivatives (with respect to u_j) to be nonnegative definite.

3.3.2 Necessary Conditions for SFNE

The state-feedback NE can be obtained in various different ways. One could again use the approach above, but paying attention to the fact that in the optimal control problem faced by a generic player, the other players' strategies are now dependent on the current value of the state. A second approach would be to adapt to this problem the method used in optimal control to directly obtain state-feedback controls (i.e., dynamic programming). We discuss here both approaches, first in this subsection the former. The Hamiltonian for player j is again:

$$H_j(x, u, \lambda_j, t) = g_j(x, u, t) + \lambda_j(t) f(x, u, t). \tag{2.22}$$

---

[6] We use the convention that $\lambda_j(t) f(\cdot, \cdot, \cdot)$ is the scalar product of two n dimensional vectors $\lambda_j(t)$ and $f(\cdots)$.

The controls u_i for $i \in M \setminus j$ are now defined by the state-feedback rules $\gamma_i^*(x,t)$. Along the equilibrium trajectory $\{x^*(t) : t \in [t^0, T]\}$, the optimal control of player j is $u_j^*(t) = \gamma_j^*(x^*(t), t)$. Then, as the counterpart of (2.18) and (2.19), we have $\lambda_j(\cdot)$ satisfying (as a necessary condition)

$$\dot{\lambda}_j(t) = -\left(\frac{\partial}{\partial x} H_j + \sum_{i \in M \setminus j} \frac{\partial}{\partial u_i} H_j \frac{\partial}{\partial x} \gamma_i^* \right) |_{x^*(t), u^*(t), t}, \tag{2.23}$$

$$\lambda_j(T) = \frac{\partial}{\partial x} S_j |_{x^*(T), T}, \tag{2.24}$$

where the second term in (2.23), involving a summation, is a reflection of the fact that H_j depends on x not only through g_j and f but also through the strategies of the other players. The presence of this extra term clearly makes the necessary condition for the state-feedback solution much more complicated than for open-loop solution.

Again, $u_j^*(t) = \gamma_j^*(x^*(t), t)$ maximizes the Hamiltonian H_j for each t, with all other variables fixed at equilibrium:

$$u_j^*(t) = \arg \max_{u_j \in U_j} H_j(x^*(t), u_j, \gamma_{-j}^*(x^*(t), t), \lambda_j, t). \tag{2.25}$$

If the solution to the maximization problem above is in the interior of U_j, then as in (2.21) a necessary condition is for the first derivative to vanish at u_j^* for all t, that is,

$$\frac{\partial}{\partial u_j} H_j |_{x^*(t), u_j^*(t), \gamma_{-j}^*(x^*(t), t), t} = 0, \tag{2.26}$$

and for the Hessian matrix of second derivatives (with respect to u_j) to be nonnegative definite.

Remark 3. The summation term in (2.23) is absent in three important cases: (i) in optimal control problems ($m = 1$), since $\frac{\partial}{\partial u} H \frac{\partial u}{\partial x} = 0$; (ii) in two-person zero-sum differential games, because $H_1 \equiv -H_2$ so that for player 1, $\frac{\partial}{\partial u_2} H_1 \frac{\partial u_2}{\partial x} = -\frac{\partial}{\partial u_2} H_2 \frac{\partial u_2}{\partial x} = 0$, and likewise for player 2; and (iii) in open-loop nonzero-sum differential games, because $\frac{\partial u_j}{\partial x} = 0$. It would also be absent in nonzero-sum differential games with state-feedback information structure that are **strategically equivalent** (Başar and Olsder 1999) to (i) (single objective) team problems (which in turn are equivalent to single-player optimal control problems) or (ii) two-person zero-sum differential games.

3.4 Constructing an SFNE Using a Sufficient Maximum Principle

As alluded to above, the necessary conditions for an SFNE as presented are not very useful to compute a state-feedback Nash equilibrium, as one has to infer the form of the partial derivatives of the equilibrium strategies, in order to write the adjoint equations (2.24). However, as an alternative, the sufficient maximum principle given below can be a useful tool when one has an a priori guess of the class of equilibrium strategies (see, Haurie et al. 2012, page 249).

Theorem 1. *Assume that the terminal reward functions S_j are continuously differentiable and concave, and let $X \subset \mathbb{R}^n$ be a state constraint set where the state $x(t)$ belongs for all t. Suppose that an m-tuple $\gamma^* = \left(\gamma_1^*, \ldots, \gamma_m^*\right)$ of state-feedback strategies $\gamma_j : X \times [t^0, T] \mapsto \mathbb{R}^{m_j}$, $j \in M$, is such that*

(i) *$\gamma^*(x, t)$ is continuously differentiable in x almost everywhere, and piecewise continuous in t;*

(ii) *$\gamma^*(x, t)$ generates at (x^0, t^0) a unique trajectory $x^*(\cdot) : [t^0, T] \mapsto X$, solution of*

$$\dot{x}(t) = f(x(t), \gamma^*(x, t), t), \quad x(t^0) = x^0,$$

which is absolutely continuous and remains in the interior of X;

(iii) *there exist m **costate vector** functions $\lambda_j(\cdot) : [t^0, T] \mapsto \mathbb{R}^n$, which are absolutely continuous and such that, for all $j \in M$, if we define the Hamiltonians*

$$\begin{aligned} &H_j(x(t), [u_j, u_{-j}], \lambda_j(t), t) \\ &= g_j\left(x(t), [u_j, u_{-j}], t\right) + \lambda_j(t) f(x(t), [u_j, u_{-j}], t), \end{aligned}$$

and the equilibrium Hamiltonians

$$\mathcal{H}_j^*(x^*(t), \lambda_j(t), t) = \max_{u_j \in U_j} H_j(x^*(t), [u_j, \gamma_{-j}^*(x^*(t), t)], \lambda_j(t), t), \tag{2.27}$$

the maximum in (2.27) is reached at $\gamma_j^*(x^*(t), t)$, i.e.,

$$\mathcal{H}_j^*(x^*(t), \lambda_j(t), t) = \max_{u_j \in U_j} H_j(x^*(t), \gamma^*(x^*(t), t), \lambda_j(t), t); \tag{2.28}$$

(iv) *the functions $x \mapsto \mathcal{H}_j^*(x, \lambda_j(t), t)$ where $\mathcal{H}_j^*$ is defined as in (2.27), but at position (t, x), are continuously differentiable and concave for all $t \in [t^0, T]$ and $j \in M$;*

(v) *the costate vector functions $\lambda_j(\cdot)$, $j \in M$, satisfy the following adjoint differential equations for almost all $t \in [t^0, T]$,*

$$\dot{\lambda}_j(t) = -\frac{\partial}{\partial x}\mathcal{H}_j^*|_{(x^*(t),\lambda_j(t),t)}, \tag{2.29}$$

along with the ***transversality conditions***

$$\lambda_j(T) = \frac{\partial}{\partial x} S_j|_{(x^*(T),T)}. \tag{2.30}$$

Then, $\left(\gamma_1^*, \ldots, \gamma_m^*\right)$ *is an SFNE at* (x^0, t^0).

3.5 Constructing an SFNE Using Hamilton-Jacobi-Bellman Equations

We now discuss the alternative dynamic programming approach which delivers the state-feedback solution directly without requiring synthesis or guessing of the solution. The following theorem captures the essence of this effective tool for SFNE (see, Başar and Olsder 1999, page 322; Haurie et al. 2012, page 252).

Theorem 2. *Suppose that there exists an* m*-tuple* $\gamma^* = \left(\gamma_1^*, \ldots, \gamma_m^*\right)$ *of state-feedback laws, such that*

(i) *for any admissible initial point* (x^0, t^0)*, there exists a unique, absolutely continuous solution* $t \in [t^0, T] \mapsto x^*(t) \in X \subset \mathbb{R}^n$ *of the differential equation*

$$\dot{x}^*(t) = f(x^*(t), \gamma_1^*(t, x^*(t)), \ldots, \gamma_m^*\left(t, x^*(t)\right), t), \quad x^*(t^0) = x^0;$$

(ii) *there exist continuously differentiable* value functionals $V_j^* : X \times [t^0, T] \mapsto \mathbb{R}$, *such that the following coupled Hamilton-Jacobi-Bellman (HJB) partial differential equations are satisfied for all* $(x, t) \in X \times [t^0, T]$

$$-\frac{\partial}{\partial t} V_j^*(x,t) = \max_{u_j \in U_j} \Big\{ g_j\left(x, [u_j, \gamma_{-j}^*(x,t)], t\right) + \frac{\partial}{\partial x} V_j^*(x,t) f(x, \left[u_j, \gamma_{-j}^*(x,t)\right], t) \Big\} \tag{2.31}$$

$$= g_j\left(x, [\gamma^*(x,t)], t\right) + \frac{\partial}{\partial x} V_j^*(x,t) f(x, \gamma^*(x,t), t); \tag{2.32}$$

(iii) *the boundary conditions*

$$V_j^*(x, T) = S_j(x), \tag{2.33}$$

are satisfied for all $x \in X$ *and* $j \in M$.

Then, $\gamma_j^*(x,t)$, is a maximizer of the right-hand side of the HJB equation for player j, and the m-tuple $\left(\gamma_1^*,\ldots,\gamma_m^*\right)$ is an SFNE at every initial point $(x^0,t^0) \in X \times [t^0,T]$.

Remark 4. Note that once a complete set of value functionals, $\{V_j,\ j \in M\}$, is identified, then (2.31) directly delivers the Nash equilibrium strategies of the players in state-feedback form. Hence, in this approach one does not have to guess the SFNE but rather the structure of each player's value function; this can be done in a number of games, with one such class being linear-quadratic differential games, as we will see shortly. Also note that Theorem 2 provides a set of sufficient conditions for SFNE, and hence once a set of strategies are found satisfying them, we are assured of their SFNE property. Finally, since the approach entails dynamic programming, it directly follows from (2.31) that a natural restriction of the set of SFNE strategies obtained for the original differential game to a shorter interval $[s,T]$, with $s > t^0$, constitutes an SFNE for the differential game which is similarly formulated but on the shorter time interval $[s,T]$. Hence, the SFNE is subgame perfect and strongly time consistent.

3.6 The Infinite-Horizon Case

Theorems 1 and 2 were stated under the assumption that the time horizon is finite. If the planning horizon is infinite, then the transversality or boundary conditions, that is, $\lambda_j(T) = \frac{\partial S_j}{\partial x}(x(T),T)$ in Theorem 1 and $V_j(x(t),T) = S_j(x(t),T)$ in Theorem 2, have to be modified. Below we briefly state the required modifications, and work out a scalar example in the next subsection to illustrate this.

If the time horizon is infinite, the dynamic system is autonomous (i.e., f does not explicitly depend on t) and the objective functional of player j is as in (2.7), then the transversality conditions in Theorem 1 are replaced by the limiting conditions:

$$\lim_{t\to+\infty} e^{-\rho_j t} q_j(t) = 0, \quad \forall j \in M, \tag{2.34}$$

where $q_j(t) = e^{\rho_j t}\lambda_j(t)$ is the so-called current-value costate variable. In the coupled set of HJB equations of Theorem 2, the value function $V_j^\star(x,t)$ is multiplicatively decomposed as

$$V_j^\star(x,t) = e^{-\rho_j t}\mathcal{V}_j^\star(x), \tag{2.35}$$

and the boundary condition (2.33) is replaced by

$$V_j^\star(x,t) \to 0, \text{ when t } \to \infty, \tag{2.36}$$

which is automatically satisfied if $\mathcal{V}_j^\star(x)$ is bounded.

3.7 Examples of Construction of Nash Equilibria

We consider here a two-player infinite-horizon differential game with scalar linear dynamics and quadratic payoff functions, which will provide an illustration of the results of the previous subsection, also to be viewed as illustration of Theorems 1 and 2 in the infinite-horizon case.

Let $u_j(t)$ be the scalar control variable of player j, $j = 1, 2$, and $x(t)$ be the state variable, with $t \in [0, \infty)$. Let player j's optimization problem be given by

$$\max_{u_j} \left\{ J_j = \int_0^\infty e^{-\rho t} \left(u_j(t) \left(\kappa - \frac{1}{2} u_j(t) \right) - \frac{1}{2} \varphi x^2(t) \right) dt \right\},$$
$$\text{such that} \quad \dot{x}(t) = u_1(t) + u_2(t) - \alpha x(t), \quad x(0) = x^0,$$

where φ and κ are positive parameters, $0 < \alpha < 1$, and $\rho > 0$ is the discount parameter. This game has the following features: (i) the objective functional of player j is quadratic in the control and state variables and only depends on the player's own control variable; (ii) there is no interaction (coupling) either between the control variables of the two players or between the control and the state variables; (iii) the game is fully symmetric across the two players in the state and the control variables; and (iv) by adding the term $e^{-\rho t} \left(u_i(t) \left(\kappa - \frac{1}{2} u_i(t) \right) \right)$, $i \neq j$, to the integrand of J_j, for $j = 1, 2$, we can make the two objective functions identical:

$$J := \int_0^\infty e^{-\rho t} \left(u_1(t) \left(\kappa - \frac{1}{2} u_1(t) \right) + u_2(t) \left(\kappa - \frac{1}{2} u_2(t) \right) - \frac{1}{2} \varphi x^2(t) \right) dt. \tag{2.37}$$

The significance of this last feature will become clear shortly when we discuss the OLNE (next). Throughout the analysis below, we suppress the time argument when no ambiguity may arise.

Open-Loop Nash Equilibrium (OLNE). We first discuss the significance of feature (iv) exhibited by this scalar differential game. Note that when the information structure of the differential game is open loop, adding to the objective function of a player (say, player 1) terms that involve only the control of the other player (player 2) does not alter the optimization problem faced by player 1. Hence, whether player j maximizes J_j, or J given by (2.37), makes no difference as far as the OLNE of the game goes. Since this applies to both players, it readily follows that every OLNE of the original differential game is also an OLNE of the single-objective optimization problem (involving maximization of J by each player). In such a case, we say that the two games are *strategically equivalent*, and note that the second game (described by the single-objective functional J) is a dynamic team. Now, Nash equilibrium (NE) in teams corresponds to person-by-optimality, and not team optimality (which means joint optimization by members of the team), but when every person-by-person optimal solution is also team optimal (the reverse implication is always true), then one can obtain all NE of games strategically equivalent to a particular dynamic team by solving for team optimal (equivalently, globally optimal) solutions of the team. Further, when solving for team-optimal solutions in deterministic teams, whether the information structure is open loop or state feedback does not make any difference, as mentioned earlier. In the particular dynamic team of this example, since J is strictly concave in u_1 and u_2 and x (jointly), and the state equation is linear, every person-by-person optimal solution is indeed team optimal, and because of strict concavity the problem admits a unique globally optimal solution. Hence, the OLNE of the original game exists and is unique.

Having established the correspondence with a deterministic concave team and thereby the existence of a unique OLNE, we now turn to our main goal here, which is to apply the conditions obtained earlier for OLNE to the differential game at hand. Toward that end, we introduce the current-value Hamiltonian of player j:

$$\mathcal{H}_j(x, \lambda, u_1, u_2) = u_j \left(\kappa - \frac{1}{2}u_j\right) - \frac{1}{2}\varphi x^2 + q_j(u_1 + u_2 - \alpha x), \;\; i = 1, 2,$$

where $q_j(t)$ is the current-value costate variable, at time t, defined as

$$q_j(t) = e^{\rho_j t}\lambda_j(t). \tag{2.38}$$

Being strictly concave in u_j, $\mathcal{H}_j$ admits a unique maximum, achieved by

$$u_j = \kappa + q_j, \; j = 1, 2. \tag{2.39}$$

Note that the Hamiltonians of both players are strictly concave in x, and hence the equilibrium Hamiltonians also are. Then, the equilibrium conditions read:

$$\dot{q}_j = \rho q_j - \frac{\partial}{\partial x}\mathcal{H}_j = (\rho + \alpha)q_j + \varphi x, \quad \lim_{t\to+\infty} e^{-\rho t}q_j(t) = 0, \; j = 1, 2,$$
$$\dot{x} = 2\kappa + q_1 + q_2 - \alpha x, \quad x(0) = x^0.$$

It is easy to see that $q_1(t) = q_2(t) =: q(t), \forall t \in [0, \infty)$, and therefore $u_1(t) = u_2(t), \forall t \in [0, \infty)$. This is not surprising given the symmetry of the game. We then have a two-equation differential system in x and q:

$$\begin{pmatrix} \dot{x} \\ \dot{q} \end{pmatrix} = \begin{pmatrix} -\alpha & 2 \\ \varphi & \rho + \alpha \end{pmatrix} \begin{pmatrix} x \\ q \end{pmatrix} + \begin{pmatrix} 2\gamma \\ 0 \end{pmatrix}.$$

We look for the solution of this system converging to the steady state which is given by

$$(x_{ss}, q_{ss}) = \left(\frac{2\kappa(\alpha + \rho)}{\alpha^2 + \alpha\rho + 2\varphi}, -\frac{2\kappa\varphi}{\alpha^2 + \alpha\rho + 2\varphi}\right).$$

The solution can be written as

$$x(t) = (x^0 - x_{ss})e^{\mu_1 t} + x_{ss},$$
$$q(t) = -(x^0 - x_{ss})\frac{2\varphi}{2\alpha + \rho + \sqrt{(2\alpha + \rho)^2 + 8\varphi}}e^{\mu_1 t} + q_{ss},$$

where μ_1 is the negative eigenvalue of the matrix associated with the differential equations system and is given by

$$\mu_1 = \frac{1}{2}(\rho - \sqrt{(2\alpha + \rho)^2 + 8\varphi}).$$

Using the corresponding expression for $q(t)$ for q_j in (2.39) leads to the OLNE strategies (which are symmetric).

State-Feedback Nash Equilibrium (SFNE). The strategic equivalence between the OL differential game and a team problem we established above does not carry over to the differential game with state-feedback information structure, since adding any term to J_1 that involves control u_2 of player 2 will alter the optimization problem faced by player 1, since u_2 depends on u_1 through the state x. Hence, the example system is a genuine game under SF information, and therefore the only way to obtain its SFNE would be to resort to Theorem 2 in view of the extension to the infinite horizon as discussed in Sect. 3.6. The HJB equation for player j, written for the current-value function $\mathcal{V}_j(x) = e^{\rho t} V_j(t, x)$, is

$$\rho\mathcal{V}_j(x) = \max_{u_j}\left[u_j\left(\kappa - \frac{1}{2}u_j\right) - \frac{1}{2}\varphi x^2 + \frac{\partial}{\partial x}\mathcal{V}_j(x)\,(u_1 + u_2 - \alpha x)\right]. \qquad (2.40)$$

Being strictly concave in u_j, the RHS of (2.40) admits a unique maximum, with the maximizing solution being

$$u_j(x) = \kappa + \frac{\partial}{\partial x}\mathcal{V}_j(x). \qquad (2.41)$$

Given the symmetric nature of this game, we focus on symmetric equilibrium strategies. Taking into account the linear-quadratic specification of the differential game, we make the informed guess that the current-value function is quadratic (because the game is symmetric, and we focus on symmetric solutions, the value function is the same for both players), given by

$$\mathcal{V}_j(x) = \frac{a}{2}x^2 + bx + c, \quad j = 1, 2,$$

where a, b, c are parameters yet to be determined. Using (2.41) then leads to $u_j(x) = \kappa + ax + b$. Substituting this into the RHS of (2.40), we obtain

$$\frac{1}{2}(3a^2 - 2a\alpha - \varphi)x^2 + (3ab - b\alpha + 2a\kappa)x + \frac{1}{2}(3b^2 + 4b\kappa + \kappa^2).$$

The LHS of (2.40) reads

$$\rho\left(\frac{a}{2}x^2 + bx + c\right),$$

and equating the coefficients of x^2, x and the constant term, we obtain three equations in the three unknowns, a, b, and c. Solving these equations, we get the following coefficients for the noncooperative value functions:

$$a = \frac{\rho + 2\alpha \pm \sqrt{(\rho + 2\alpha)^2 + 16\varphi}}{6},$$

$$b = \frac{-2a\kappa}{3a - (\rho + \alpha)},$$

$$c = \frac{\kappa^2 + 4b\kappa + 3b^2}{2\rho}.$$

Remark 5. The coefficient a is the root of a second-degree polynomial having two roots: one positive and one negative. The selection of the negative root

$$a = \frac{\rho + 2\alpha - \sqrt{(\rho + 2\alpha)^2 + 16\varphi}}{6},$$

guarantees the global stability of the state trajectory. The resulting noncooperative equilibrium state trajectory is given by

$$x^*(t) = \left[x^0 + \frac{2(\kappa + b)}{2a - \alpha}\right]e^{(2a-\alpha)t} - \frac{2(\kappa + b)}{2a - \alpha}.$$

The state dynamics of the game has a globally asymptotically stable steady state if $2a - \alpha < 0$. It can be shown that to guarantee this inequality and therefore global asymptotic stability, the only possibility is to choose $a < 0$.

3.8 Linear-Quadratic Differential Games (LQDGs)

We have seen in the previous subsection, within the context of a specific scalar differential game, that linear-quadratic structure (linear dynamics and quadratic payoff functions) enables explicit computation of both OLNE and SFNE strategies (for the infinite-horizon game). We now take this analysis a step further, and discuss the general class of linear-quadratic (LQ) games, but in finite horizon, and show that the LQ structure leads (using the necessary and sufficient conditions obtained earlier for OLNE and SFNE, respectively) to computationally feasible equilibrium strategies. Toward that end, we first make it precise in the definition that follows the class of LQ differential games under consideration (we in fact define a slightly larger class of DGs, namely, affine-quadratic DGs, where the state dynamics are driven by also a known exogenous input). Following the definition, we discuss characterization of the OLNE and SFNE strategies, in that order. Throughout, x' denotes the transpose of a vector x, and B' denotes the transpose of a matrix B.

Definition 6. An m-player differential game of fixed prescribed duration $[0, T]$ is of the *affine-quadratic* type if $U_j = \mathbb{R}^{p_j}$ $(j \in M)$ and

$$f(t,x,u) = A(t)x + \sum_{i \in M} B_i(t)u_i + c(t),$$

$$g_j(t,x,u) = -\frac{1}{2}\left(x'Q_j(t)x + \sum_{i \in M} u_i' R_j^i(t)u_i\right),$$

$$S_j(x) = -\frac{1}{2}x_j' Q_j^f x,$$

where $A(\cdot)$, $B_j(\cdot)$, $Q_j(\cdot)$, $R_j^i(\cdot)$ are matrices of appropriate dimensions, $c(\cdot)$ is an n-dimensional vector, all defined on $[0, T]$, and with continuous entries $(i, j \in M)$. Furthermore, Q_j^f, $Q_j(\cdot)$ are symmetric, $R_i^j(\cdot) > 0$ $(j \in M)$, and $R_j^j(\cdot) \geq 0$ $(i \neq j,\ i, j \in M)$.

An affine-quadratic differential game is of the *linear-quadratic* type if $c \equiv 0$.

3.8.1 OLNE

For the affine-quadratic differential game formulated above, let us further assume that $Q^i(\cdot) \geq 0$, $Q_f^i \geq 0$. Then, under the open-loop information structure, player j's payoff function $J_j([u_j, u_{-j}]; x^0, t^0 = 0)$, defined by (2.3), is a strictly concave function of $u_j(\cdot)$ for all permissible control functions $u_{-j}(\cdot)$ of the other players and for all $x^0 \in \mathbb{R}^n$. This then implies that the necessary conditions for OLNE derived in Sect. 3.3.1 are also sufficient, and every solution set of the first-order conditions provides an OLNE. Now, the Hamiltonian for player j is

$$H_j(x,u,\lambda_j,t) = -\frac{1}{2}\left(x'Q_jx + \sum_{i \in M} u_j' R_j^i u_i\right) + \lambda_j\left(Ax + c + \sum_{i \in M} B_i u_i\right),$$

whose maximization with respect to $u_j(t) \in \mathbb{R}^{p_j}$ yields the unique relation

$$u_j^*(t) = R_j^j(t)^{-1} B_j(t)'\lambda_j(t), \quad j \in M. \tag{i}$$

Furthermore, the costate equations are

$$\dot{\lambda}_j = Q_j x^* - A'\lambda_j; \quad \lambda_j;(T) = -Q_j^f x(T) \quad (j \in M), \tag{ii}$$

and the optimal state trajectory is generated by

$$\dot{x}^* = Ax^* + c - \sum_{i \in M} B_i R_i^{i^{-1}} B_i' \lambda_i \,; \quad x^*(0) = x^0. \tag{iii}$$

This set of differential equations constitutes a two-point boundary value problem, the solution of which can be written, without any loss of generality, as $\{\lambda_j(t) = -K_j(t)x^*(t) - k_j(t),\ j \in M; x^*(t), t \in [0,T]\}$ where $K_j(\cdot)$ are $(n \times n)$-dimensional matrices and $k_j(\cdot)$ are n-dimensional vectors. Now, substituting $\lambda_j = -K_j x^* - k_j$ $(j \in M)$ into the costate equations (ii), we can arrive at the conclusion that K_j $(j \in M)$ and k_j $(j \in M)$ should then satisfy, respectively, the following two sets of matrix and vector differential equations:

$$\begin{gathered} \dot{K}_j + K_j A + A'K_j + Q_j^j - K_j \textstyle\sum_{i \in M} B_i R_i^{i^{-1}} B_i' K_i = 0; \\ K_j(T) = Q_j^f \quad (j \in M), \end{gathered} \tag{2.42}$$

and

$$\begin{gathered} \dot{k}_j + A'k_j + K_j c - K_j \textstyle\sum_{i \in M} B_i R_i^{i^{-1}} B_i' k_i = 0; \\ k_j(T) = 0 \quad (j \in M). \end{gathered} \tag{2.43}$$

The expressions for the OLNE strategies can then be obtained from (i) by substituting $\lambda_j = -K_j x^* - k_j$, and likewise the associated state trajectory for x^* follows from (iii).

The following theorem now captures this result (see, Başar and Olsder 1999, pp. 317–318).

Theorem 3. *For the m-player affine-quadratic differential game with $Q_j(\cdot) \geq 0$, $Q_j^f \geq 0$ $(j \in M)$, let there exist a unique solution set $\{K_j,\ j \in M\}$ to the coupled set of matrix Riccati differential equations (2.42). Then, the differential game admits a* unique OLNE solution *given by*

$$\gamma_j^*(x^0, t) \equiv u_j^*(t) = -R_i^i(t)^{-1} B_i'(t)[K_j(t)x^*(t) + k_j(t)] \quad (j \in M),$$

where $\{k_j(\cdot),\ j \in M\}$ solve uniquely the set of linear differential equations (2.43) and $x^*(\cdot)$ denotes the corresponding OLNE state trajectory, generated by (iii), which can be written as

$$x^*(t) = \Phi(t,0)x_0 + \int_0^t \Phi(t,\sigma)\eta(\sigma)\,\mathrm{d}\sigma,$$
$$\frac{\mathrm{d}}{\mathrm{d}t}\Phi(t,\sigma) = F(t)\Phi(t,\sigma); \quad \Phi(\sigma,\sigma) = I,$$
$$F(t) := A - \textstyle\sum_{i\in M} B_i {R_i^i}^{-1} B_i' K_i(t),$$
$$\eta(t) := c(t) - \textstyle\sum_{i\in M} B_i {R_i^i}^{-1} B_i' k_i(t).$$

Remark 6 (Nonexistence and multiplicity of OLNE). Note that the existence of OLNE for the affine-quadratic differential game hinges on the existence of a solution to the set of coupled matrix Riccati equations (2.42), since the second differential equation (2.43) always admits a solution, being linear in k_i's. Further, the OLNE is unique, whenever the matrix solution to (2.42) is unique. It is within the realm of possibility, however, that an OLNE may not exist, just as a Nash equilibrium may not exist in static quadratic games (reaction planes may not have a common point of intersection) or there may be multiple OLNEs (using the earlier analogy to static games, reaction planes may have more than one point of intersection). Note also that even for the LQDG (i.e., when $c \equiv 0$), when we have $k_j \equiv 0,\ j \in M$, still the same possibilities (of nonexistence or multiplicity of OLNE) are valid.

An important point to note regarding the OLNE in Theorem 3 above is that the solution does not depend on all the parameters that define the affine-quadratic differential game, particularly the matrices $\{R_j^i,\ i \neq j,\ i,j = 1,2\}$. Hence, the OLNE would be the same if g_j were replaced by

$$\tilde{g}_j(t,x,u_j) = -\frac{1}{2}\left(x'Q_j(t)x + u_j' R_j^j(t)u_j\right).$$

This is in fact not surprising in view of our earlier discussion in Sect. 3.7 on strategic equivalence. Under open-loop information structure, adding to $g_j(t,x,u)$ of one game any function of u_{-j} generates another game that is strategically equivalent to the first one and hence has the same set of OLNE strategies, and in this particular case, adding the term $(1/2)\sum_{i\neq j} u_i' R_j^i(t)u_i$ to g_j generates $\tilde{g}_j$. We can now go a step further, and subtract the term $(1/2)\sum_{i\neq j} u_i' R_i^i(t)u_i$ from $\tilde{g}_j$, and assuming also that the state weighting matrices $Q_j(\cdot)$ and Q_j^f are the same across all players (i.e., $Q(\cdot)$ and Q^f, respectively), we arrive at a single-objective function for all players (where we suppress dependence on t in the weighting matrices):

$$J_j(u(\cdot);x^0) =: J(u(\cdot);x^0) = -\frac{1}{2}\left(\int_0^T (x'Qx + \sum_{i\in M} u_i' R_i^i u_i)\,dt + x(T)'Q^f x(T)\right). \tag{2.44}$$

Hence, the affine-quadratic differential game where $Q_j(\cdot)$ and Q_j^f are the same across all players is strategically equivalent to a team problem, which, being deterministic, is in fact an optimal control problem. Letting $u := ({u_1}', \ldots, u_m')'$,

$B := (B_1, \ldots, B_m)$, and $R = \text{diag}(R_1^1, \ldots, R_m^m)$, this affine-quadratic optimal control problem has state dynamics

$$\dot{x} = A(t)x + B(t)u(t) + c(t),\ x(0) = x^0,$$

and payoff function

$$J(u(\cdot); x^0) = -\frac{1}{2}\left(\int_0^T (x'Q(t)x + u'R(t)u)\,dt + x(T)'Q^f x(T)\right),$$

where $R(\cdot) > 0$. Being strictly concave (and affine-quadratic), this optimal control problem admits a unique globally optimal solution, given by[7]

$$u^*(t) = -R(t)^{-1}B(t)'[K(t)x^*(t) + k(t)],\ t \geq 0,$$

where $K(\cdot)$ is the unique nonnegative-definite solution of the matrix Riccati equation:

$$\dot{K} + KA + A'K + Q - KBR^{-1}B'K = 0,\ K(T) = Q^f, \tag{2.45}$$

$k(\cdot)$ uniquely solves

$$\dot{k} + A'k + Kc - KBR^{-1}B'k = 0, k(T) = 0, \tag{2.46}$$

and $x^*(\cdot)$ is generated by (3), with

$$F(t) = A - BR^{-1}B'K(t), \quad \eta(t) = c(t) - BR^{-1}B'k(t), t \geq 0.$$

Note that for each block component of u, the optimal control can be written as

$$\gamma_j^*(x^0, t) = u_j^*(t) = -R_j^j(t)^{-1}B_j(t)'[K(t)x^*(t) + k(t)],\ t \geq 0, j \in M, \tag{2.47}$$

which by strategic equivalence is the unique OLNE. The following corollary to Theorem 3 summarizes this result.

Corollary 1. *The special class of affine-quadratic differential games with open-loop information structure, where in Definition 6, $Q_j = Q \geq 0\,\forall j \in M$ and $Q_j^f = Q^f \geq 0\,\forall j \in M$, is strategically equivalent to a strictly concave optimal control problem and admits a unique OLNE, given by (2.47), where $K(\cdot)$ is the unique nonnegative-definite solution of (2.45), $k(\cdot)$ uniquely solves (2.46), and $x^*(\cdot)$ is the unique OLNE state trajectory as defined above.*

[7]This is a standard result in optimal control, which can be found in any standard text, such as Bryson et al. (1975).

Remark 7 (Strategic equivalence and symmetry). A special class of affine-quadratic differential games which fits into the framework covered by Corollary 1 is the class of *symmetric differential games*, where the players are indistinguishable (with B_j, Q_j, Q_j^f, R_j^j being the same across all players, that is, index j free). Hence, symmetric affine-quadratic differential games, with $Q_j = Q \geq 0 \, \forall j \in M$, $Q_j^f = Q^f \geq 0 \, \forall j \in M$, $R_j^j = \bar{R} > 0 \, \forall j \in M$, and $B_j = \bar{B} \, \forall j \in M$, admit a unique OLNE:

$$\gamma_j^*(x^0, t) = u_j^*(t) = -\bar{R}(t)^{-1}\bar{B}(t)'[K(t)x^*(t) + k(t)], \; t \geq 0, j \in M,$$

where $K(\cdot)$ and $k(\cdot)$ uniquely solve

$$\dot{K} + KA + A'K + Q - mK\bar{B}\bar{R}^{-1}\bar{B}'K = 0, \; K(T) = Q^f, \tag{2.48}$$

and

$$\dot{k} + A'k + Kc - mK\bar{B}\bar{R}^{-1}\bar{B}'k = 0, \; k(T) = 0,$$

and $x^*(\cdot)$ is as defined before.

Remark 8 (Zero-sum differential games). A special class of nonzero-sum differential games is zero-sum differential games, where in the general framework, $m = 2$ and $J_2 \equiv -J_1 =: J$. The (two) players in this case have totally opposing objectives, and hence what one would be minimizing, the other one would be maximizing. Nash equilibrium in this case corresponds to the *saddle-point equilibrium*, and if (γ_1^*, γ_2^*) is one such pair of strategies, with player 1 as minimizer (of J) and player 2 as maximizer, they satisfy the pair of saddle-point inequalities:

$$\bar{J}(\gamma_1^*, \gamma_2) \leq \bar{J}(\gamma_1^*, \gamma_2^*) \leq \bar{J}(\gamma_1, \gamma_2^*), \;\; \forall \gamma_j \in \Gamma_j, j = 1, 2. \tag{2.49}$$

Affine-quadratic zero-sum differential games are defined as in Definition 6, with $m = 2$ and (suppressing dependence on the time variable t)

$$\begin{aligned} g_2 \equiv -g_1 &=: g(x, u_1, u_2, t) \\ &= \frac{1}{2}(x'Qx + u_1'R_1u_1 - u_2'R_2u_2), \;\; Q \geq 0, R_i > 0, i = 1, 2, \end{aligned}$$

$$S_2(x) \equiv -S_1(x) =: S(x) = \frac{1}{2}x'Q^f x, \;\; Q^f \geq 0.$$

Note, however, that this formulation cannot be viewed as a special case of two-player affine-quadratic nonzero-sum differential games with nonpositive-definite weighting on the states and negative-definite weighting on the controls of the players in their payoff functions (which makes the payoff functions strictly concave in individual players' controls – making their individual maximization problems automatically well defined), because here maximizing player (player 2 in this case) has nonnegative-definite weighting on the state, which brings up the possibility of player 2's optimization problem to be unbounded. To make the game well defined, we have to assure that it is convex-concave. Convexity of J in u_1 is readily satisfied, but for concavity in u_2, we have to impose an additional condition. It turns out that (see Başar and Bernhard 1995; Başar and Olsder 1999) a practical way of checking strict concavity of J in u_2 is to assure that the following matrix Riccati differential equation has a continuously differentiable nonnegative-definite solution over the interval $[0, T]$, that is, there are no conjugate points:

$$\dot{\hat{S}} + \hat{S}A + A'\hat{S} + Q + \hat{S}B_2R_2^{-1}B_2'\hat{S} = 0, \;\; \hat{S}(T) = Q^f. \tag{2.50}$$

Then, one can show that the game admits a unique saddle-point solution in open-loop policies, which can be obtained directly from Theorem 3 by noticing that $K_2 = -K_1 =: \hat{K}$ and $k_2 = -k_1 =: \hat{k}$, which satisfy

$$\dot{\hat{K}} + \hat{K}A + A'\hat{K} + Q - \hat{K}(B_1R_1^{-1}B_1' - B_2R_2^{-1}B_2')\hat{K} = 0, \;\; \hat{K}(T) = Q^f, \tag{2.51}$$

and

$$\dot{\hat{k}} + A'\hat{k} + \hat{K}c - \hat{K}(B_1R_1^{-1}B_1' - B_2R_2^{-1}B_2')\hat{k} = 0, \;\; \hat{k}(T) = 0. \tag{2.52}$$

Under the condition of existence of a well-defined solution to (2.50), the matrix Riccati differential equation (2.51) admits a unique continuously differentiable nonnegative-definite solution, and the open-loop saddle-point (OLSP) strategies for the players, satisfying (2.49), are given by

$$\gamma_1^*(x^0, t) = -R_1^{-1}B_1'[\hat{K}(t)x^*(t) + \hat{k}(t)],$$
$$\gamma_2^*(x^0, t) = R_2^{-1}B_2'[\hat{K}(t)x^*(t) + \hat{k}(t)], \; t \geq 0,$$

where $x^*(\cdot)$ is the saddle-point state trajectory, generated by

$$\dot{x} = (A - (B_1'R_1^{-1}B_1' - B_2R_2^{-1}B_2')\hat{K})x - (B_1'R_1^{-1}B_1 - B_2'R_2^{-1}B_2)\hat{k} + c, \; x(0) = x^0.$$

Hence, the existence of an OLSP hinges on the existence of a nonnegative-definite solution to the matrix Riccati differential equation (2.50), which as indicated is related to the nonexistence of conjugate points in the interval $[0, T]$,[8] which in turn is related to whether the game in the infinite-dimensional function space (Hilbert

[8]The existence of a conjugate point in $[0, T)$ implies that there exists a sequence of policies by the maximizer which can drive the value of the game arbitrarily large, that is, the upper value of the game is infinite.

space in this case) is convex-concave or not, as mentioned earlier. For details, we refer to Başar and Bernhard (1995).

3.8.2 SFNE

We now turn to affine-quadratic differential games (cf. Definition 6) with state-feedback information structure. We have seen earlier (cf. Theorem 2) that the CLNE strategies can be obtained from the solution of coupled HJB partial differential equations. For affine-quadratic differential games, these equations can be solved explicitly, since their solutions admit a general quadratic (in x) structure, as we will see shortly. This also readily leads to a set of SFNE strategies which are expressible in closed form. The result is captured in the following theorem, which follows from Theorem 2 by using in the coupled HJB equations the structural specification of the affine-quadratic game (cf. Definition 6), testing the solution structure $V_j(t,x) = -\frac{1}{2}x'Z_j(t)x - x'\zeta_j(t) - n_j(t),\ j \in M$, showing consistency, and equating like powers of x to arrive at differential equations for Z_j, ζ_j, and n_j (see, Başar and Olsder 1999, pp. 323–324).

Theorem 4. *For the m-player affine-quadratic differential game introduced in Definition 6, with $Q_j(\cdot) \geq 0$, $Q_j^f \geq 0$ $(j \in M)$, let there exist a set of matrix-valued functions $Z_j(\cdot) \geq 0$, $j \in M$, satisfying the following m-coupled matrix Riccati differential equations:*

$$\begin{gathered}\dot{Z}_j + Z_j\tilde{F} + \tilde{F}'Z_j + \textstyle\sum_{i\in M} Z_i B_i R_i^{i^{-1}} R_j^i R_i^{i^{-1}} B_i' Z_i + Q_j = 0;\\ Z_j(T) = Q_j^f,\end{gathered} \tag{2.53}$$

where

$$\tilde{F}(t) := A(t) - \sum_{i\in M} B_i(t) R_i^i(t)^{-1} B_i(t)' Z_i(t). \tag{2.54}$$

Then, under the state-feedback information structure, the differential game admits an SFNE solution, affine in the current value of the state, given by

$$\gamma_j^*(x,t) = -R_j^j(t)^{-1} B_j(t)'[Z_j(t)x(t) + \zeta_j(t)], \quad j \in M, \tag{2.55}$$

where ζ_j $(j \in M)$ are obtained as the unique solution of the coupled linear differential equations

$$\dot{\zeta}_j + \tilde{F}'\zeta_j + \sum_{i\in M} Z_i B_i R_i^{i^{-1}} R_j^i R_i^{i^{-1}} B_i'\zeta_i + Z_j\beta = 0;\ \zeta_j(T) = 0, \tag{2.56}$$

with

$$\beta := c - \sum_{i \in M} B_i R_i^{i^{-1}} B_i' \zeta_i . \tag{2.57}$$

The corresponding values of the payoff functionals are

$$\bar{J}_j^* = V_j(x^0, 0) = -\frac{1}{2} x^{0\prime} Z_j(0) x^0 - x^{0\prime} \zeta_j(0) - n_j(0), \quad j \in M, \tag{2.58}$$

where $n_j(\cdot)$ $(j \in M)$ are obtained as unique continuously differentiable solutions of

$$\dot{n}_j + \beta' \zeta_j + \frac{1}{2} \sum_{i \in M} \zeta_i' B_i R_i^{i^{-1}} R_j^i R_i^{i^{-1}} B_i' \zeta_i = 0; \; n_j(T) = 0. \tag{2.59}$$

Remark 9. Note that the "nonnegative-definiteness" requirement imposed on $Z_j(\cdot)$ is a consequence of the fact that $V_j(x,t) \geq 0 \; \forall x \in \mathbb{R}^n, t \in [0,T]$, this latter feature being due to the eigenvalue restrictions imposed a priori on $Q_j(\cdot)$, Q_j^f, and $R_j^i(\cdot)$, $i, j \in M$. Finally, the corresponding "Nash" values for the payoff functionals follow from the fact that $V_j(x,t)$ is the value function for player j at SFNE, at any point (x,t). We also note that Theorem 4 provides only one set of SFNE strategies for the affine-quadratic game under consideration, and it does not attribute any uniqueness feature to this solution set. What can be shown, however, is the uniqueness of SFNE when the players are restricted at the outset to affine memoryless state-feedback strategies (Başar and Olsder 1999).

Remark 10. The result above extends readily to more general affine-quadratic differential games where the payoff functions of the players contain additional terms that are linear in x, that is, with g_j and S_j in Definition 6 extended, respectively, to

$$g_j = -\frac{1}{2}\left(x'[Q_j(t)x + 2l_j(t)] + \sum_{i \in M} u_j' R_i^j u_i \right); \quad S_j(x) = -\frac{1}{2} x'[Q_j^f x + 2l_j^f],$$

where $l_j(\cdot)$ is a known n-dimensional vector-valued function, continuous on $[0,T]$, and l_j^f is a fixed n-dimensional vector, for each $j \in M$. Then, the statement of Theorem 4 remains intact, with only the differential equation (2.56) that generates $\zeta_j(\cdot)$ now reading:

$$\dot{\zeta}_j + \tilde{F}' \zeta_j + \sum_{i \in M} Z_i B_i R_i^{i^{-1}} R_j^i R_i^{i^{-1}} B_i' \zeta_i + Z_j \beta + l_j = 0; \; \zeta_j(T) = l_j^f .$$

When comparing SFNE with the OLNE, one question that comes up is whether there is the counterpart of Corollary 1 in the case of SFNE. The answer is no, because adding additional terms to g_j that involve controls of other players generally leads to a different optimization problem faced by player j, since u_i's for $i \neq j$ depend on x and through it on u_j. Hence, in general, a differential game with state-feedback information structure cannot be made strategically equivalent to a team (and hence optimal control) problem. One can, however, address the issue of simplification of the set of sufficient conditions (particularly the coupled matrix Riccati differential equations (2.53)) when the game is symmetric. Let us use the same setting as in Remark 7, but also introducing a common notation $\hat{R}$ for the weighting matrices $R^i_j, i \neq j$ for the controls of the other players appearing in player j's payoff function, and for all $j \in M$ (note that this was not an issue in the case of open-loop information since R^i_j's, $i \neq j$ were not relevant to the OLNE), and focusing on symmetric SFNE, we can now rewrite the SFNE strategy (2.55) for player j as

$$\gamma_j^*(x,t) = -\bar{R}(t)^{-1}\bar{B}(t)'[Z(t)x(t) + \zeta(t)], \quad j \in M, \tag{2.60}$$

where $Z(\cdot) \geq 0$ solves

$$\dot{Z} + Z\tilde{F} + \tilde{F}'Z + Z\bar{B}\bar{R}^{-1}\bar{B}'Z + \sum_{i \neq j, i \in M} Z\bar{B}\bar{R}^{-1}\hat{R}\bar{R}^{-1}\bar{B}'^f, \tag{2.61}$$

with (from (2.54))

$$\tilde{F} := A(t) - m\bar{B}(t)\bar{R}(t)^{-1}\bar{B}(t)'Z(t). \tag{2.62}$$

Substituting this expression for $\tilde{F}$ into (2.61), we arrive at the following alternative (more revealing) representation:

$$\dot{Z} + ZA + A'Z - Z\bar{B}\bar{R}^{-1}\bar{B}'Z + Q - (m-1)Z\bar{B}\bar{R}^{-1}[2\bar{R} - \hat{R}]\bar{R}^{-1}\bar{B}'Z$$
$$= 0; \quad Z(T) = Q^f. \tag{2.63}$$

Using the resemblance to the matrix Riccati differential equation that arises in standard optimal control (compare it with the differential equation (2.48) for K in Remark 7), we can conclude that (2.63) admits a unique continuously differentiable nonnegative-definite solution whenever the condition

$$2\bar{R} - \hat{R} > 0, \tag{2.64}$$

holds. This condition can be interpreted as players placing relatively more weight on their self-controls (in their payoff functions) than on each of the other individual players. In fact, if the weights are equal, then $\bar{R} = \hat{R}$, and (2.63) becomes equivalent to (2.48); this is of course not surprising since a symmetric game with $\bar{R} = \hat{R}$ is essentially an optimal control problem (players have identical payoff functions), for which OL and SF solutions have the same underlying Riccati differential equations.

Now, to complete the characterization of the SFNE for the symmetric differential game, we have to write down the differential equation for ζ_j in Theorem 4, that is (2.56), using the specifications imposed by symmetry. Naturally, it now becomes independent of the index j and can be simplified to the form below:

$$\dot{\zeta} + \left[A' - Z\bar{B}\bar{R}^{-1}[(2m-1)\bar{R} - (m-1)\hat{R}]\bar{R}^{-1}\bar{B}'\right]\zeta + Zc = 0;\ \zeta(T) = 0. \tag{2.65}$$

We following corollary to Theorem 4 now captures the main points of the discussion above.

Corollary 2. *For the symmetric affine-quadratic differential game introduced above, let the matrix Riccati differential equation* (2.63) *admit a unique continuously differentiable nonnegative-definite solution* $Z(\cdot)$. *Then the game admits a CLNE solution, which is symmetric across all players, and given by*

$$\gamma_j^*(x,t) = -\bar{R}(t)^{-1}\bar{B}(t)'[Z(t)x(t) + \zeta(t)],\ t \geq 0, j \in M,$$

where $\zeta(\cdot)$ *is generated uniquely by* (2.65). *If, furthermore, the condition* (2.64) *holds, then* $Z(\cdot)$ *exists and is unique.*

Remark 11 (Zero-sum differential games with SF information). The counterpart of Remark 8 on saddle-point equilibrium can also be derived under state-feedback information structure, this time specializing Theorem 4 to the two-player zero-sum differential game. Using the same setting as in Remark 8, it follows by inspection from Theorem 4 that $Z_2 = -Z_1 =: \hat{Z}$ and $\zeta_2 = \zeta_1 =: \hat{\zeta}$, where the differential equations satisfied by $\hat{Z}$ and $\hat{\zeta}$ are precisely the ones satisfied by $\hat{K}$ and $\hat{k}$ in the OL case, that is, (2.51) and (2.52), respectively. Under the condition of the existence of well-defined (unique continuously differentiable nonnegative-definite) solution to the matrix Riccati differential equation (2.51), the state-feedback saddle-point (SFSP) strategies for the players, satisfying (2.49), are given by (directly from Theorem 4)

$$\gamma_1^*(x,t) = -R_1^{-1}B_1'[\hat{K}(t)x(t)+\hat{k}(t)],\ \gamma_2^*(x,t) = R_2^{-1}B_2'[\hat{K}(t)x(t)+\hat{k}(t)],\ t \geq 0.$$

Note that these are in the same form as the OLSP strategies, with the difference being that they are now functions of the *actual* current value of the state instead of the *computed* value (as in the OLSP case). Another difference between the OLSP and SFSP is that the latter does not require an a priori concavity condition to be imposed, and hence whether there exists a solution to (2.50) is irrelevant under state-feedback information; this condition is replaced by the existence of a solution to (2.51), which is less restrictive Başar and Bernhard (1995). Finally, since the forms of the OLSP and SFSP strategies are the same, they generate the same state trajectory (and hence lead to the same value for J), provided that the corresponding existence conditions are satisfied.

4 Stackelberg Equilibria

In the previous sections, the assumption was that the players select their strategies simultaneously, without any communication. Consider now a different scenario: a two-player game where one player, the leader, makes her decision before the other player, the follower.[9] Such a sequence of moves was first introduced by von Stackelberg in the context of a duopoly output game; see, von Stackelberg (1934).

Denote by L the leader and by F the follower. Suppose that $u_L(t)$ and $u_F(t)$ are, respectively, the control vectors of L and F. The control constraints $u_L(t) \in U_L$ and $u_F(t) \in U_F$ must be satisfied for all t. The state dynamics and the payoff functionals are given as before by (2.1), (2.2), and (2.3), where we take the initial time to be $t^0 = 0$, without any loss of generality. As with the Nash equilibrium, we will define an open-loop Stackelberg equilibrium (OLSE). We will also introduce what is called feedback (or Markovian)-Stackelberg equilibrium (FSE), which uses state feedback information and provides the leader only time-incremental lead advantage.

4.1 Open-Loop Stackelberg Equilibria (OLSE)

When both players use open-loop strategies, μ_L and μ_F, their control paths are determined by $u_L(t) = \mu_L(x^0, t)$ and $u_F(t) = \mu_F(x^0, t)$, respectively. Here μ_j denotes the open-loop strategy of player j.

The game proceeds as follows. At time $t = 0$, the leader announces her control path $u_L(\cdot)$ for $t \in [0, T]$. Suppose, for the moment, that the follower believes in this announcement. The best she can do is then to select her own control path $u_F(\cdot)$ to maximize the objective functional

$$J_F = \int_0^T g_F\,(x(t), u_L(t),\ u_F(t), t)\,dt + S_F(x(T)), \tag{2.66}$$

subject to the state dynamics

$$\dot{x}(t) = f\,(x(t), u_L(t), u_F(t), t) \quad x(0) = x^0, \tag{2.67}$$

and the control constraint

$$u_F(t) \in U_F. \tag{2.68}$$

[9]The setup can be easily extended to the case of several followers. A standard assumption is then that the followers play a (Nash) simultaneous-move game vis-a-vis each other, and a sequential game vis-a-vis the leader (Başar and Olsder 1999).

This is a standard optimal control problem. To solve it, introduce the follower's Hamiltonian

$$\begin{aligned}
&H_F(x(t), \lambda_F(t), u_F(t), u_L(t), t) \\
&\qquad = g_F(x(t), u_F(t), u_L(t), t) + \lambda_F f(x(t),\ u_F(t), u_L(t), t),
\end{aligned}$$

where the adjoint variable $\lambda_F = \lambda_F(t)$ is an n-vector. Suppose that the Hamiltonian H_F is strictly concave in $u_F \in U_F$, where U_F is a convex set. Then the maximization of H_F with respect to u_F, for $t \in [0, T]$, uniquely determines $u_F(t)$ as a function of t, x, u_L, and λ_F, which we write as

$$u_F(t) = R(x(t), t, u_L(t), \lambda_F(t)). \tag{2.69}$$

This defines the follower's *best reply (response)* to the leader's announced time path $u_L(\cdot)$.

The follower's costate equations and their boundary conditions in this maximization problem are given by

$$\begin{aligned}
\dot{\lambda}_F(t) &= -\frac{\partial}{\partial x} H_F, \\
\lambda_F(T) &= \frac{\partial}{\partial x} S_j\left(x(T)\right).
\end{aligned}$$

Substituting the best response function R into the state and costate equations yields a two-point boundary-value problem. The solution of this problem, $(x(t), \lambda_F(t))$, can be inserted into the function R. This represents the follower's optimal behavior, given the leader's announced time path $u_L(\cdot)$.

The leader can replicate the follower's arguments. This means that, since she knows everything the follower does, the leader can calculate the follower's best reply R to any $u_L(\cdot)$ that she may announce. The leader's problem is then to select a control path $u_L(\cdot)$ that maximizes her payoff given F's response, that is, maximization of

$$J_L = \int_0^T g_L\left(x(t),\ u_L(t), R\left(x(t), t, u_L(t), \lambda_F(t)\right), t\right) dt + S_L(x(T)), \tag{2.70}$$

subject to

$$\begin{aligned}
\dot{x}(t) &= f(x(t), u_L(t), R(x(t), t, u_L(t), \lambda_F(t)), t), \quad x(0) = x^0, \\
\dot{\lambda}_F(t) &= -\frac{\partial}{\partial x} H_F\left(x(t),\ u_L(t), R\left(x(t), t, u_L(t), \lambda_F(t)\right)\right), t), \\
\lambda_F(T) &= \frac{\partial}{\partial x} S_F\left(x(T)\right),
\end{aligned}$$

and the control constraint

$$u_L(t) \in U_L.$$

Note that the leader's dynamics include two state equations, one governing the evolution of the original state variables x and a second one accounting for the evolution of λ_F, the adjoint variables of the follower, which are now treated as state variables. Again, we have an optimal control problem that can be solved using the maximum principle. To do so, we introduce the leader's Hamiltonian

$$\begin{aligned} H_L \left(x(t), u_L(t), R\left(x(t), t, \ u_L(t), \lambda_F(t)\right), \lambda_L(t), \theta(t)\right) \\ = g_L \left(x(t), u_L(t), R\left(x(t), t, u_L(t), \lambda_F(t)\right), t\right) \\ + \lambda_L(t) f\left(x(t), u_L(t), R(x(t), t, u_L(t), \lambda_F(t)), t\right) \\ + \theta(t) \left(-\frac{\partial}{\partial x} H_F\left(x(t), \lambda_F(t), R\left(x(t), t, u_L(t), \lambda_F(t)\right), u_L(t), t\right)\right), \end{aligned}$$

where $\lambda_L = \lambda_L(t)$ is the n-vector of costate variables appended to the state equation for $x(t)$, with the boundary conditions

$$\lambda_L(T) = \frac{\partial}{\partial x} S_L\left(x(T)\right),$$

and $\theta = \theta(t)$ is the vector of n costate variables appended to the state equation for $\lambda_F(t)$, satisfying the initial condition

$$\theta(0) = 0.$$

This initial condition is a consequence of the fact that $\lambda_F(0)$ is "free," i.e., unrestricted, being free of any soft constraint in the payoff function, as opposed to $x(T)$ which enters a terminal reward term. The following theorem now collects all this for the OLSE (see, Başar and Olsder 1999, pp. 409–410).

Theorem 5. *For the two-player open-loop Stackelberg differential game formulated in this subsection, let $u_L^*(t) = \mu_L^*(x^0, t)$ be the leader's open-loop equilibrium strategy and $u_F^*(t) = \mu_F^*(x^0, t)$ be the follower's. Let the solution to the follower's optimization problem of maximizing J_F given by (2.66) subject to the state equation (2.67) and control constraint (2.68) exist and be uniquely given by (2.69). Then,*

(i) *The leader's open-loop Stackelberg strategy μ_L^* maximizes (2.70) subject to the given control constraint and the $2n$-dimensional differential equation system for x and λ_F (given after (2.70)) with mixed boundary specifications.*

(ii) *The follower's open-loop Stackelberg strategy* μ_F^* *is* (2.69) *with* u_L *replaced by* u_L^*.

Remark 12. In the light of the discussion in this subsection leading to Theorem 5, it is possible to write down a set of necessary conditions (based on the maximum principle) which can be used to solve for L's open-loop strategy μ_L^*. Note that, as mentioned before, in this maximization problem in addition to the standard state (differential) equation with specified initial conditions, we also have the costate differential equation with specified terminal conditions, and hence the dynamic constraint for the maximization problem involves a $2n$-dimensional differential equation with mixed boundary conditions (see the equations for x and λ_F following (2.70)). The associated Hamiltonian is then H_L, defined prior to Theorem 5, which has as its arguments two adjoint variables, λ_L and θ, corresponding to the differential equation evolutions for x and λ_F, respectively. Hence, from the maximum principle, these new adjoint variables satisfy the differential equations:

$$\begin{aligned}
\dot{\lambda}_L(t) &= -\frac{\partial}{\partial x} H_L\left(x(t),\ u_L(t), R\left(x(t), t, u_L(t), \lambda_F(t)\right), \lambda_L(t), \theta(t), t\right), \\
\lambda_L(T) &= \frac{\partial}{\partial x} S_L\left(x(T)\right), \\
\dot{\theta}(t) &= -\frac{\partial}{\partial \lambda_F} H_L\left(x(t),\ u_L(t), R\left(x(t), t, u_L(t), \lambda_F(t)\right), \lambda_L(t), \theta(t), t\right), \\
\theta(0) &= 0.
\end{aligned}$$

Finally, $u_L^*(t) = \mu_L^*(x^0, t)$ is obtained from the maximization of the Hamiltonian H_L (where we suppress dependence on t):

$$u_L^* = \arg \max_{u_L \in U_L} H_L(x, u_L, R(x, u_L, \lambda_F), \lambda_L, \theta).$$

4.2 Feedback Stackelberg Equilibria (FSE)

We now endow both players with state-feedback information, as was done in the case of SFNE, which is a memoryless information structure, not allowing the players to recall even the initial value of the state, x^0, except at $t = 0$. In the case of Nash equilibrium, this led to a meaningful solution, which also had the appealing feature of being subgame perfect and strongly time consistent. We will see in this subsection that this appealing feature does not carry over to Stackelberg equilibrium when the leader announces her strategy in advance for the entire duration of the game, and in fact the differential game becomes ill posed. This will force us to introduce, again under the state-feedback information structure, a different concept of Stackelberg equilibrium, called *feedback Stackelberg*, where the strong time consistency is imposed at the outset. This will then lead to a derivation that parallels the one for SFNE.

Let us first address "ill-posedness" of the classical Stackelberg solution when the players use state-feedback information, in which case their strategies are mappings from $\mathbb{R}^n \times [0, T]$ where the state-time pair (x, t) maps into U_L and U_F, for L and F, respectively. Let us denote these strategies by $\gamma_L \in \Gamma_L$ and $\gamma_F \in \Gamma_F$, respectively. Hence, the realizations of these strategies lead to the control actions (or control paths): $u_L(t) = \gamma_L(x, t)$ and $u_F(t) = \gamma_F(x, t)$, for L and F, respectively. Now, in line with the OLSE we discussed in the previous subsection, under the Stackelberg equilibrium, the leader L announces at time zero her strategy $\gamma(x, t)$ and commits to using this strategy throughout the duration of the game. Then the follower F reacts rationally to L's announcement, by maximizing her payoff function. Anticipating this, the leader selects a strategy that maximizes her payoff functional subject to the constraint imposed by the best response of F.

First let us look at the follower's optimal control problem. Using the dynamic programming approach, we have the Hamilton-Jacobi-Bellman (HJB) equation characterizing F's best response to an announced $\gamma_L \in \Gamma_L$:

$$
\begin{aligned}
-\frac{\partial}{\partial t} V_F(x, t) = \max_{u_F \in U_F} \{ & g_F(x, u_F(t), \gamma_L(x, t), t) \\
& + \frac{\partial}{\partial x} V_F(x, t) f(x(t), u_F(t), \gamma_L(x, t), t) \Big\},
\end{aligned}
$$

where V_F is the value function of F, which has the terminal condition $V_F(x, T) = S_F(x(T))$. Note that, for each fixed $\gamma_L \in \Gamma_L$, the maximizing control for F on the RHS of the HJB equation above is a function of the current time and state and hence is an element of Γ_F. Thus, F's maximization problem and its solution are compatible with the state-feedback information structure, and hence we have a well-defined problem at this stage. The dependence of this best response on γ_L, however, will be quite complex (much more than in the open-loop case), since what we have is a functional dependence in an infinite-dimensional space. Nevertheless, at least formally, we can write down this relationship as a best reaction function, $\tilde{R} : \Gamma_L \rightarrow \Gamma_F$, for the follower:

$$
\gamma_F = \tilde{R}(\gamma_L). \tag{2.71}
$$

Now, L can make this computation too, and according to the Stackelberg equilibrium concept, which is also called *global Stackelberg solution* (see, Başar and Olsder 1999), she has to maximize her payoff under the constraints imposed by this reaction function and the state dynamics that is formally

$$
\max_{\gamma_L \in \Gamma_L} J_L(\gamma_L, \tilde{R}(\gamma_L))
$$

Leaving aside the complexity of this optimization problem (which is not an optimal control problem of the standard type because of the presence of the reaction function which depends on the entire strategy of L over the full- time interval of the game), we note that this optimization problem is ill posed since for each choice of $\gamma_L \in \Gamma_L$, $J_L(\gamma_L, \tilde{R}(\gamma_L))$ is not a real number but generally a function of the initial state x^0, which is not available to L; hence, what we have is a multi-objective optimization problem, and not a single-objective one, which makes the differential game with the standard (global) Stackelberg equilibrium concept ill posed. One way around this difficulty would be to allow the leader (as well as the follower) recall the initial state (and hence modify their information sets to $\nu(x(t), x^0, t)$) or even have full memory on the state (in which case, ν is $\nu(x(s), s \leq t; t)$), which would make the game well posed, but requiring a different set of tools to obtain the solution (see, e.g., Başar and Olsder 1980; Başar and Selbuz 1979 and Chap. 7 of Başar and Olsder 1999), which also has connections to incentive designs and inducement of collusive behavior, further discussed in the next section of this chapter. We should also note that including x^0 in the information set also makes it possible to obtain the global Stackelberg equilibrium under mixed information sets, with F's information being inferior to that of L, such as $\nu_L(x(t), x^0, t)$ for L and $\nu(x^0, t)$ for F. Such a differential game would also be well defined.

Another way to resolve the ill-posedness of the global Stackelberg solution under state-feedback information structure is to give the leader only a *stagewise* (in the discrete-time context) first-mover advantage; in continuous time, this translates into an instantaneous advantage at each time t (Başar and Haurie 1984). This pointwise (in time) advantage leads to what is called a *feedback Stackelberg equilibrium (FSE)*, which is also strongly time consistent (Başar and Olsder 1999). The characterization of such an equilibrium for $j \in \{L, F\}$ involves the HJB equations

$$\left\{-\frac{\partial}{\partial t} V_j(x,t)\right\}_{j=1,2} = \mathrm{Sta}\left\{ g_j(x, [u_F, u_L]) + \frac{\partial}{\partial x} V_j(x,t) f(x, [u_F, u_L], t)\right\}_{j=1,2}, \tag{2.72}$$

where the "Sta" operator on the RHS solves, for each (x, t), for the Stackelberg equilibrium solution of the static two-player game in braces, with player 1 as leader and player 2 as follower. More precisely, the pointwise (in time) best response of F to $\gamma_L \in \Gamma_L$ is

$$\hat{R}(x, t; \gamma_L(x,t)) = \arg\max_{u_F \in U_F} \left\{ g_F(x, [u_F, \gamma_L(x,t)], t) + \frac{\partial}{\partial x} V_F(x,t) f(x, [u_F, \gamma_L(x,t)], t)\right\},$$

and taking this into account, L solves, again pointwise in time, the maximization problem:

$$\max_{u_L \in U_L} \left\{ g_L(x), ([\hat{R}(x,t;u_L), u_L], t) + \frac{\partial}{\partial x} V_L(x,t) f(x, [\hat{R}(x,t;u_L), u_L], t) \right\}.$$

Denoting the solution to this maximization problem by $u_L = \hat{\gamma}_L(x,t)$, an FSE for the game is then the pair of state-feedback strategies:

$$\left(\hat{\gamma}_L(x,t),\ \hat{\gamma}_F(x,t) = \hat{R}(x,t;\hat{\gamma}_L(x,t)) \right\} \tag{2.73}$$

Of course, following the lines we have outlined above, it should be obvious that explicit derivation of this pair of strategies depends on the construction of the value functions, V_L and V_F, satisfying the HJB equations (2.72). Hence, to complete the solution, one has to solve (2.72) for $V_L(x,t)$ and $V_F(x,t)$ and use these functions in (2.73). The main difficulty here is, of course, in obtaining explicit solutions to the HJB equations, which however can be done in some classes of games, such as those with linear dynamics and quadratic payoff functions (in which case V_L and V_F will be quadratic in x) (Başar and Olsder 1999). We provide some evidence of this solvability through numerical examples in the next subsection.

4.3 An Example: Construction of Stackelberg Equilibria

Consider the example of Sect. 3.7 but now with player 1 as the leader (from now on referred to as player L) and player 2 as the follower(player F). Recall that player j's optimization problem and the underlying state dynamics are

$$\max_{u_j} \left\{ J_j = \int_0^{\infty} e^{-\rho t} \left(u_j(t) \left(\kappa - \frac{1}{2} u_j(t) \right) - \frac{1}{2} \varphi x^2(t) \right) dt \right\},\ j = L, F, \tag{2.74}$$

$$\dot{x}(t) = u_L(t) + u_F(t) - \alpha x(t), \quad x(0) = x^0, \tag{2.75}$$

where φ and κ are positive parameters and $0 < \alpha < 1$. We again suppress the time argument henceforth when no ambiguity may arise. We discuss below both OLSE and FSE, but with horizon length infinite. This will give us an opportunity to introduce, in this context, also the infinite-horizon Stackelberg differential game.

4.3.1 Open-Loop Stackelberg Equilibrium (OLSE).

To obtain the best reply of the follower to the leader's announcement of the path $u_L(t)$, we introduce the Hamiltonian of player F:

$$H_F(x, u_L, u_F) = u_F \left(\kappa - \frac{1}{2} u_F \right) - \frac{1}{2} \varphi x^2 + q_F(u_L + u_F - \alpha x),$$

where q_F is the follower's costate variable associated with the state variable x. H_F being quadratic and strictly concave in u_F, it has a unique maximum:

$$u_F = \kappa + q_F, \tag{2.76}$$

where (from the maximum principle) q_F satisfies

$$\dot{q}_F = \rho q_F - \frac{\partial}{\partial x} H_F = (\rho + \alpha) q_F + \varphi x, \quad \lim_{t \to \infty} e^{-\rho t} q_F(t) = 0, \tag{2.77}$$

and with (2.76) used in the state equation, we have

$$\dot{x}(t) = u_L(t) + \kappa + q_F(t) - \alpha x(t), \quad x(0) = x^0. \tag{2.78}$$

Now, one approach here would be first to solve the two differential equations (2.77) and (2.78) and next to substitute the solutions in (2.76) to arrive at follower's best reply, i.e., $u_F(t) = R(x(t), u_L(t), q_F(t))$. Another approach would be to postpone the resolution of these differential equations and instead use them as dynamic constraints in the leader's optimization problem:

$$\begin{aligned}
&\max_{u_L} \left\{ J_L = \int_0^\infty e^{-\rho t} \left(u_L \left(\kappa - \frac{1}{2} u_L \right) - \frac{1}{2} \varphi x^2 \right) dt \right\} \\
&\dot{q}_F = (\rho + \alpha) q_F + \varphi x, \quad \lim_{t \to \infty} e^{-\rho t} q_F(t) = 0, \\
&\dot{x} = u_L + \kappa + q_F - \alpha x, \quad x(0) = x^0.
\end{aligned}$$

This is an optimal control problem with two state variables (q_F and x) and one control variable (u_L). Introduce the leader's Hamiltonian:

$$\begin{aligned}
H_L(x, u_L, q_F, q_L, \theta) = u_L \left(\kappa - \frac{1}{2} u_L \right) - \frac{1}{2} \varphi x^2 + \theta \left((\rho + \alpha) q_F + \varphi x \right) \\
+ q_L (u_L + \kappa + q_F - \alpha x),
\end{aligned}$$

where θ and q_L are adjoint variables associated with the two state equations in the leader's optimization problem. Being quadratic and strictly concave in u_L, H_L also admits a unique maximum, given by

$$u_L = \kappa + q_L,$$

and we have the state and adjoint equations:

$$\dot{\theta} = \rho \theta - \frac{\partial}{\partial q_F} H_L = -\theta \alpha - q_L, \quad \theta(0) = 0,$$

$$\dot{q}_L = \rho q_L - \frac{\partial}{\partial x} H_L = (\rho + \alpha) q_L + \varphi (x - \theta), \quad \lim_{t\to\infty} e^{-\rho t} q_L(t) = 0,$$
$$\dot{x} = \frac{\partial}{\partial q_L} H_L = u_L + \kappa + q_F - \alpha x, \quad x(0) = x^0,$$
$$\dot{q}_F = \frac{\partial}{\partial \theta} H_L = (\rho + \alpha) q_F + \varphi x, \quad \lim_{t\to\infty} e^{-\rho t} q_F(t) = 0.$$

Substituting the expression for u_L in the differential equation for x, we obtain a system of four differential equations, written in matrix form as follows:

$$\begin{pmatrix} \dot{\theta} \\ \dot{q}_L \\ \dot{x} \\ \dot{q}_F \end{pmatrix} = \begin{pmatrix} -\alpha & -1 & 0 & 0 \\ -\varphi & \rho + \alpha & \varphi & 0 \\ 0 & 1 & -\alpha & 1 \\ 0 & 0 & \varphi & \rho + \alpha \end{pmatrix} \begin{pmatrix} \theta \\ q_L \\ x \\ q_F \end{pmatrix} + \begin{pmatrix} 0 \\ 0 \\ 2\kappa \\ 0 \end{pmatrix}.$$

Solving the above system yields (θ, q_L, x, q_F). The last step would be to insert the solutions for q_F and q_L in the equilibrium conditions

$$u_F = \kappa + q_F, \quad u_L = \kappa + q_L,$$

to obtain the open-loop Stackelberg equilibrium controls u_L and u_F.

4.3.2 Feedback-Stackelberg Equilibrium (FSE).

To obtain the FSE, we first have to consider the infinite-horizon version of (2.72) and compute the best response of F to $u_L = \gamma_L(x)$. The maximization problem faced by F has the associated steady-state HJB equation for the current-value function $\mathcal{V}_F(x)$ (with the value function defined as $V_F(x,t) = e^{-\rho t}\mathcal{V}_F(x)$):

$$\rho \mathcal{V}_F(x) = \max_{u_F} \left[u_F \left(\kappa - \frac{1}{2} u_F \right) - \frac{1}{2} \varphi x^2 + \frac{\partial}{\partial x} \mathcal{V}_F(x) \left(u_L + u_F - \alpha x \right) \right]. \tag{2.79}$$

Maximization of the RHS yields (uniquely, because of strict concavity)

$$u_F = \kappa + \frac{\partial}{\partial x} \mathcal{V}_F(x). \tag{2.80}$$

Note that the above reaction function of the follower does not directly depend on the leader's control u_L, but only indirectly, through the state variable.

Accounting for the follower's response, the leader's HJB equation is

$$\rho \mathcal{V}_L(x) = \max_{u_L \geq 0} \left[u_L \left(\kappa - \frac{1}{2} u_L \right) - \frac{1}{2} \varphi x^2 + \frac{\partial}{\partial x} \mathcal{V}_L(x) \left(u_L + \kappa + \frac{\partial}{\partial x} \mathcal{V}_F(x) - \alpha x \right) \right], \tag{2.81}$$

where $\mathcal{V}_L(x)$ denotes the leader's current-value function. Maximizing the RHS yields

$$u_L = \kappa + \frac{\partial}{\partial x}\mathcal{V}_L(x).$$

Substituting in (2.81) leads to

$$\rho\mathcal{V}_L(x) = \left(\kappa + \frac{\partial}{\partial x}\mathcal{V}_L(x)\right)\left(\kappa - \frac{1}{2}\left(\kappa + \frac{\partial}{\partial x}\mathcal{V}_L(x)\right)\right) \tag{2.82}$$
$$-\frac{1}{2}\varphi x^2 + \frac{\partial}{\partial x}\mathcal{V}_L(x)\left(\frac{\partial}{\partial x}\mathcal{V}_L(x) + \frac{\partial}{\partial x}\mathcal{V}_F(x) + 2\kappa - \alpha x\right).$$

As the game at hand is of the linear-quadratic type, we can take the current value functions to be general quadratic. Accordingly, let

$$\mathcal{V}_L(x) = \frac{A_L}{2}x^2 + B_L x + C_L, \tag{2.83}$$

$$\mathcal{V}_F(x) = \frac{A_F}{2}x^2 + B_F x + C_F, \tag{2.84}$$

be, respectively, the leader's and the follower's current-value functions, where the six coefficients are yet to be determined. Substituting these structural forms in (2.82) yields

$$\rho\left(\frac{A_L}{2}x^2 + B_L x + C_L\right) = \frac{1}{2}\left(A_L^2 - \varphi + 2\left(A_F - \alpha\right)A_L\right)x^2$$
$$+ \left(A_L\left(B_L + B_F + 2\kappa\right) + \left(A_F - \alpha\right)B_L\right)x + \frac{1}{2}\left(\kappa^2 + B_L^2\right) + \left(B_F + 2\kappa\right)B_L.$$

Using (2.79), (2.80), and (2.83)–(2.84), we arrive at the following algebraic equation for the follower:

$$\rho\left(\frac{A_F}{2}x^2 + B_F x + C_F\right) = \frac{1}{2}\left(A_F^2 - \varphi + 2\left(A_L - \alpha\right)A_F\right)x^2$$
$$+ \left(A_F\left(B_F + B_L + 2\kappa\right) + \left(A_L - \alpha\right)B_F\right)x + \frac{1}{2}\left(\kappa^2 + B_F^2\right) + \left(B_L + 2\kappa\right)B_F.$$

By comparing the coefficients of like powers of x, we arrive at the following six-equation, nonlinear algebraic system:

$$0 = A_L^2 + \left(2A_F - 2\alpha - \rho\right)A_L - \varphi,$$
$$0 = A_L\left(B_L + B_F + 2\kappa\right) + \left(A_F - \alpha - \rho\right)B_L,$$

$$0 = \frac{1}{2}\left(\kappa^2 + B_L^2\right) + (B_F + 2\kappa) B_L - \rho C_L,$$
$$0 = A_F^2 - \varphi + (2A_L - 2\alpha - \rho) A_F,$$
$$0 = A_F (B_F + B_L + 2\kappa) + (A_L - \alpha - \rho) B_F,$$
$$0 = \frac{1}{2}\left(\kappa^2 + B_F^2\right) + (B_L + 2\kappa) B_F - \rho C_F.$$

The above system generally admits multiple solutions. One can eliminate some of these based on, e.g., convergence to an asymptotically globally stable steady state. Let the sextuple $\left(A_L^S, B_L^S, C_L^S, A_F^S, B_F^S, C_F^S\right)$ denote a solution to the above system, satisfying the additional desirable properties. Then, a pair of FSE strategies is given by

$$u_F = \kappa + \mathcal{V}_F' (x) = A_F^S x + B_F^S,$$
$$u_L = \kappa + \mathcal{V}_L' (x) = A_L^S x + B_L^S.$$

4.4 Time Consistency of Stackelberg Equilibria

When, at an initial instant of time, the leader announces a strategy she will use throughout the game, her goal is to influence the follower's strategy choice in a way that will be beneficial to her. *Time consistency* addresses the following question: given the option to re-optimize at a later time, will the leader stick to her original plan, i.e., the announced strategy and the resulting time path for her control variable? If it is in her best interest to deviate, then the leader will do so, and the equilibrium is then said to be *time inconsistent*. An inherently related question is then why would the follower, who is a rational player, believe in the announcement made by the leader at the initial time if it is not credible? The answer is clearly that she would not.

In most of the Stackelberg differential games, it turns out that the OLSE is time inconsistent, that is, the leader's announced control path $u_L(\cdot)$ is not credible. Markovian or feedback Stackelberg equilibria (SFE), on the other hand, are subgame perfect and hence time consistent; they are in fact *strongly time consistent*, which refers to the situation where the restriction of leader's originally announced strategy to a shorter time interval (sharing the same terminal time) is still SFE and regardless of what evolution the game had up to the start of that shorter interval.

The OLSE in Sect. 4.3 is time inconsistent. To see this, suppose that the leader has the option of revising her plan at time $\tau > 0$ and to choose a new decision rule $u_L(\cdot)$ for the remaining time span $[\tau, \infty)$. Then she will select a rule that satisfies $\theta(\tau) = 0$ (because this choice will fulfill the initial condition on the costate θ). It can be shown, by using the four state and costate equations $\left(\dot{x}, \dot{q}_F, \dot{q}_L, \dot{\theta}\right)$, that for some instant of time, $\tau > 0$, it will hold that $\theta(\tau) \neq 0$. Therefore, the leader will want to announce a new strategy at time τ, and this makes the original strategy time inconsistent, i.e., the new strategy does not coincide with the restriction of the original strategy to the interval $[\tau, \infty)$.

Before concluding this subsection, we make two useful observations.

Remark 13. Time consistency (and even stronger, strong time consistency) of FSE relies on the underlying assumption that the information structure is state feedback and hence without memory, that is, at any time t, the players do **not** remember the history of the game.

Remark 14. In spite of being time inconsistent, the OLSE can still be a useful solution concept for some short-term horizon problems, where it makes sense to assume that the leader will not be tempted to re-optimize at an intermediate instant of time.

5 Memory Strategies and Collusive Equilibria

5.1 Implementing Memory Strategies in Differential Games

As mentioned earlier, by memory strategies we mean that the players can, at any instant of time, recall any specific past information. The motivation for using memory strategies in differential games is in reaching through an equilibrium a desirable outcome that is not obtainable noncooperatively using open-loop or state-feedback strategies. Loosely speaking, this requires that the players *agree* (implicitly, or without taking on any binding agreement) on a desired trajectory to follow throughout the game (typically a cooperative solution) and are willing to implement a punishment strategy if a deviation is observed. Richness of an information structure, brought about through incorporation of memory, enables such monitoring.

If one party realizes, or remembers, that, in the past, the other party deviated from an agreed-upon strategy, it implements some pre-calculated punishment. Out of the fear of punishment, the players adhere to the Pareto-efficient path, which would be unobtainable in a strictly noncooperative game.

A punishment is conceptually and practically attractive only if it is *effective*, i.e., it deprives a player of the benefits of a defection, and *credible*, i.e., it is in the best interest of the player(s) who did not defect to implement this punishment. In this section, we first introduce the concept of non-Markovian strategies and the resulting Nash equilibrium and next illustrate these concepts through a simple example.

Consider a two-player infinite-horizon differential game, with state equation

$$\dot{x}(t) = f\left(x(t),\, u_1(t),\, u_2(t), t\right), \quad x(0) = x^0.$$

To a pair of controls $(u_1(t), u_2(t))$, there corresponds a unique trajectory $x(\cdot)$ emanating from x^0. Player j's payoff is given by

$$J_j(u_1(t), u_2(t); x^0) = \int_0^\infty e^{-\rho_j t} g_j(x(t), u_1(t), u_2(t), t)\, dt, \ j = 1, 2,$$

where $g_j(x(t), u_1(t), u_2(t), t)$ is taken to be bounded and continuously differentiable.[10] As before, the control set of player j is U_j and the state set X is identical to $\mathbb{R}^n$.

Heretofore, a strategy has been defined as a mapping from player's information space to her control set. Unfortunately, this direct approach poses formidable mathematical difficulties in the present context; therefore, we will define a strategy as an infinite sequence of approximate constructions, called δ-strategies. For player j, consider the sequence of times $t_j = i\delta, i = 0, 1, \ldots,$ where δ is a fixed positive number. For any time interval $[t_j, t_{i+1})$, let $\mathcal{U}_j^i$ be the set of measurable control functions $u_{j,i} : [t_j, t_{i+1}) \to U_j$, and let $\mathcal{U}^i = \mathcal{U}_1^i \times \mathcal{U}_2^i$. A δ-*strategy* for player j is a sequence $\Delta_j^\delta = \left(\Delta_{j,i}\right)_{i=0,1,\ldots,}$ of mappings

$$\Delta_{j,0} \in \mathcal{U}_j^0,$$
$$\Delta_{j,i} = \mathrm{U}^0 \times \mathrm{U}^1 \times \ldots \times \mathrm{U}^{i-1} \to \mathcal{U}_j^j \ \text{ for } \ j = 1, 2, \ldots.$$

A *strategy for player* j is an infinite sequence of δ-strategies:

$$\Delta_j = \left\{\Delta_j^{\delta_n} : \delta_n \to 0, n = 1, 2, \ldots\right\}.$$

Note that this definition implies that the information set of player j at time t is

$$\{(u_1(s), u_2(s)), 0 \le s < t\},$$

that is, the entire control history up to (but not including) time t. So when players choose δ-*strategies*, they are using, at successive sample times t_j, the accumulated information to generate a pair of measurable controls $\left(u_1^\delta(\cdot), u_2^{\delta'}(\cdot)\right)$ which, in turn, generate a unique trajectory $x^{\bar{\delta}}(\cdot)$ and thus, a unique outcome $w^{\bar{\delta}} = \left(w_1^{\bar{\delta}}, w_2^{\bar{\delta}}\right) \in \mathbb{R}^2$, where $\bar{\delta} = (\delta, \delta')$, and

$$w_j^{\bar{\delta}} = \int_0^\infty e^{-\rho_j t} g_j(x^{\bar{\delta}}(t), u_1^\delta(t), u_2^{\delta'}(t), t)\, dt.$$

[10]This assumption allows us to use the strong-optimality concept and avoid introducing additional technicalities.

An *outcome* of the strategy pair $\bar{\Delta}$ is a pair $\bar{w} \in \mathbb{R}^2$, which is a *limit* of the sequence $\left\{w^{\bar{\delta}_n}\right\}$ of the outcomes of δ-strategy pairs $\bar{\Delta}^{\bar{\delta}_n} = \left(\Delta_1^{\delta_n}, \Delta_2^{\delta_n'}\right)$ when n tends to infinity. With a strategy pair, the initial state and time are thus associated with a set $v\left(t^0, x^0; \bar{\Delta}\right)$ of possible outcomes. (Note that we have used the obvious extension to a non-zero initial time t^0.) The game is well defined if, for any strategy pair $\bar{\Delta}$ and any initial conditions $\left(t^0, x^0\right)$, the set of outcomes $v\left(t^0, x^0; \bar{\Delta}\right)$ is nonempty.[11]

Definition 7. A strategy pair $\bar{\Delta}^*$ is a Nash equilibrium at $\left(t^0, x^0\right)$ if, and only if,

1. the outcome set $v\left(t^0, x^0; \bar{\Delta}\right)$ reduces to a singleton $w^* = \left(w_1^*, w_2^*\right)$;
2. for all strategy pairs $\bar{\Delta}^{(1)} \triangleq \left(\Delta_1, \Delta_2^*\right)$ and $\bar{\Delta}^{(2)} \triangleq \left(\Delta_1^*, \Delta_2\right)$, the following holds for $j = 1, 2$:

$$(w_1, w_2) \in v\left(t^0, x^0; \bar{\Delta}^{(i)}\right) \Rightarrow w_j \leq w_j^*.$$

The equilibrium condition for the strategy pair is valid only at $\left(t^0, x^0\right)$. This implies, in general, that the Nash equilibrium that was just defined is not subgame perfect.

Definition 8. A strategy pair $\bar{\Delta}^*$ is a subgame-perfect Nash equilibrium at $\left(t^0, x^0\right)$ if, and only if,

1. given a control pair $\bar{u}(\cdot) : [t^0, t) \to U_1 \times U_2$ and the state $x(t)$ reached at time t, we define the prolongation of $\bar{\Delta}^*$ at $(t, x(t))$ as $\left\{\Delta^{\prime * \delta_n} : \delta_n \to 0, n = 1, 2, \ldots\right\}$ defined by

$$\begin{aligned}&\Delta^{\prime \delta_n}\left(\bar{u}_{[t,t+\delta_n]}, \ldots, \bar{u}_{[t+i\delta_n, t+(i+1)\delta_n]}\right)\\ &= \Delta^{\delta_n}\left(\bar{u}_{[0,\delta_n]}, \bar{u}_{[\delta_n, 2\delta_n]} \ldots, \bar{u}_{[t+i\delta_n, t+(i+1)\delta_n]}\right);\end{aligned}$$

2. the prolongation of $\bar{\Delta}^*$ at $(t, x(t))$ is again an equilibrium at $(t, x(t))$.

Before providing an illustrative example, we make a couple of points in the following remark.

Remark 15. 1. The information set was defined here as the entire control history. An alternative definition is $\{x(s), 0 \leq s < t\}$, that is, each player bases her decision on the entire past state trajectory. Clearly, this definition requires less

[11]Here, and in the balance of this section, we depart from our earlier convention of state-time ordering (x, t), and use the reverse ordering (t, x).

memory capacity and hence may be an attractive option, particularly when the differential game involves more than two players. (See Tolwinski et al. 1986 for details.)

2. The consideration of memory strategies in differential games can be traced back to Varaiya and Lin (1963), Friedman (1971), and Krassovski and Subbotin (1977). Their setting was (mainly) zero-sum differential games, and they used memory strategies as a convenient tool for proving the existence of a solution. Başar used memory strategies in the 1970s to show how richness of and redundancy in information structures could lead to *informationally nonunique* Nash equilibria (Başar 1974, 1975, 1976, 1977) and how the richness and redundancy can be exploited to solve for global Stackelberg equilibria (Başar 1979, 1982; Başar and Olsder 1980; Başar and Selbuz 1979) and to obtain incentive designs (Başar 1985). The exposition above follows Tolwinski et al. (1986) and Haurie and Pohjola (1987), where the setting is nonzero-sum differential games and the focus is on the construction of cooperative equilibria.

5.2 An Example

Consider a two-player differential game where the evolution of the state is described by

$$\dot{x}(t) = (1 - u_1(t))\, u_2(t), \quad x(0) = x^0 > 0, \tag{2.85}$$

where $0 < u_j(t) < 1$. The players maximize the following objective functionals:

$$J_1(u_1(t), u_2(t); x^0) = \alpha \int_0^\infty e^{-\rho t} \left(\ln u_1(t) + x(t)\right) dt,$$

$$J_2(u_1(t), u_2(t); x^0) = (1 - \alpha) \int_0^\infty e^{-\rho t} \left(\ln (1 - u_1(t))\,(1 - u_2(t)) + x(t)\right) dt,$$

where $0 < \alpha < 1$ and $0 < \rho \leq 1/4$.

Suppose that the two players wish to implement a cooperative solution noncooperatively by using non-Markovian strategies and threats.

Step 1: Determine Cooperative Outcomes. Assume that these outcomes are given by the joint maximization of the sum of players' payoffs. To solve this optimal control problem, we introduce the current-value Hamiltonian (we suppress the time argument):

$$\mathcal{H}(u_1, u_2, x, \lambda) = \alpha \ln u_1 + (1 - \alpha) \ln (1 - u_1)(1 - u_2) + x + q\,(1 - u_1)\, u_2,$$

where q is the current-value adjoint variable associated with the state equation (2.85). Necessary and sufficient optimality conditions are

$$\dot{x} = (1 - u_1) u_2, \quad x(0) = x^0 > 0,$$
$$\dot{q} = \rho q - 1, \quad \lim_{t\to\infty} e^{-\rho t} q(t) = 0,$$
$$\frac{\partial \mathcal{H}}{\partial u_1} = \frac{\alpha}{u_1} - \frac{(1-\alpha)}{(1-u_1)} - q u_2 = 0,$$
$$\frac{\partial \mathcal{H}}{\partial u_2} = -\frac{(1-\alpha)}{(1-u_2)} - q(1-u_1) = 0.$$

It is easy to verify that the unique optimal solution is given by

$$\left(u_1^*, u_2^*\right) = \left(\alpha\rho, \frac{1-\rho}{1-\alpha\rho}\right), \quad x^*(t) = x^0 + (1-\rho)\, t,$$
$$J_1\left(u_1^*(\cdot), u_2^*(\cdot); x^0\right) = \frac{\alpha}{\rho}\left(\ln \alpha\rho + x^0 + \frac{1-\rho}{\rho}\right),$$
$$J_2\left(u_1^*(\cdot), u_2^*(\cdot); x^0\right) = \frac{1-\alpha}{\rho}\left(\ln (1-\alpha)\,\rho + x^0 + \frac{1-\rho}{\rho}\right).$$

Note that both optimal controls satisfy the constraints $0 < u_j(t) < 1, \; j = 1, 2$.

Step 2: Compute Nash-Equilibrium Outcomes. As the game is of the linear-state variety,[12] open-loop and state-feedback Nash equilibria coincide. We therefore proceed with the derivation of the OLNE, which is easier to solve. To determine this equilibrium, we first write the players' current-value Hamiltonians:

$$\mathcal{H}_1(u_1, u_2, x, q_1) = \alpha(\ln u_1 + x) + q_1(1-u_1) u_2,$$
$$\mathcal{H}_2(u_1, u_2, x, q_2) = (1-\alpha)(\ln(1-u_1)(1-u_2) + x) + q_2(1-u_1) u_2,$$

where q_j is the costate variable attached by player j to the state equation (2.85). Necessary conditions for a Nash equilibrium are

$$\dot{x} = (1 - u_1) u_2, \quad x(0) = x^0 > 0,$$
$$\dot{q}_1 = \rho q_1 - \alpha, \quad \lim_{t\to\infty} e^{-\rho t} q_1(t) = 0,$$
$$\dot{q}_2 = \rho q_2 - (1-\alpha), \quad \lim_{t\to\infty} e^{-\rho t} q_2(t) = 0,$$
$$\frac{\partial}{\partial u_1}\mathcal{H}_1 = \frac{\alpha}{u_1} - q_1 u_2 = 0,$$

[12]In a linear-state differential game, the objective functional, the salvage value and the dynamics are linear in the state variables. For such games, it holds that a feedback strategy is constant, i.e., independent of the state and hence open-loop and state-feedback Nash equilibria coincide.

$$\frac{\partial}{\partial u_2}\mathcal{H}_2 = -\frac{(1-\alpha)}{(1-u_2)} - q_2(1-u_1) = 0.$$

It is easy to check that the Nash equilibrium is unique and is given by

$$(\bar{u}_1, \bar{u}_2) = \left(\frac{1-k}{2}, \frac{1+k}{2}\right), \quad \bar{x}(t) = x^0 + \left(\frac{1+k}{4}\right)t,$$

$$J_1\left(\bar{u}_1(\cdot), \bar{u}_2(\cdot); x^0\right) = \frac{\alpha}{\rho}\left(\ln\left(\frac{1-k}{2}\right) + x^0 + \frac{1+k}{2\rho} - 1\right),$$

$$J_2\left(\bar{u}_1(\cdot), \bar{u}_2(\cdot); x^0\right) = \frac{1-\alpha}{\rho}\left(\ln\rho + x^0 + \frac{1+k}{2\rho} - 1\right),$$

where $k = \sqrt{1-4\rho}$. Note that the equilibrium controls satisfy the constraints $0 < u_j(t) < 1$, $j = 1, 2$, and as expected in view of the game structure, they are constant over time.

Step 3: Construct a Collusive Equilibrium. We have thus so far obtained

$$\left(w_1^*, w_2^*\right) = \left(J_1\left(u_1^*(\cdot), u_2^*(\cdot); x^0\right), J_2\left(u_1^*(\cdot), u_2^*(\cdot); x^0\right)\right),$$

$$(\bar{w}_1, \bar{w}_2) = \left(J_1\left(\bar{u}_1(\cdot), \bar{u}_2(\cdot); x^0\right), J_2\left(\bar{u}_1(\cdot), \bar{u}_2(\cdot); x^0\right)\right).$$

Computing the differences

$$w_1^* - \bar{w}_1 = \frac{\alpha}{\rho}\left(\ln\left(\frac{1-k}{2\alpha\rho}\right) + \frac{1+\rho^2-3\rho+k}{2\rho}\right),$$

$$w_2^* - \bar{w}_2 = \frac{1-\alpha}{\rho}\left(\ln(1-\alpha) + \frac{1-3\rho+k}{2\rho}\right),$$

we note that they are independent of the initial state x^0 and that their signs depend on the parameter values. For instance, if we have the following restriction on the parameter values:

$$0 < \alpha < \min\left(\frac{1-k}{2\rho\exp\left(\frac{3\rho-1-\rho^2-k}{2\rho}\right)}, 1-\exp\left(\frac{3\rho-1+k}{2\rho}\right)\right),$$

then $w_1^* > \bar{w}_1$ and $w_2^* > \bar{w}_2$. Suppose that this is true. What remains to be shown is then that by combining the cooperative (Pareto-optimal) controls with the state-feedback (equivalent to open-loop, in this case) Nash strategy pair,

$$(\gamma_1(x), \gamma_2(x)) = (\bar{u}_1, \bar{u}_2) = \left(\frac{1-k}{2}, \frac{1+k}{2}\right),$$

we can construct a subgame-perfect equilibrium strategy in the sense of Definition 8.

Consider a strategy pair

$$\bar{\Delta}_j = \left\{ \bar{\Delta}_j^{*\delta_n} : \delta_n \to 0, n = 1, 2, \ldots \right\},$$

where, for $j = 1, 2$, $\bar{\Delta}_j^{*\delta}$ is defined as follows:

$$\Delta_j^{*\delta} = \left(\Delta_{j,i}\right)_{i=0,1,2,\ldots,}$$

with

$$\Delta_{j,0}^{*} = u_{j,0}^{*}(\cdot),$$

$$\Delta_{j,i}^{*} = \begin{cases} u_{j,i}^{*}(\cdot), & \text{if } \bar{u}(s) = \bar{u}^{*}(s) \text{ for almost } s \leq i\delta, \\ \varphi_j(x(j\delta)) = \bar{u}_j, & \text{otherwise,} \end{cases}$$

for $i = 1, 2, \ldots,$ where $u_{j,i}^{*}(\cdot)$ denotes the restriction truncation of $u_{j,}^{*}(\cdot)$ to the subinterval $[i\delta, (i+1)\delta]$, $i = 0, 1, 2, \ldots$, and $x(i\delta)$ denotes the state observed at time $t = i\delta$.

The strategy just defined is known as a *trigger strategy*. A statement of the trigger strategy, as it would be made by a player, is "At time t, I implement my part of the optimal solution if the other player has never cheated up to now. If she cheats at t, then I will retaliate by playing the state-feedback Nash strategy from t onward." It is easy to show that this trigger strategy constitutes a subgame-perfect equilibrium.

Remark 16. It is possible in this differential-game setting to define a retaliation period of finite length, following a deviation. Actually, the duration of this period can be designed to discourage any player from defecting. Also, in the above development and example, we assumed that a deviation is instantaneously detected. This may not necessarily be the case, and in such situations we can consider a detection lag. For an example of a trigger strategy with a finite retaliation period and detection lag, see Hämäläinen et al. (1984).

6 Conclusion

This chapter has provided an overview of the theory of nonzero-sum differential games formulated in continuous time and without any stochastic elements. Only noncooperative aspects of the theory have been covered, primarily under two different solution concepts: Nash equilibrium and Stackelberg equilibrium and several of their variants. The importance of information structures in such dynamic games has been emphasized, with special focus on open-loop and state-feedback information structures. The additional degrees of freedom memory strategies bring in in inducing specific behavior on the part of the players has also been discussed, and several special structures of differential games, such as linear-quadratic (or affine-quadratic) games, symmetric games, and zero-sum differential games, have also been covered, with some illustrative examples. The chapter has also emphasized the important role strategic equivalence plays in solvability of some classes of differential games.

There are several other issues very relevant to the topic and material of this chapter, which are covered by selected other chapters in the *Handbook*. These involve dynamic games described in discrete time, concave differential games with coupled state constraints defined over infinite horizon, dynamic games with an infinite number of players (more precisely, mean-field games), zero-sum differential games (with more in-depth analysis than the coverage in this chapter), games with stochastic elements (more precisely, stochastic games), mechanism designs, and computational methods, to list just a few.

References

Başar T (1974) A counter example in linear-quadratic games: existence of non-linear Nash solutions. J Optim Theory Appl 14(4):425–430

Başar T (1975) Nash strategies for M-person differential games with mixed information structures. Automatica 11(5):547–551

Başar T (1976) On the uniqueness of the Nash solution in linear-quadratic differential games. Int J Game Theory 5(2/3):65–90

Başar T (1977) Informationally nonunique equilibrium solutions in differential games. SIAM J Control 15(4):636–660

Başar T (1979) Information structures and equilibria in dynamic games. In: Aoki M, Marzollo A (eds) New trends in dynamic system theory and economics. Academic, New York, pp 3–55

Başar T (1982) A general theory for Stackelberg games with partial state information. Large Scale Syst 3(1):47–56

Başar T (1985) Dynamic games and incentives. In: Bagchi A, Jongen HTh (eds) Systems and optimization. Volume 66 of Lecture Notes in Control and Information Sciences. Springer, Berlin, pp 1–13

Başar T (1989) Time consistency and robustness of equilibria in noncooperative dynamic games. In: Van der Ploeg F, de Zeeuw A (eds) Dynamic policy games in economics, North Holland, Amsterdam/New York, pp 9–54

Başar T, Bernhard P (1995) H^∞-optimal control and related minimax design problems: a dynamic game approach, 2nd edn. Birkhäuser, Boston

Başar T, Haurie A (1984) Feedback equilibria in differential games with structural and modal uncertainties. In: Cruz JB Jr (ed) Advances in large scale systems, Chapter 1. JAE Press Inc., Connecticut, pp 163–201

Başar T, Olsder GJ (1980) Team-optimal closed-loop Stackelberg strategies in hierarchical control problems. Automatica 16(4):409–414

Başar T, Olsder GJ (1999) Dynamic noncooperative game theory. SIAM series in Classics in Applied Mathematics. SIAM, Philadelphia

Başar T, Selbuz H (1979) Closed-loop Stackelberg strategies with applications in the optimal control of multilevel systems. IEEE Trans Autom Control AC-24(2):166–179

Bellman R (1957) Dynamic programming. Princeton University Press, Princeton

Blaquière A, Gérard F, Leitmann G (1969) Quantitative and qualitative games. Academic, New York/London

Bryson AE Jr, Ho YC (1975) Applied optimal control. Hemisphere, Washington, DC

Carlson DA, Haurie A, Leizarowitz A (1991) Infinite horizon optimal control: deterministic and stochastic systems, vol 332. Springer, Berlin/New York
Case J (1969) Toward a theory of many player differential games. SIAM J Control 7(2):179–197
Dockner E, Jørgensen S, Long NV, Sorger G (2000) Differential games in economics and management science. Cambridge University Press, Cambridge
Engwerda J (2005) Linear-quadratic dynamic optimization and differential games. Wiley, New York
Friedman A (1971) Differential games. Wiley-Interscience, New York
Hämäläinen RP, Kaitala V, Haurie A (1984) Bargaining on whales: a differential game model with Pareto optimal equilibria. Oper Res Lett 3(1):5–11
Haurie A, Pohjola M (1987) Efficient equilibria in a differential game of capitalism. J Econ Dyn Control 11:65–78
Haurie A, Krawczyk JB, Zaccour G (2012) Games and dynamic games. World scientific now series in business. World Scientific, Singapore/Hackensack
Isaacs R (1965) Differential games. Wiley, New York
Jørgensen S, Zaccour G (2004) Differential games in marketing. Kluwer, Boston
Krassovski NN, Subbotin AI (1977) Jeux differentiels. Mir, Moscow
Lee B, Markus L (1972) Foundations of optimal control theory. Wiley, New York
Leitmann G (1974) Cooperative and non-cooperative many players differential games. Springer, New York
Mehlmann A (1988) Applied differential games. Springer, New York
Selten R (1975) Reexamination of the perfectness concept for equilibrium points in extensive games. Int J Game Theory 4:25–55
Sethi SP, Thompson GL (2006)Optimal control theory: applications to management science and economics. Springer, New York. First edition 1981
Starr AW, Ho YC (1969) Nonzero-sum differential games, Part I. J Optim Theory Appl 3(3): 184–206
Starr AW, Ho YC (1969) Nonzero-sum differential games, Part II. J Optim Theory Appl 3(4): 207–219
Tolwinski B, Haurie A, Leitmann G (1986) Cooperative equilibria in differential games. J Math Anal Appl 119:182–202
Varaiya P, Lin J (1963) Existence of saddle points in differential games. SIAM J Control Optim 7:142–157
von Stackelberg H (1934) Marktform und Gleischgewicht. Springer, Vienna [An English translation appeared in 1952 entitled The theory of the market economy. Oxford University Press, Oxford]
Yeung DWK, Petrosjan L (2005) Cooperative stochastic differential games. Springer, New York

Infinite Horizon Concave Games with Coupled Constraints

3

Dean Carlson, Alain Haurie, and Georges Zaccour

Contents

D. Carlson (✉)
Mathematical Reviews, American Mathematical Society, Ann Arbor, MI, USA
e-mail: dac@ams.org

A. Haurie
ORDECSYS and University of Geneva, Geneva, Switzerland

GERAD-HEC Montréal PQ, Montreal, QC, Canada
e-mail: ahaurie@gmail.com

G. Zaccour
Department of Decision Sciences, GERAD, HEC Montréal, Montreal, QC, Canada
e-mail: Georges.Zaccour@gerad.ca

T. Başar, G. Zaccour (eds.), *Handbook of Dynamic Game Theory*,
https://doi.org/10.1007/978-3-319-44374-4_34

Abstract

In this chapter, we expose a full theory for infinite-horizon concave differential games with coupled state-constraints. Concave games provide an attractive setting for many applications of differential games in economics, management science and engineering, and state coupling constraints happen to be quite natural features in many of these applications. After recalling the results of Rosen (1965) regarding existence and uniqueness of equilibrium of concave game with coupling contraints, we introduce the classical model of Ramsey and presents the Hamiltonian systems approach for its treatment. Next, we extend to a differential game setting the Hamiltonian systems approach and this formalism to the case of coupled state-constraints. Finally, we extend the theory to the case of discounted rewards.

Keywords

Concave Games · Coupling Constraints · Differential Games · Global Change Game · Hamilonian Systems · Oligopoly Game · Ramsey Model · Rosen Equilibrium

1 Introduction and Motivation

Suppose that players in a game face a joint, or coupled constraint on their decision variables. A natural question is how could they reach a solution that satisfies such constraint? The answer to this question depends on the mode of play. If the game is played cooperatively and the players can coordinate their strategies, then the problem can be solved by following a two-step procedure, i.e., by first optimizing the joint payoff under the coupled constraint and next by allocating the total payoff using one of the many available solution concepts of cooperative games. The optimal solution gives the actions that must be implemented by the players, as well as the status of the constraint (binding or not). If, for any reason, cooperation is not feasible, then the players should settle for an equilibrium under the coupled constraint. In a seminal paper (Rosen 1965), Rosen proposed in 1965 the concept of normalized equilibrium to deal with a class of noncooperative games with coupled constraints. In a nutshell, to obtain such an equilibrium, one appends to the payoff of each player a penalty term associated with the non-satisfaction of the coupled constraint, defined through a common Lagrange multiplier divided by a weight that is specific to each player.

When the game is dynamic, a variety of coupling constraints can be envisioned. To start with, the players' controls can be coupled. For instance, in a deregulated electricity industry where firms choose noncooperatively their outputs, an energy regulation agency may implement a renewable portfolio standard program, which typically requires that a given percentage of the total electricity delivered to the market is produced from renewable sources. Here, the constraint is on the collective industry's output and not on individual quantities, and it must be satisfied at each period of time. A coupling constraint may alternatively be on the state, reflecting that what matters is the accumulation process and not (necessarily only) the instantaneous actions. A well-known example is in international environmental treaties where countries (players) are requested to keep the global cumulative emissions of GHGs (the state variable) below a given value, either at each period of time or only at the terminal date of the game. The implicit assumptions here are (i) the environmental damage is essentially due to the accumulation of pollution, that is, not only to the flows of emissions, and (ii) the adoption of new less polluting production technologies and the change of consumption habits can be achieved over time and not overnight, and therefore it makes economic sense to manage the long-term concerns rather than the short-term details.

In this chapter, we expose a full theory for infinite horizon concave differential games with coupled state constraints. The focus on coupled state constraints can be justified by the fact that, often, a coupled constraint on controls can be reformulated as a state constraint via the introduction of some extra auxiliary state variables. An example will be provided to illustrate this approach. Concave games, which were the focus of Rosen's theory, provide an attractive setting for many applications of differential games in economics, management science, and engineering. Further, state coupling constraints in infinite horizon are quite natural features in many of these applications. As argued a long time ago by Arrow and Kurz in (1970), infinite horizon is a natural assumption in economic (growth) models because any chosen finite date is essentially arbitrary.

This chapter is organized as follows: Sect. 2 recalls the main results of Rosen's paper (Rosen 1965), Sect. 3 introduces the classical model of Ramsey and presents the Hamiltonian systems approach for its treatment, Sect. 4 extends to a differential game setting the Hamiltonian systems approach, and Sect. 5 extends the formalism to the case of coupled state constraints. An illustrative example is provided in Sect. 6, and the extension of the whole theory to the case of discounted rewards is presented in Sect. 7. Finally, Sect. 8 briefly concludes this chapter.

2 A Refresher on Concave *m*-Person Games with Coupled Constraints

To recall the main elements and results for concave games with coupled constraints developed in Rosen (1965), consider a game in normal form. Let $M = \{1, \ldots, m\}$ be the set of players. Each player $j \in M$ controls the action $x_j \in X_j$, where X_j is a compact convex subset of $\mathbb{R}^{m_j}$ and m_j is a given integer. Player j receives a payoff $\psi_j(x_1, \ldots, x_j, \ldots, x_m)$ that depends on the actions chosen by all the players. The reward function $\psi_j : X_1 \times \cdots \times X_j, \cdots \times X_m \to \mathbb{R}$ is assumed continuous in each x_i, for $i \neq j$, and concave in x_j. A *coupled constraint* set is defined as a proper subset $\mathcal{X}$ of $X_1 \times \cdots \times X_j \times \cdots \times X_m$. The constraint is that the joint action $\mathbf{x} = (x_1, \ldots, x_m)$ must be in $\mathcal{X}$.

Definition 1. An equilibrium under the coupled constraint set $\mathcal{X}$ is defined as an m-tuple $(x_1^*, \ldots, x_j^*, \ldots, x_m^*) \in \mathcal{X}$ such that for each player $j \in M$

$$\psi_j(x_1^*, \ldots, x_j^*, \ldots, x_m^*) \geq \psi_j(x_1^*, \ldots, x_j, \ldots, x_m^*) \quad \text{for all } x_j \in X_j \text{ s.t. } (x_1^*, \ldots, x_j, \ldots, x_m^*) \in \mathcal{X}. \tag{3.1}$$

For brevity, a coupled constraint equilibrium will be referred to by CCE.

Coupled constraints means that each player's strategy space may depend on the strategy of the other players. This may look awkward in a noncooperative game where the players cannot enter into communication or coordinate their actions. However, the concept is mathematically well defined. Further, some authors (see, e.g., Facchinei et al. (2007, 2009), Facchinei and Kanzow (2007), Fukushima (2009), Harker (1991), von Heusingen and Kanzow (2006) or Pang and Fukushima (2005)) call a coupled constraints equilibrium a **generalized Nash equilibrium** or GNE. See the paper by Facchinei et al. (2007) for a comprehensive survey on GNE and numerical solutions for this class of equilibria. Among other topics, the survey includes complementarity formulations of the equilibrium conditions and solution methods based on variational inequalities. Krawczyk (2007) also provides a comprehensive survey of numerical methods for the computation of coupled constraint equilibria. For applications of coupled constraint equilibrium in environmental economics, see Haurie and Krawczyk (1997), Tidball and Zaccour (2005, 2009), in electricity markets see Contreras et al. (2004) and Hobbs and Pang (2007), and see Kesselman et al. (2005) for an application to Internet traffic.

2.1 Existence of an Equilibrium

At a coupled constraint equilibrium point, no player can improve his payoff by a unilateral change in his strategy while keeping the combined vector in $\mathcal{X}$. To show that such a solution exists in a concave game, it is convenient to reformulate the equilibrium conditions as a *fixed-point* condition for a *point-to-set mapping*.[1] For that purpose, introduce the *global reaction function* $\theta : \mathcal{X}\times\mathcal{X}\times\mathbb{R}^m_+ \to \mathbb{R}$ defined by

$$\theta(\mathbf{x}, \mathbf{v}, \mathbf{r}) = \sum_{j=1}^{m} r_j \psi_j(x_1, \ldots, v_j, \ldots, x_m), \tag{3.2}$$

where the coefficients $r_j > 0$, $j = 1, \ldots, m$, are arbitrary positive weights given to each player's payoff. The precise role of this weighting scheme will be explained later. One may assume that $\sum_{j=1}^{m} r_j = 1$.

As defined in (3.2), the function $\theta(\mathbf{x}, \mathbf{v}, \mathbf{r})$ is continuous in $\mathbf{x}$ and concave in $\mathbf{v}$ for every fixed $\mathbf{x}$. This function is helpful to characterize an equilibrium through a fixed-point property, as shown in the following result, proved in Rosen (1965).

Lemma 1. *Let* $\mathbf{x}^* \in \mathcal{X}$ *be such that*

$$\theta(\mathbf{x}^*, \mathbf{x}^*, \mathbf{r}) = \max_{\mathbf{x}\in\mathcal{X}} \theta(\mathbf{x}^*, \mathbf{x}, \mathbf{r}). \tag{3.3}$$

Then, $\mathbf{x}^*$ *is a coupled-constraint equilibrium.*

To make the fixed-point argument more precise, introduce a coupled-reaction *mapping.*

Definition 2. The point-to-set mapping

$$\Gamma(\mathbf{x}, \mathbf{r}) = \{\mathbf{v} | \theta(\mathbf{x}, \mathbf{v}, \mathbf{r}) = \max_{\mathbf{w}\in\mathcal{X}} \theta(\mathbf{x}, \mathbf{w}, \mathbf{r})\}, \tag{3.4}$$

is called the coupled reaction mapping associated with the positive weighting $\mathbf{r}$. A fixed point of $\Gamma(\cdot, \mathbf{r})$ is a vector $\mathbf{x}^*$ such that $\mathbf{x}^* \in \Gamma(\mathbf{x}^*, \mathbf{r})$.

Lemma 1 shows that a fixed point of $\Gamma(\cdot, \mathbf{r})$ is a coupled constraint equilibrium. The proof of existence of a coupled constraint equilibrium requires the use of the Kakutani fixed-point theorem, which is recalled below.

[1] A point-to-set mapping, or *correspondence*, is a multivalued function that assigns vectors to sets. Similarly to the "usual" point-to-point mappings, these functions can also be continuous, upper semicontinuous, etc.

Theorem 1. *Let $\Phi : A \to 2^A$ be a point-to-set mapping with a closed graph,*[2] *where A is a compact subset (i.e., a closed and bounded subset) of $\mathbb{R}^m$. Then, there exists a fixed point for Φ, i.e., there exists $x^* \in \Phi(x^*)$ for some $x^* \in A$.*

The existence theorem for coupled constraint equilibrium follows as a direct application of the Kakutani fixed-point theorem.

Theorem 2. *Let the mapping $\Gamma(\cdot, \mathbf{r})$ be defined through (3.4). For any positive weighting $\mathbf{r}$, there exists a fixed point of $\Gamma(\cdot, \mathbf{r})$, i.e., a point $\mathbf{x}^*$ s.t. $\mathbf{x}^* \in \Gamma(\mathbf{x}^*, \mathbf{r})$. Hence, a coupled constraint equilibrium exists.*

2.2 Normalized Equilibria

2.2.1 Karush-Kuhn-Tucker Multipliers

Suppose that $\mathcal{X}$, the coupled constraint set (3.1), can be defined by a set of inequalities

$$h_k(\mathbf{x}) \geq 0, \quad k = 1, \ldots, p,$$

where $h_k : X_1 \times \ldots \times X_m \to \mathbb{R}$, $k = 1, \ldots, p$ are given concave functions. Assume further that the payoff functions $\psi_j(\cdot)$ as well as the constraint functions $h_k(\cdot)$ are continuously differentiable and satisfy a constraint qualification condition[3] so that Karush-Kuhn-Tucker multipliers exist for each of the implicit single agent optimization problems defined below.

Assume all players, other than Player j, use their strategies x^*_ℓ, $\ell \in M \setminus \{j\}$, while Player j uses x_j, and denote the corresponding joint strategy vector by $[\mathbf{x}^{*-j}, x_j]$ to refer to the decision vector where all players i, other than j, play x^*_i, while Player j uses x_j. Then, the equilibrium conditions (3.1) define a single agent optimization problem with a concave objective function and a convex compact admissible set. Under the assumed constraint qualification, there exists a vector of Karush-Kuhn-Tucker multipliers $\lambda_j = (\lambda_{jk})_{k=1,\ldots,p}$ such that the Lagrangian

$$\mathcal{L}_j([\mathbf{x}^{*-j}, x_j], \lambda_j) = \psi_j([\mathbf{x}^{*-j}, x_j]) + \sum_{k=1\ldots p} \lambda_{jk} h_k([\mathbf{x}^{*-j}, x_j]), \tag{3.5}$$

[2]This mapping has the closed graph property if, whenever the sequence $\{x_k\}_{k=1,2,\ldots}$ converges in $\mathbb{R}^m$ toward x^0 then any accumulation point y^0 of the sequence $\{y_k\}_{k=1,2,\ldots}$ in $\mathbb{R}^n$, where $y_k \in \Phi(x_k)$, $k = 1, 2, \ldots$, is such that $y^0 \in \Phi(x_0)$.

[3]Known from mathematical programming, see, e.g., Mangasarian (1969).

verifies, at the optimum, the following conditions:

$$0 = \frac{\partial}{\partial x_j}\mathcal{L}_j([\mathbf{x}^{*-j}, x_j^*], \lambda_j), \tag{3.6}$$

$$0 \le \lambda_j, \tag{3.7}$$

$$0 = \lambda_{jk} h_k([\mathbf{x}^{*-j}, x_j^*]) \quad k = 1, \ldots, p, \tag{3.8}$$

$$0 \le h_k([\mathbf{x}^{*-j}, x_j^*]) \quad k = 1, \ldots, p. \tag{3.9}$$

Definition 3. The equilibrium is *normalized* if the multiplier λ_j is colinear with a common vector λ_0, that is,

$$\lambda_j = \frac{1}{r_j}\lambda_0, \text{ for all } j \in M, \tag{3.10}$$

where the coefficient $r_j > 0$ is a weight assigned to player j.

Observe that the common multiplier λ_0 is associated with the following implicit mathematical programming problem:

$$\mathbf{x}^* = \text{argmax}_{\mathbf{x}\in\mathcal{X}}\theta(\mathbf{x}^*, \mathbf{x}, \mathbf{r}). \tag{3.11}$$

To see this, it suffices to write down the Lagrangian of this problem, that is,

$$\mathcal{L}_0(\mathbf{x}, \lambda_0) = \sum_{j\in M} r_j \psi_j([\mathbf{x}^{*-j}, x_j]) + \sum_{k=1\ldots p} \lambda_{0k} h_k(\mathbf{x}), \tag{3.12}$$

and the first-order necessary conditions

$$0 = \frac{\partial}{\partial x_j}\left\{ r_j \psi_j(\mathbf{x}^*) + \sum_{k=1,\ldots,p} \lambda_{0k} h_k(\mathbf{x}^*) \right\}, \quad j \in M, \tag{3.13}$$

$$0 \le \lambda_0, \tag{3.14}$$

$$0 = \lambda_{0k} h_k(\mathbf{x}^*) \quad k = 1, \ldots, p, \tag{3.15}$$

$$0 \le h_k(\mathbf{x}^*) \quad k = 1, \ldots, p. \tag{3.16}$$

Then, the relationship (3.10) is clear.

2.2.2 An Economic Interpretation

In a mathematical programming problem, a Karush-Kuhn-Tucker multiplier can be interpreted as the marginal cost or *shadow price* associated with the right-hand side of a constraint. More precisely, the multiplier measures the sensitivity of the optimal solution to marginal changes of this right-hand side. The multiplier permits also a *price decentralization* in the sense that, through an ad hoc pricing mechanism, the optimizing agent is induced to satisfy the constraints.

In a normalized equilibrium, the shadow cost interpretation is not so apparent; however, the *price decomposition principle* is still valid. Once the common multiplier has been defined with the associated weighting $r_j > 0$, $j = 1, \ldots, m$, the coupled constraint will be satisfied by equilibrium-seeking players, playing without the coupled constraint but using the Lagrangians as payoffs

$$\mathcal{L}_j([\mathbf{x}^{*-j}, x_j], \lambda_j) = \psi_j([\mathbf{x}^{*-j}, x_j]) + \frac{1}{r_j} \sum_{k=1\ldots p} \lambda_{0k} h_k([\mathbf{x}^{*-j}, x_j]),$$
$$j = 1, \ldots, m.$$

The common multiplier permits then an "implicit pricing" of the common constraint so that the latter remains compatible with the equilibrium structure. However, to be useful this result necessitates uniqueness.

2.3 Uniqueness of Equilibrium

In a mathematical programming framework, uniqueness of an optimum results from strict concavity of the objective function to be maximized. In a game, uniqueness of the equilibrium requires a more stringent concavity condition, called *diagonal strict concavity* by Rosen (1965).

Consider the following function:

$$\sigma(\mathbf{x}, \mathbf{r}) = \sum_{j=1}^{m} r_j \psi_j(\mathbf{x}), \tag{3.17}$$

sometimes referred to as the *joint payoff*. Define the *pseudo-gradient* of this function as the vector

$$g(\mathbf{x}, \mathbf{r}) = \begin{pmatrix} r_1 \dfrac{\partial}{\partial x_1} \psi_1(\mathbf{x}) \\ r_2 \dfrac{\partial}{\partial x_2} \psi_2(\mathbf{x}) \\ \vdots \\ r_m \dfrac{\partial}{\partial x_m} \psi_m(\mathbf{x}) \end{pmatrix}. \tag{3.18}$$

Note that this expression is composed of the partial gradients of the different player payoffs with respect to the decision variables of the corresponding player.

Definition 4. The function $\sigma(\mathbf{x}, \mathbf{r})$ is **diagonally strictly concave** on $\mathcal{X}$ if the following condition holds:

$$(\mathbf{x}_2 - \mathbf{x}_1)' g(\mathbf{x}_1, \mathbf{r}) + (\mathbf{x}_1 - \mathbf{x}_2)' g(\mathbf{x}_2, \mathbf{r}) > 0, \tag{3.19}$$

for every $\mathbf{x}_1$ and $\mathbf{x}_2$ in $\mathcal{X}$.

A sufficient condition for $\sigma(\mathbf{x}, \mathbf{r})$ to be *diagonally strictly concave* is that the symmetric matrix[4] $[G(\mathbf{x}, \mathbf{r}) + G(\mathbf{x}, \mathbf{r})']$ be negative definite for any $\mathbf{x}$ in $\mathcal{X}$, where $G(\mathbf{x}, \mathbf{r})$ is the Jacobian of $g(\mathbf{x}, \mathbf{r})$ with respect to $\mathbf{x}$. The Uniqueness theorem proved by Rosen is now stated as:[5]

Theorem 3. *If $\sigma(\mathbf{x}, \mathbf{r})$ is **diagonally strictly concave** on the convex set $\mathcal{X}$, with the assumptions ensuring existence of the Karush-Kuhn-Tucker multipliers, then there exists a unique normalized equilibrium for the weighting scheme* $\mathbf{r} > 0$.

3 A Refresher on Hamiltonian Systems

A second pillar of a complete theory of open-loop differential games with coupled constraints is the theory of Hamiltonian systems. The objective here is not to provide a full introduction to this theory, but to recall what is needed in the sequel. It is done below using the well-known Ramsey problem, which has led to considerable development in economic dynamics using the theory of Hamiltonian systems in the calculus of variations (Cass and Shell 1976). Next, one hints at the extension of the theory to infinite horizon concave open-loop differential games, with coupled state constraints.

3.1 Ramsey Problem

Consider a stock of capital $x(t)$ at time $t > 0$ that evolves over time according to the differential equation

$$\dot{x}(t) = f(x(t)) - \delta x(t) - c(t), \tag{3.20}$$

[4]The expression in the square brackets is sometimes referred to as the pseudo-Hessian of $\sigma(\mathbf{x}, \mathbf{r})$.

[5]See Rosen's paper (1965) or Haurie et al. (2012) pp. 61–62.

with fixed initial condition $x(0) = x_0$, and where $f : [0, \infty) \to \mathbb{R}$ is a continuous production function, $\delta > 0$ is the rate of depreciation of capital, and $c(t)$ denotes consumption at time t satisfying $c(t) \leq f(x(t)) - \delta x(t)$. The agent reward is measured by accumulated utility of consumption described by the integral functional

$$W^T(c(\cdot)) := \int_0^T \mathcal{U}(c(t))\, dt, \tag{3.21}$$

where $\mathcal{U} : [0, +\infty) \to \mathbb{R}$ is a concave function. Ramsey's goal was to determine the rate of consumption that maximizes $\lim_{T\to\infty} W^T(c(\cdot))$. The first difficulty in analyzing this problem is that it is not well defined, since for most admissible consumption schedules $c(\cdot)$, the improper integral

$$W(c(\cdot)) = \int_0^\infty \mathcal{U}(c(t))\, dt = \lim_{T\to\infty} \int_0^T \mathcal{U}(c(t))\, dt,$$

does not converge. If instead the decision maker maximizes a stream of discounted utility of consumption given by

$$W_\rho(c(\cdot)) = \int_0^T e^{-\rho t}\mathcal{U}(c(t))\, dt,$$

where ρ is the discount rate $(0 < \rho < 1)$, then convergence of the integral is guaranteed by assuming $\mathcal{U}(c)$ is bounded. However, Ramsey dismisses this case as unethical for problems involving different generations of agents in that the discount rate weights a planner's decision toward the present generation at the expense of the future ones. (Those concerns are particularly relevant when dealing with global environmental change problems, like those created by GHG[6] long-lived accumulation in the atmosphere and oceans.) To deal with the undiscounted case, Ramsey introduced what he referred to as maximal sustainable rate of enjoyment or "bliss." One can view bliss, denoted by B, as an "optimal steady state" or the optimal value of the mathematical program, referred to here as the optimal steady-state problem

$$B = \max\{U(c) : c = f(x) - \delta x\}, \tag{3.22}$$

which is assumed to have a unique solution $(\bar{x}, \bar{c})$. The objective can now be formulated as

$$W_B(c(\cdot)) = \int_0^\infty (\mathcal{U}(c(t)) - B)\ dt,$$

[6]Greenhouse gases.

which can be made finite by assuming that there exists a consumption rate that can steer the capital stock from the value x_0 to $\bar{x}$ in a finite time. That is, there exists a time T and a consumption rate $\hat{c}(t)$ defined on $[0, T]$ so that at time T the solution $\hat{x}(t)$ of (3.20) with $c(t) = \hat{c}(t)$ satisfies $\hat{x}(T) = \bar{x}$. This is related to the so-called turnpike property, which is common in economic growth models.

To formulate the Ramsey problem as a problem of Lagrange in the parlance of calculus of variations, introduce the integrand $L(x, z) = \mathcal{U}(z - f(x) + \delta x) - B$. Consequently, the optimization problem becomes an Infinite horizon problem of Lagrange consisting of maximizing

$$J(x(\cdot)) = \int_0^\infty L(x(t), \dot{x}(t))\, dt = \int_0^\infty [\mathcal{U}(\dot{x}(t) - f(x(t)) + \delta x(t)) - B]\, dt, \tag{3.23}$$

over all admissible state trajectories $x(\cdot) : [0, \infty) \to \mathbb{R}$ satisfying the fixed initial condition $x(0) = x_0$. This formulation portrays the Ramsey model as an infinite horizon problem of calculus of variations, which is a well-known problem.

Now, consider this problem in more generality by letting $L : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ be a twice differentiable, concave function and consider the infinite horizon problem of maximizing

$$J(x) = \int_0^\infty L(x(t), \dot{x}(t))\, dt, \tag{3.24}$$

over all admissible state trajectories $x(\cdot) : [0, +\infty) \to \mathbb{R}^n$ satisfying a fixed initial condition $x(0) = x_0$, where $x_0 \in \mathbb{R}^n$ is given. The standard optimality conditions for this problem is that an optimal solution $x(\cdot)$ must satisfy the Euler-Lagrange equations

$$\frac{d}{dt}\left(\left.\frac{\partial L}{\partial z_i}\right|_{(x(t),\dot{x}(t))} \right) = \left.\frac{\partial L}{\partial x_i}\right|_{(x(t),\dot{x}(t))}, \quad i = 1, 2, \ldots, \tag{3.25}$$

which for convenience are rewritten as

$$p_i(t) = \partial_{z_i} L(x(t), \dot{x}(t)), \quad \dot{p}_i(t) = \partial_{x_i} L(x(t), \dot{x}(t)), \quad i = 1, 2, \ldots, n, \tag{3.26}$$

where ∂_{z_i} and ∂_{x_i} denote the partial derivatives of $L(x, z)$ with respect to the z_i-th and x_i-th variables, respectively. For the Ramsey model described above, this set of equations reduces to the single equation

$$\frac{d}{dt}\mathcal{U}'(c(t)) = -\mathcal{U}'(c(t))(f'(x(t)) - \delta),$$

where $c(t) = \dot{x}(t) - f(x(t)) - \delta x(t)$, which is the familiar Ramsey's rule.

From a dynamical systems point of view, it is convenient to express the Euler-Lagrange equations in an equivalent Hamiltonian form. To do so, define the

Hamiltonian $H : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ through the formula

$$H(x, p) = \sup_z \{L(x, z) + p'z\}. \tag{3.27}$$

When L is twice continuously differentiable and concave, it follows that

$$H(x, p) = L(x, \hat{z}) + p'\hat{z},$$

where $\hat{z}$ is related to p and x through the formulas $p_i = -\partial_{z_i} L(x, \hat{z})$ for $i = 1, 2, \ldots, n$. Further, one also has that

$$\partial_{x_i} H(x, p) = \partial_{x_i} L(x, \hat{z}) \quad \text{and} \quad \partial_{p_i} H(x, p) = \hat{z}_i, \quad i = 1, 2 \ldots n.$$

From this, if x^* is a solution of the Euler-Lagrange equations it follows that the pair $(x^*(t), p^*(t))$, where $p^*(t) = -\partial_{z_i} L(x^*(t), \dot{x}^*(t))$ satisfies the Hamiltonian system

$$\dot{x}_i^*(t) = \partial_{p_i} H(x^*(t), p^*(t)), \quad \dot{p}_i{}^*(t) = \partial_{x_i} H(x^*(t), p^*(t)), \quad i = 1, 2, \ldots n. \tag{3.28}$$

The optimal steady-state problem that arises in the Ramsey model is equivalently expressed as

$$L(\bar{x}, 0) = \max\{L(x, 0)\} = \max\{\mathcal{U}(c) : c = f(x) - \delta x\}. \tag{3.29}$$

This leads to considering the analogous problem for the general Lagrange problem. A necessary and sufficient condition for $\bar{x}$ to be an optimal steady state is that $\partial_{x_i} L(\bar{x}, 0) = 0$. Setting $\bar{p}_i = \partial L_{z_i}(\bar{x}, 0)$ gives a steady-state pair for the Hamiltonian dynamical system. That is,

$$0 = \partial_{p_i} H(\bar{x}, \bar{p}), \quad 0 = \partial_{p_i} H(\bar{x}, \bar{p}), \quad i = 1, 2, \ldots, p. \tag{3.30}$$

When $L(x, z)$ is concave, it follows that the pair $(\bar{x}, \bar{p})$ is a saddle point for the Hamiltonian system (3.28) giving rise to a stable manifold and an unstable manifold.[7] This leads to the *global asymptotic stability* (GAS) property for optimal trajectories $x(\cdot)$ and costate $p(\cdot)$, that is, the optimal steady state, actually the bliss point, is a global attractor for all optimal trajectories emanating from the different possible initial states.

For finite horizon problems in this setting, the turnpike property says that the optimal steady-state trajectory is an attractor for the optimal state trajectory. Thus the turnpike property states that for all sufficiently long terminal times T the optimal state trajectory spends most of its time near the optimal steady state. More precisely

[7] For a discussion of these ideas, see, Rockafellar (1973).

there exists a time $T_0 > 0$ such that for any time $T \geq T_0$ and each $\epsilon > 0$ there exists a constant $B_\epsilon > 0$ (independent of time T) so that

$$\text{meas}\{t \in [0, T] : \|x^*(t) = \bar{x}\| > \epsilon\} \leq B_\epsilon.$$

For an introduction to these ideas for both finite and infinite horizon problems, see Carlson et al. (1991, Chap. 3). For more details in the infinite horizon case, see the monograph by Cass and Shell (1976).

Remark 1. In concave problems of Lagrange, one can use the theory of convex analysis to allow for nonsmoothness of the integrand L. In this case, the Hamiltonian H is concave in x and convex in p, and the Hamiltonian system may be written as

$$\dot{x}_i^*(t) \in \partial_{p_i} H(x^*(t), p^*(t)), \quad \dot{p}_i{}^*(t) \in \partial_{x_i} H(x^*(t), p^*(t)), \quad i = 1, 2, \ldots n, \tag{3.31}$$

where now the derivative notation ∂ refers to the set of subgradients. A similar remark applies to (3.26). Therefore, in the rest of the chapter, the notation for the nonsmooth case will be used.[8]

3.2 Toward Competitive Models

Motivated by the study of optimal economic growth models, a general theory of optimal control over an infinite time horizon has been developed. See Arrow and Kurz (1970) and Cass and Shell (1976) for a presentation of the economic models and to the book Carlson et al. (1991) for a comprehensive discussion of the optimal control problem. An extension of this theory to open-loop differential games is natural both as an economic paradigm and as an optimization problem. Indeed, it is appealing to also consider the real case where several firms compete in a market through their production capacities. One of the first infinite horizon models of competition among a few firms is due to Brock (1977). His model assumed decoupled dynamics, i.e., each firm controls its own accumulation of production capacity. The firms are only coupled in the payoff functionals, and more specifically through the demand function.

An attempt to extend the *global asymptotic stability* (GAS) conditions of state and costate trajectories, known as the *turnpike property*, to open-loop differential games (OLDG) is also reported in Brock (1977). A set of sufficient conditions for obtaining GAS results in infinite horizon OLDG has been proposed by Haurie and Leitmann in (1984). To define the equilibrium, the authors use the idea of overtaking optimality of the response of each player to the controls chosen by the opponents. Conditions for GAS are given in terms of a so-called *vector Lyapunov function*

[8] Readers who do not feel comfortable with nonsmooth analysis should, e.g., view the notation $\dot{p} \in \partial_x H$ as $\dot{p} = \frac{\partial}{\partial x} H$.

applied to the *pseudo-Hamiltonian system* resulting from the necessary optimality conditions. Knowing that such a GAS property holds permits in particular the development of numerical methods for solving infinite horizon differential games, the use of this property in the analysis of competition has been well illustrated in Hämäläinen et al. (1985) where a transboundary fisheries model, with several nations exploiting the same biomass, has been studied and solved numerically, using the asymptotic steady state as terminal conditions. Now, to take into account coupled state constraints, one needs to extend to a dynamic setting the results obtained by Rosen (1965) for concave static games. This leads to a rather complete theory with proofs of existence, uniqueness, and asymptotic stability (i.e., turnpike property) of overtaking equilibrium programs for a class of games satisfying a strong concavity assumption, namely, strict diagonal concavity.[9]

The rest of this chapter presents a comprehensive theory concerning the existence, uniqueness, and GAS of equilibrium solutions, for a class of concave infinite horizon open-loop differential games with coupled state constraints. While pointwise state constraints generally pose no particular difficulties in establishing the existence of an equilibrium (or optimal solution), they do create problems when the relevant necessary conditions are utilized to determine the optimal solution. Indeed, in these situations the adjoint variable will possibly have a discontinuity whenever the optimal trajectory hits the boundary of the constraint (see the survey paper Hartl et al. 1995). The proposed approach circumvents this difficulty by introducing a relaxed asymptotic formulation of the coupled state constraint. This relaxation of the constraint, which uses an asymptotic steady-state game with coupled constraints à la Rosen, proves to be well adapted to the context of economic growth models with global environmental constraints.

4 Open-Loop Differential Games Played Over Infinite Horizon

This section introduces the concept of an *overtaking equilibrium* for a class of noncooperative infinite horizon differential games.

4.1 The Class of Games of Interest

Consider a controlled system defined by the following data:

- An infinite time horizon $t \in [0, \infty)$.

[9] A discrete time version of the model studied in this chapter can be found in Carlson and Haurie (1996) where a turnpike theory for discrete time competitive processes is developed.

- A set of players $M \doteq \{1, \ldots, m\}$. Each of the m players is represented at time t by a state $x_j(t) \in \mathbb{R}^{n_j}$, where n_j is a given positive integer. Let $n \doteq n_1 + \ldots + n_m$.[10]
- For each $j \in M$, a trajectory for player j is defined as an absolutely continuous function $x_j(\cdot) = (x_j(t) \in \mathbb{R}^{n_j} : t \geq 0)$ that satisfies whatever constraints are imposed on the model. An M-trajectory is defined as $\mathbf{x}(\cdot) = (\mathbf{x}(t) : t \geq 0) \doteq ((x_j(t))_{j \in M} : t \geq 0)$.
- Along an M-trajectory, a reward accumulation process for player j is defined as

$$\phi_j^T(\mathbf{x}(\cdot)) = \int_0^T L_j(\mathbf{x}(t), \dot{x}_j(t))\, dt, \quad \forall\, T > 0, \tag{3.32}$$

where $L_j : \mathbb{R}^n \times \mathbb{R}^{n_j} \to \mathbb{R} \cup \{-\infty\}$, $j \in M$ are given functions and $\dot{x}_j(t) = \frac{d}{dt} x_j(t)$. The expression $L_j(\mathbf{x}(t), \dot{x}_j(t))$ represents a reward rate for player j when the M-trajectory is $x(t)$ and the velocity for player j is $\dot{x}_j(t)$.

Remark 2. A control decoupling is assumed since only the velocity $\dot{x}_j(t)$ enters in the definition of the reward rate of player j. This is not a very restrictive assumption for most applications in economics.

Remark 3. Adopting a nonsmooth analysis framework means that there is indeed no loss of generality in adopting this generalized calculus of variations formalism instead of a state equation formulation of each player's dynamics. For details on the transformation of a fully fledged control formulation, including state and control constraints, into a generalized calculus of variations formulation, see Brock and Haurie (1976), Feinstein and Luenberger (1981), or Carlson et al. (1991). In this setting the players' strategies (or control actions) are applied through their derivatives $\dot{x}_j(\cdot)$. Once a fixed initial state is imposed on the players, each of their strategies give rise to a unique state trajectory (via the Fundamental Theorem of Calculus). Thus one is able to say that each player chooses his/her state trajectory $x_j(\cdot)$ to mean that in reality the strategy $\dot{x}_j(\cdot)$ is chosen.

4.2 The Overtaking Equilibrium Concept

The overtaking equilibrium concept is a version of the Nash equilibrium concept which is adapted to the consideration of an infinite time horizon. Given an M-trajectory $\mathbf{x}^*(\cdot)$, let $[\mathbf{x}^{*(j)}(\cdot); x_j(\cdot)]$ denote the M-trajectory obtained when player j changes unilaterally his trajectory to x_j.

[10] To distinguish between functions of time t and their images, the notation $x(\cdot)$ will always denote the function, while $x(t)$ will denote its value at time t (i.e., a vector).

Definition 5. An M-trajectory $\mathbf{x}^*(\cdot)$ is an **overtaking equilibrium** at $\mathbf{x}^o$ if

1. $\mathbf{x}^*(0) = \mathbf{x}^o$.
2. $\liminf_{T\to\infty}(\phi_j^T(\mathbf{x}^*(\cdot)) - \phi_j^T([\mathbf{x}^{*(j)}(\cdot); x_j(\cdot)]) \geq 0$, for all trajectories $x_j(\cdot)$ such that $x_j(0) = x_j^o$ for all $j \in M$.

Remark 4. The notation $[\mathbf{x}^{*(j)}; x_j]$ is to emphasize that the focus is on player j. That is, $[\mathbf{x}^{*(j)}; x_j] \doteq (x_1^*, x_2^*, \ldots, x_{j-1}^*, x_j, x_{j+1}^*, \ldots, x_m^*)$.

Remark 5. The consideration of the *overtaking optimality* concept for economic growth problems is due to von Weizsäcker (1965), and its extension to the *overtaking equilibrium* concept in dynamic games has been first proposed by Rubinstein in (1979). The concept has also been used in Haurie and Leitmann (1984) and Haurie and Tolwinski (1985).

4.3 Optimality Conditions

This section deals with conditions that guarantee that all the overtaking equilibrium trajectories, emanating from different initial states, bunch together at infinity. As in the theory of optimal economic growth and, more generally, the theory of asymptotic control of convex systems (see Brock and Haurie 1976), one may expect the steady state to play an important role in the characterization of the equilibrium solution. More precisely, one expects the steady-state equilibrium to be unique and to define an attractor for all equilibrium trajectories, emanating from different initial states.

First, recall the necessary optimality conditions for open-loop (overtaking) equilibrium. These conditions are a direct extension of the celebrated *maximum principle* established by Halkin (1974) for infinite horizon control problems but written for the case of convex systems. Introduce for $j \in M$ and $p_j \in \mathbb{R}^{n_j}$ the *Hamiltonians* $H_j : \mathbb{R}^n \times \mathbb{R}^{n_j} \to \mathbb{R} \cup \{-\infty\}$, defined as

$$H_j(\mathbf{x}, p_j) = \sup_{z_j}\{L_j(\mathbf{x}, z_j) + p_j' z_j\}. \tag{3.33}$$

Here p_j is called a j-*supporting costate* vector. The function

$$\mathbf{p}(t) = ((p_j(t))_{j\in M} \ : \ t \geq \ 0)$$

will be called an M-*costate trajectory*.

Assumption 1. *For each j, the Hamiltonian $H_j(\mathbf{x}, p_j)$ is concave in x_j and convex in p_j.*

For an overtaking equilibrium, the following necessary conditions hold (see Carlson and Haurie 1995 for proof):

Theorem 4. *If* $\mathbf{x}^*$ *is an* ***overtaking equilibrium*** *at initial state* $\mathbf{x}^o$, *then there exists an absolutely continuous* M*-costate trajectory* $\mathbf{p}^*$ *such that*

$$\dot{x}_j^*(t) \in \partial_{p_j} H_j(\mathbf{x}^*(t), p_j^*(t)), \tag{3.34}$$

$$\dot{p}_j^*(t) \in -\partial_{x_j} H_j(\mathbf{x}^*(t), p_j^*(t)), \tag{3.35}$$

for all $j \in M$.

The relations (3.34) and (3.35) are also called a *pseudo-Hamiltonian system* in Haurie and Leitmann (1984). These conditions are incomplete since only initial conditions are specified for the M-trajectories and no transversality conditions are given for their associated M-costate trajectories. In the single-player case, this system is made complete by invoking the *turnpike property* which provides an asymptotic transversality condition. Due to the coupling among the players, the system (3.35) considered here does not fully enjoy the rich geometric structure found in the classical optimization setting (e.g., the saddle point behavior of Hamiltonian systems in the autonomous case). Below, one gives conditions under which the turnpike property holds for these pseudo-Hamiltonian systems.

4.4 A Turnpike Result

One of the fundamental assumptions that underlines these developments is now introduced. It is directly linked to the *strict diagonal concavity assumption* made by Rosen (1965) in his study of concave static games. Recall first the definition given by Rosen. Let $\mathbf{x} = (x_j)_{j=1,\ldots,m} \in \mathbb{R}^{n_1} \times \cdots \times \mathbb{R}^{n_m}$ and consider m continuously differentiable functions $\phi_j(\mathbf{x})$. Let $\nabla_j \phi_j(\mathbf{x})$ denote the gradient of $\phi_j(\mathbf{x})$ with respect to x_j. The sum

$$\sigma(\mathbf{x}) = \sum_{j=1}^{m} \phi_j(\mathbf{x}),$$

is said to be *diagonally strictly concave,* if for every $\mathbf{x}^0$ and $\mathbf{x}^1$ one has

$$\sum_{j=1,\ldots,m} (x_j^1 - x_j^0)'(\nabla_j \phi_j(\mathbf{x}^0) - \nabla_j \phi_j(\mathbf{x}^1)) > 0.$$

The assumption concerns the sum of the Hamiltonians. It is also formulated in a slightly more general way to take care of the possible nondifferentiability of these functions.

Assumption 2 (Strict diagonal concavity-convexity assumption (SDCCA)). *Assume that the combined Hamiltonian $\sum_{j\in M} H_j(\mathbf{x}, p_j)$ is strictly diagonally concave in x, convex in p. That is,*

$$\sum_{j\in M}\left[(\hat{p}_j-\tilde{p}_j)'(\hat{\pi}_j-\tilde{\pi}_j)+(\hat{x}_j-\tilde{x}_j)'(\hat{\xi}_j-\tilde{\xi}_j)\right]>0, \tag{3.36}$$

for all $(\hat{x}_j, \tilde{x}_j, \hat{p}_j, \tilde{p}_j)$ and $(\hat{\pi}_j, \tilde{\pi}_j, \hat{\xi}_j, \tilde{\xi}_j)$, such that

$$\hat{\pi}_j \in \partial_{p_j} H_j(\hat{\mathbf{x}}, \hat{p}_j), \quad \tilde{\pi}_j \in \partial_{p_j} H_j(\tilde{\mathbf{x}}, \tilde{p}_j), \tag{3.37}$$

$$\hat{\xi}_j \in -\partial_{x_j} H_j(\hat{\mathbf{x}}, \hat{p}_j), \quad \tilde{\xi}_j \in -\partial_{x_j} H_j(\tilde{\mathbf{x}}, \tilde{p}_j). \tag{3.38}$$

A direct consequence of Assumption 2 is the following lemma,[11] which gives insight into the definition of a Lyapunov function providing a sufficient condition for GAS.

Lemma 2. *Let $\hat{\mathbf{x}}(\cdot)$ and $\tilde{\mathbf{x}}(\cdot)$ be two overtaking equilibria at $\hat{\mathbf{x}}^o$ and $\tilde{\mathbf{x}}^o$, respectively, with their respective associated M-costate trajectories $\hat{\mathbf{p}}(\cdot)$ and $\tilde{\mathbf{p}}(\cdot)$. Then, under Assumptions 1 and 2, the inequality*

$$\sum_{j\in M}\frac{d}{dt}\left[(\hat{p}_j(t)-\tilde{p}_j(t))'(\hat{x}_j(t)-\tilde{x}_j(t))\right]>0, \tag{3.39}$$

holds.

A turnpike theorem for overtaking equilibria follows[12] under a strengthening of the inequality (3.39).

Definition 6. The overtaking equilibrium M-program $\hat{\mathbf{x}}(\cdot)$ is called **strongly diagonally supported** by the M-costate trajectory $\hat{\mathbf{p}}(\cdot)$ if, for every $\varepsilon > 0$ there exists a $\delta > 0$, such that for all $t \geq 0$, $\|\mathbf{x}-\hat{\mathbf{x}}(t)\| + \|\mathbf{p}-\hat{\mathbf{p}}(t)\| > \varepsilon$ implies

$$\sum_{j\in M}\left[\left(\frac{d}{dt}\hat{p}_j(t)-\pi_j\right)'(\hat{x}_j(t)-x_j)+\left(\frac{d}{dt}\hat{x}_j(t)-\xi_j\right)'(\hat{p}_j(t)-p_j)\right]>\delta, \tag{3.40}$$

for all (x_j, p_j) and (π_j, ξ_j), such that

$$\pi_j \in \partial_{p_j} H(\mathbf{x}, p_j), \tag{3.41}$$

[11] See Carlson and Haurie (1995) for a proof.

[12] See Carlson and Haurie (1995) for a proof.

$$\xi_j \in -\partial_{x_j} H_j(\mathbf{x}, p_j). \tag{3.42}$$

Remark 6. The stricter inequality (3.40) is obtained as a consequence of Assumption (2) or inequality (3.39) when the state variable $\mathbf{x}$ remains in a compact set; this is known as the Atsumi lemma in the single-player case.

Theorem 5. *Let $\hat{\mathbf{x}}(\cdot)$ with its associated M-costate trajectory $\hat{\mathbf{p}}(\cdot)$ be a strongly diagonally supported overtaking equilibrium at $\hat{\mathbf{x}}^o$, such that*

$$\limsup_{t\to\infty} \|(\hat{\mathbf{x}}(t), \hat{\mathbf{p}}(t))\| < \infty.$$

Let $\tilde{\mathbf{x}}(\cdot)$ be another overtaking equilibrium at $\tilde{\mathbf{x}}^o$ with $\tilde{\mathbf{p}}(\cdot)$ its associated M-costate trajectory such that

$$\limsup_{t\to\infty} \|(\tilde{\mathbf{x}}(t), \tilde{\mathbf{p}}(t))\| < \infty.$$

Then,

$$\lim_{t\to\infty} \|(\hat{\mathbf{x}}(t) - \tilde{\mathbf{x}}(t), \hat{\mathbf{p}}(t) - \tilde{\mathbf{p}}(t))\| = 0. \tag{3.43}$$

Remark 7. This turnpike result is very much in the spirit of the turnpike theory of McKenzie (1976) since it is established for a nonconstant turnpike. The special case of constant turnpikes will be considered in more detail in Sect. 4.6.

4.5 Conditions for SDCCA

The following lemma shows how the central assumption SDCCA can be checked on the data of the differential game.

Lemma 3. *Assume $L_j(\mathbf{x}, z_j)$ is concave in (x_j, z_j) and assume that the total reward function $\sum_{j\in M} L_j(\mathbf{x}, z_j)$ is diagonally strictly concave in $(\mathbf{x}, \mathbf{z})$, i.e., satisfies*

$$\sum_{j\in M} [(z_j^1 - z_j^0)'(\zeta_j^1 - \zeta_j^0) + (x_j^1 - x_j^0)'(\eta_j^1 - \eta_j^0)] > 0, \tag{3.44}$$

for all

$$\eta_j^1 \in \partial_{x_j} L_j(\mathbf{x}^1, z_j^1) \quad \zeta_j^1 \in \partial_{z_j} L_j(\mathbf{x}^1, z_j^1),$$

$$\eta_j^0 \in \partial_{x_j} L_j(\mathbf{x}^0, z_j^0) \quad \zeta_j^0 \in \partial_{z_j} L_j(\mathbf{x}^0, z_j^0).$$

Then Assumption 2 holds true as it follows directly from the concavity of the $L_j(\mathbf{x}, z_j)$ *in* (x_j, z_j) *and* (3.44).

When the functions $L_j(\mathbf{x}, z_j)$ are smooth, explicit conditions for the total reward function, $\sum_{j \in M} L_j(\mathbf{x}, z_j)$, to be diagonally strictly concave in $(\mathbf{x}, \mathbf{z})$ are given in Rosen (1965).[13]

4.6 The Steady-State Equilibrium

The optimality conditions (3.34) and (3.35) define an autonomous Hamiltonian system, and the possibility arises that there exists a *steady-state equilibrium*. That is, a pair $(\bar{\mathbf{x}}, \bar{\mathbf{p}}) \in \mathbb{R}^n \times \mathbb{R}^n$ that satisfies

$$0 \in \partial_{p_j} H_j(\bar{\mathbf{x}}, \bar{p}_j),$$
$$0 \in -\partial_{x_j} H_j(\bar{\mathbf{x}}, \bar{p}_j).$$

When a unique steady-state equilibrium exists, the turnpike properties discussed above provide conditions for when the pair $(\bar{\mathbf{x}}, \bar{\mathbf{p}})$ becomes an attractor for all bounded (overtaking) equilibria.

For a steady-state equilibrium $(\bar{\mathbf{x}}, \bar{\mathbf{p}})$, the strong diagonal support property for $(\bar{\mathbf{x}}, \bar{\mathbf{p}})$ holds if for each $\epsilon > 0$, there exists $\delta > 0$ so that whenever $\|\mathbf{x} - \bar{\mathbf{x}}\| + \|\mathbf{p} - \bar{\mathbf{p}}\| > \epsilon$, one has

$$\sum_{j \in M} \left[\left(p_j - \bar{p}_j\right)' \pi_j + \left(x_j - \bar{x}_j\right)' \xi_j \right] > \delta, \tag{3.45}$$

for all $j \in M$ and pairs $\left(\pi_j, \xi_j\right)$ satisfying

$$\pi_j \in \partial_{p_j} H_j \left(\mathbf{x}, p_j\right) \text{ and } \xi_j \in -\partial_{x_j} H_j \left(\mathbf{x}, p_j\right).$$

With this strong diagonal support property, the following theorem can be proved (see Carlson and Haurie 1995).

Theorem 6. *Assume that* $(\bar{\mathbf{x}}, \bar{\mathbf{p}})$ *is a unique steady-state equilibrium that has the strong diagonal support property given in* (3.45)*. Then, for any* M*-program* $\mathbf{x}(\cdot)$ *with an associated* M*-costate trajectory* $\mathbf{p}(\cdot)$ *that satisfies*

$$\limsup_{t \to \infty} \| \left(\mathbf{x}(t), \mathbf{p}(t)\right) \| < \infty,$$

[13] In Theorem 6, page 528 (for an explanation of the terminology, please see pp. 524–528). See also Sect. 2.3.

one has

$$\lim_{t\to\infty} \| (\mathbf{x}(t) - \bar{\mathbf{x}}, \mathbf{p}(t) - \bar{\mathbf{p}}) \| = 0.$$

Remark 8. The above results extend the classical asymptotic turnpike theory to a dynamic game framework with separated dynamics. The fact that the players interact only through the state variables and not the control ones is essential. An indication of the increased complexities of coupled state and control interactions may be seen in Haurie and Leitmann (1984).

4.7 Existence of Equilibria

In this section, Rosen's approach is extended to show existence of equilibria for the class of games considered, under sufficient smoothness and compactness conditions. Basically, the existence proof is reduced to a fixed-point argument for a point-to-set mapping constructed from an associated class of infinite horizon concave optimization problems.

4.8 Existence of Overtaking Equilibria in the Undiscounted Case

The proof of existence of an overtaking equilibrium for undiscounted dynamic competitive processes uses extensively sufficient overtaking optimality conditions for single-player optimization problems (see Carlson et al. 1991, Chap. 2). For this appeal to sufficiency conditions, the existence of a bounded attractor to all good trajectories is important. This is the reason why this existence theory is restricted to autonomous systems, for which a steady-state equilibrium provides such an attractor.

Remark 9. The existence of overtaking optimal control for autonomous systems (in discrete or continuous time) can be established through a reduction to finite reward argument (see, e.g., Carlson et al. 1991). There is a difficulty in extending this approach to the case of dynamic open-loop games. It comes from the inherent time-dependency introduced by the other players' decisions. One circumvents this difficulty by implementing a reduction to finite rewards for an associated class of infinite horizon concave optimization problems.

Assumption 3. *There exists a unique steady-state equilibrium* $\bar{\mathbf{x}} \in \mathbb{R}^n$ *and a corresponding constant* M*-costate trajectory* $\bar{\mathbf{p}} \in \mathbb{R}^n$ *satisfying*

$$0 \in \partial_{p_j} H_j(\bar{\mathbf{x}}, \bar{p}_j),$$
$$0 \in \partial_{x_j} H_j(\bar{\mathbf{x}}, \bar{p}_j).$$

Assumption 4. *There exists* $\epsilon_0 > 0$ *and* $S > 0$ *such that for any* $\tilde{\mathbf{x}} \in \mathbb{R}^n$ *satisfying* $\|\tilde{\mathbf{x}} - \bar{\mathbf{x}}\| < \epsilon_0$*, there exists an* M*-program* $w(\tilde{\mathbf{x}}, \cdot)$ *defined on* $[0, S]$ *such that* $w(\tilde{\mathbf{x}}, 0) = \tilde{\mathbf{x}}$ *and* $w(\tilde{\mathbf{x}}, S) = \bar{\mathbf{x}}$.

This assumption, which is a controllability assumption, states that in a neighborhood of the steady-state equilibrium, $\bar{\mathbf{x}}$, the system can be driven to the steady state (i.e., turnpike) in a uniform length of time (in this case S) and still maintain the uncoupled constraints. In order to achieve the existence result, one must assure that all admissible M-trajectories lie in a compact set. Additionally, one must assume that their rates of growth are not too large. Thus, the following additional assumption is required.

Assumption 5. *For each* $j \in M$ *there exists a closed bounded set*

$$X_j \subset \mathbb{R}^{n_j} \times \mathbb{R}^{n_j},$$

such that each M*-program,* $\mathbf{x}(\cdot)$ *satisfies* $\left(x_j(t), \dot{x}_j(t)\right) \in X_j$ *a.e.* $t \geq 0$.

Introduce the following:

- Let Ω denote the set of all M-trajectories that start at $\mathbf{x}^o$ and converge to $\bar{\mathbf{x}}$, the unique steady-state equilibrium.
- Define the family of functionals $\theta^T : \Omega \times \Omega \to \mathbb{R}$, $T \geq 0$, by the formula

$$\theta^T(\mathbf{x}(\cdot), \mathbf{y}(\cdot)) \doteq \int_0^T \left[\sum_{j \in M} L_j([\mathbf{x}^{(j)}(t), y_j(t)], \dot{y}_j(t)) \right] dt.$$

The set Ω can be viewed as a subset of all bounded continuous functions in $\mathbb{R}^n$ endowed with the topology of uniform convergence on bounded intervals.

Definition 7. Let $\mathbf{x}(\cdot), \mathbf{y}(\cdot) \in \Omega$. One says that $\mathbf{y}(\cdot) \in \Gamma(\mathbf{x}(\cdot))$ if

$$\liminf_{T \to \infty} \left(\theta^T(\mathbf{x}(\cdot), \mathbf{y}(\cdot)) - \theta^T(\mathbf{x}(\cdot), \mathbf{z}(\cdot))\right) \geq 0,$$

for all M-trajectories $\mathbf{z}(\cdot)$ such that $z_j(0) = x_j^0$, $j \in M$. That is, $\mathbf{y}(\cdot)$ is an overtaking optimal solution of the infinite horizon optimization problem whose objective functional is defined by $\theta^T(\mathbf{x}(\cdot), \cdot)$. Therefore, $\Gamma(\mathbf{x}(\cdot))$ can be viewed as the set of *optimal responses* by all players to an M-program $\mathbf{x}(\cdot)$.

One can then prove the following theorem (see Carlson and Haurie (1995) for a long proof):

Theorem 7. *Under Assumptions 3, 4, and 5, there exists an overtaking equilibrium for the infinite horizon dynamic game.*

4.9 Uniqueness of Equilibrium

In Rosen's paper (1965), the strict diagonal concavity condition was introduced essentially in order to have uniqueness of equilibria. In Sect. 4.6, a similar assumption has been introduced to get asymptotic stability of equilibrium trajectories. Indeed, this condition also leads to uniqueness.[14]

Theorem 8. *Suppose that the assumptions of Theorem 6 hold. Then, there exists an overtaking equilibrium at* $\mathbf{x}^o$*, and it is unique.*

5 Coupled State Constraints

One proceeds now to the extension of the concept of *normalized equilibria*, also introduced by Rosen in (1965), to the infinite horizon differential game framework; this is for differential games with *coupled state constraints*.[15]

5.1 The Model and Basic Hypotheses

The game considered in this section appends a coupled state constraint to the general model considered above. That is, as before the accumulated reward function up to time T takes the form

$$\phi_j^T(\mathbf{x}(\cdot)) = \int_0^T L_j(\mathbf{x}(t), \dot{x}_j(t))\, dt, \tag{3.46}$$

where the integrands $L_j : \mathbb{R}^n \times \mathbb{R}^{n_j} \to \mathbb{R}$ are continuous in all variables and concave and continuously differentiable in the variables (x_j, z_j), and in addition it is assumed that an M-program $\mathbf{x}(\cdot) : [0, +\infty) \to \mathbb{R}^n$ satisfies the constraints

$$\mathbf{x}(0) = \mathbf{x}_0 \in \mathbb{R}^n, \tag{3.47}$$

$$\big(x_j(\mathrm{t}), \dot{x}_j(\mathrm{t})\big) \in X_j \subset \mathbb{R}^{2n_j} \quad \text{a.e. } t \geq 0,\ j \in M, \tag{3.48}$$

$$h_l(\mathbf{x}(t)) \geq 0 \quad \text{for } t \geq 0,\ l = 1, 2, \ldots k, \tag{3.49}$$

[14] See Carlson and Haurie (1995).

[15] The usefulness of the concept has been shown in Haurie (1995) and Haurie and Zaccour (1995) for the analysis of the policy coordination of an oligopoly facing a global environmental constraint.

in which the sets X_j, $j \in M$ are convex and compact and the functions $h_l : \mathbb{R}^n \to \mathbb{R}$ are continuously differentiable and concave, $l = 1, 2, \ldots, k$. One refers to the problem described by (3.46), (3.47), and (3.48) as the uncoupled constraint problem (UCP) and to the problem described by (3.46), (3.47), (3.48), and (3.49) as the coupled constraint problem (CCP). The consideration of a pointwise state constraint (3.49) complicates significantly the model and the characterization of an optimal (equilibrium) solution. However, in the realm of environmental management problems, the satisfaction of a global pollution constraint is often a "long term" objective of the regulating agency instead of a strict compliance to a standard at all points of time. This observation motivates the following definition of a relaxed version of the coupled constraint for an infinite horizon game (the notation meas $[\cdot]$ is used to denote Lebesgue measure).

Definition 8. An M-program $\mathbf{x}(\cdot)$ (see Definition 9 below) is said to asymptotically satisfy the coupled state constraint (3.49) if the following holds: for each $\varepsilon > 0$, there exists a number $B(\varepsilon) > 0$ such that

$$\text{meas}\,[\{t \geq 0 : h_l(\mathbf{x}(t)) < -\varepsilon\}] < B(\varepsilon) \text{ for } l = 1, 2, \ldots k. \tag{3.50}$$

Remark 10. This definition is inspired from the turnpike theory in optimal economic growth (see, e.g., Carlson et al. 1991). As in the case of a turnpike, for any $\varepsilon > 0$ the asymptotically admissible M-program will spend most of its journey in the ε-vicinity of the admissible set.

Definition 9. An M-program $\mathbf{x} : [0, +\infty) \to \mathbb{R}^n$ is said to be

1. Admissible for the uncoupled constrained equilibrium problem (UCP) if $\mathbf{x}(\cdot)$ is locally absolutely continuous, $\phi_j^T(\mathbf{x}(\cdot))$ is finite for all $T \geq 0$, and the constraints (3.47) and (3.48) hold.
2. Asymptotically admissible for the coupled constrained equilibrium problem (CCP) if it is admissible for UCP and asymptotically satisfies the coupled constraint (3.49).
3. A pointwise admissible trajectory for CCP if $\mathbf{x}(\cdot)$ is locally absolutely continuous, $\phi_j^T(\mathbf{x}(\cdot))$ is finite for all $T \geq 0$ and the constraints (3.47), (3.48), and (3.49) hold.

Remark 11. When no confusion arises, one refers to only admissible M-programs with no explicit reference to the uncoupled constraint or coupled constraint problem.

Definition 10. Let $\mathbf{x}^*(\cdot) : [0, +\infty) \to \mathbb{R}^n$ be an admissible M-program. For $j = 1, 2, \ldots m$, one defines the set of admissible responses

$$A_j(\mathbf{x}^*(\cdot)) = \left\{ y_j : [0, +\infty) \to \mathbb{R}^{n_j} : \quad [\mathbf{x}^{*(j)}(\cdot), y_j(\cdot)] \text{ is admissible} \right\},$$

where $[\mathbf{x}^{*(j)}(\cdot), y_j(\cdot)] = (x_1^*(\cdot), x_2^*(\cdot), \ldots, x_{j-1}^*(\cdot), y_j(\cdot), x_{j+1}^*(\cdot), \ldots, x_p^*(\cdot))$.

Finally, the two following types of equilibria are considered in this section.

Definition 11. An admissible M-program $\mathbf{x}^*(\cdot)$ is called

1. An *overtaking Nash equilibrium* if for every $\epsilon > 0$, $j = 1, 2, \ldots m$, and $y_j \in A_j(\mathbf{x}^*(\cdot))$ there exists $T_j = T(j, \epsilon, y) > 0$ such that

$$\phi_j^T(\mathbf{x}^*(\cdot)) > \phi_j^T([\mathbf{x}^{*(j)(\cdot)}, y_j(\cdot)]) - \epsilon, \tag{3.51}$$

for all $T \geq T_j$ (or equivalently,

$$\limsup_{T\to\infty} \left[\phi_j^T([\mathbf{x}^{*(j)}(\cdot), y_j(\cdot)]) - \phi_j^T(\mathbf{x}^*(\cdot))\right] \leq 0, \tag{3.52}$$

for all $y_j \in A_j(\mathbf{x}^*(\cdot))$.

2. An *average reward optimal Nash equilibrium* if for $j = 1, 2, \ldots m$, one has

$$\limsup_{T\to\infty} \frac{1}{T}\left[\phi_j^T([\mathbf{x}^{*(j)}(\cdot), y_j(\cdot)]) - \phi_j^T(\mathbf{x}^*(\cdot))\right] \leq 0, \tag{3.53}$$

for all $y_j \in A_j(\mathbf{x}^*(\cdot))$.

Remark 12. If $\mathbf{x}^*(\cdot)$ is an overtaking Nash equilibrium, then it is easy to see that it is an average reward Nash equilibrium as well. To see this, we note that by (3.51) for all $T \geq T_j$, one has

$$\frac{1}{T}\left[\phi_j^T([\mathbf{x}^{*(j)}(\cdot), y_j(\cdot)]) - \phi_j^T(\mathbf{x}^*(\cdot))\right] < \frac{\epsilon}{T},$$

which, upon taking the limit superior, immediately implies (3.53) holds.

5.2 The Steady-State Normalized Equilibrium Problem

To describe the associated steady-state game, define the m scalar functions $\Phi_j : \mathbb{R}^n \to \mathbb{R}$ by the formulae

$$\Phi_j(\mathbf{x}) = L_j(\mathbf{x}, 0), \tag{3.54}$$

where, as usual, $\mathbf{x} = (x_1, x_2, \ldots x_m)$ and consider the concave game of finding a Nash equilibrium for the reward functions Φ_j subject to the constraints

$$(x_j, 0) \in X_j \quad j = 1, 2, \dots m, \tag{3.55}$$

$$h_l(\mathbf{x}) \geq 0. \quad l = 1, 2, \dots k, \tag{3.56}$$

This is a static game of Rosen studied in Sect. 2. In the remainder of this chapter, it will be shown that the solution of the steady-state game, in particular the vector of Karush-Kuhn-Tucker multipliers, can be used to define an equilibrium solution, in a weaker sense, to the dynamic game.

5.3 Equilibria for the Coupled Dynamic Game

Introduce the following additional assumptions:

Assumption 6. *There exists a fixed* $\mathbf{r} = (r_1, r_2, \dots, r_m) \in \mathbb{R}^n$*, with* $r_j > 0$ *so that the associated steady-state game has a unique normalized Nash equilibrium, denoted by* $\bar{\mathbf{x}}$*, with multipliers* $\mu_j = \mu_0 / r_j$*.*

Remark 13. Conditions under which this assumption holds were given in the previous section. An equivalent form for writing this condition is that there exists a unique normalized steady-state equilibrium, say $\bar{\mathbf{x}}$, a Lagrange multiplier $\mu_0 \in \mathbb{R}^k$, and an associated multiplier, say $\bar{\mathbf{p}} \in \mathbb{R}^n$, satisfying

$$\begin{aligned}
0 &\in \partial_{p_j} \widetilde{H}_j(\bar{\mathbf{x}}, \bar{p}_j; \mu_0), \\
0 &\in \partial_{x_j} \widetilde{H}_j(\bar{\mathbf{x}}, \bar{p}_j; \mu_0), \\
\mu_0' h(\bar{\mathbf{x}}) &= 0, \\
h(\bar{\mathbf{x}}) &\leq 0, \\
\mu_0 &\geq 0,
\end{aligned}$$

in which

$$\begin{aligned}
\widetilde{H}_j(\mathbf{x}, p_j; \mu_0) &\doteq \sup_z \left\{ p_j' z + L_j(\mathbf{x}, z) + \frac{1}{r_j} \sum_{l=1}^{k} \mu_{0l} h_l(\mathbf{x}) \right\}, \\
&= \sup_z \left\{ p_j' z + L_j(\mathbf{x}, z) \right\} + \frac{1}{r_j} \sum_{l=1}^{k} \mu_{0l} h_l(\mathbf{x}), \\
&= H_j(\mathbf{x}, p_j) - \frac{1}{r_j} \sum_{l=1}^{k} \mu_{0l} h_l(\mathbf{x}),
\end{aligned}$$

where the supremum above is over all $z \in \mathbb{R}^{n_j}$ for which $(x_j, z_j) \in X_j$.

In addition to the above assumptions, one must also assume the controllability assumption 4 in a neighborhood of the unique, normalized, steady-state equilibrium $\bar{\mathbf{x}}$.

With these assumptions in place, one considers an associated uncoupled dynamic game that is obtained by adding to the accumulated rewards (3.46) a penalty term, which imposes a cost to a player for violating the coupled constraint. To do this, use the multipliers defining the normalized steady-state Nash equilibrium. That is, consider the new accumulated rewards

$$\tilde{\phi}_j^T(\mathbf{x}(\cdot)) = \int_0^T \left[L_j(\mathbf{x}(t), \dot{x}_j(t)) + \frac{1}{r_j} \sum_{l=1}^{k} \mu_{0l} h_l(\mathbf{x}(t)) \right] dt, \tag{3.57}$$

where one assumes that the initial condition (3.47) and the uncoupled constraints (3.48) are included. Note here that the standard method of Lagrange is not used since a constant multiplier has been introduced instead of a continuous function as is usually done for variational problems. The fact that this is an autonomous infinite horizon problem and the fact that the asymptotic turnpike property holds allow one to use, as a constant multiplier, the one resulting from the solution of the steady-state game and then obtain an asymptotically admissible dynamic equilibrium. Under the above assumptions, the following result is a direct consequence of Theorem 3.1 of Carlson and Haurie (2000).

Theorem 9. *The associated uncoupled dynamic game with costs given by* (3.57) *and constraints* (3.47) *and* (3.48) *has an overtaking Nash equilibrium, say* $\mathbf{x}^*(\cdot)$, *which additionally satisfies*

$$\lim_{t\to\infty} \mathbf{x}^*(t) = \bar{\mathbf{x}}. \tag{3.58}$$

Remark 14. Under an additional strict diagonal concavity assumption, the above theorem insures the existence of a **unique** overtaking Nash equilibrium. The specific hypothesis required is that the combined reward function $\sum_{j=1}^{p} r_j L_j(\mathbf{x}, z_j)$ is diagonally strictly concave in $(\mathbf{x}, \mathbf{z})$, i.e., verifies

$$\sum_{j=1}^{p} r_j \left[(z_j^1 - z_j^0)' \left(\frac{\partial}{\partial_{z_j}} L_j(\mathbf{x}^0, z_j^0) - \frac{\partial}{\partial_{z_j}} L_j(\mathbf{x}^1, z_j^1) \right) \right.$$
$$\left. + (x_j^1 - x_j^0)' \left(\frac{\partial}{\partial_{x_j}} L_j(\mathbf{x}^0, z_j^0) - \frac{\partial}{\partial_{x_j}} L_j(\mathbf{x}^1, z_j^1) \right) \right] > 0, \tag{3.59}$$

for all pairs

$$(\mathbf{x^1}, \mathbf{z^1}) \quad \text{and} \quad (\mathbf{x^0}, \mathbf{z^0}).$$

Remark 15. Observe that the equilibrium M-program guaranteed in the above result is not generally pointwise admissible for the dynamic game with coupled constraints. On the other hand, since it enjoys the turnpike property (3.58), it is an easy matter to see that it is asymptotically admissible for the original coupled dynamic game. Indeed, since the functions $h_l(\cdot)$ are continuous and since (3.58) holds, for every $\varepsilon > 0$, there exists $T(\varepsilon) > 0$ such that for all $t > T(\varepsilon)$ one has

$$h_l(\bar{\mathbf{x}}) - \varepsilon < h_l(\mathbf{x}^*(t)) < h_l(\bar{\mathbf{x}}) + \varepsilon \leq \varepsilon,$$

for all $l = 1, 2, \ldots, k$ since $h_l(\bar{\mathbf{x}}) \leq 0$. Thus, one can take $B(\varepsilon) = T(\varepsilon)$ in (3.50). As a consequence of this fact, one is led directly to investigate whether this M-program is some sort of equilibrium for the original problem. This, in fact, is the following theorem.[16]

Theorem 10. *Under the above assumptions, the overtaking Nash equilibrium for the UCP* (3.47)*,* (3.48)*,* (3.57) *is also an averaging Nash equilibrium for the coupled dynamic game when the coupled constraints* (3.49) *are interpreted in the asymptotic sense as described in Definition* 8*.*

This section is concluded by stating that the long-term average reward for each player for the Nash equilibrium given in the above theorem is precisely the steady-state reward obtained from the solution of the steady-state equilibrium problem. That is, the following result holds.[17]

Theorem 11. *For the averaging Nash equilibrium given by Theorem* 10*, one has*

$$\lim_{T \to +\infty} \frac{1}{T} \int_0^T L_j(\mathbf{x}^*(t), \dot{x}_j^*(t))\, dt = L_j(\bar{\mathbf{x}}, 0).$$

6 A Global Change Game

As an example of the application of the theory presented above, consider the following model à la Ramsey describing an international global climate change control. In this example, the optimal control formalism will be used, instead of the calculus of variations. This will provide the occasion to look at the theory presented above when the model is stated in terms of control variables and state equations instead of trajectories and velocities.

[16] See Carlson and Haurie (2000).

[17] See Carlson and Haurie (2000).

Consider a world formed of m countries, indexed by $j = 1, \dots m$, with stationary population level L_j, stock of productive capital $K_j(t)$, and stock of mitigating capital $M_j(t)$. The productive capital stock evolves according to the following differential equation:

$$\dot{K}_j(t) = I_{K_j}(t) - \alpha_j K_j(t), \quad K_j(0) \text{ given},$$

where $I_{K_j}(t)$ is the investment at t and α_j is the depreciation rate. Similarly, the evolution of mitigation capacity according to

$$\dot{M}_j(t) = I_{M_j}(t) - \beta_j M_j(t), \quad M_j(0) \text{ given},$$

where $I_{M_j}(t)$ is the investment in abatement (mitigation) capacity at t and β_j is the depreciation rate. The GHG emissions are defined as a function $E_j(K_j(t), M_j(t))$, which is increasing in K_j and decreasing in M_j. GHGs accumulate according to the equation

$$\dot{G}(t) = \sum_{j=1}^{m} E_j(K_j(t), M_j(t)) - \delta G(t), \quad G(0) \text{ given}, \tag{3.60}$$

where δ is nature's absorption rate.

The output of the economy is $ELF(G(t)) f_j(K_j(t))$, where the economic loss factor $ELF(\cdot)$ function takes its values between 0 and 1 and is decreasing in G. The output is shared between consumption and investments in the two types of capital, that is,

$$ELF_j(G(t)) f_j(K_j(t)) = C_j(t) + I_{K_j}(t) + I_{M_j}(t).$$

The payoff of country j, $j = 1, \dots m$, over the time interval $[0, T]$, is given by

$$\Phi_j(\cdot) = \int_0^T L_j U_j(C_j(t)/L_j)\, dt,$$

where $U_j(C_j(t)/L_j)$ measures the per capita utility from consumption.

As the differential equation in (3.60) is linear, it can be decomposed into m accumulation equations as follows:

$$\dot{G}_j(t) = E_j(K_j(t), M_j(t)) - \delta G_j(t), \tag{3.61}$$

with

$$G(t) = \sum_{j=1}^{m} G_j(t). \tag{3.62}$$

6.1 The Overtaking Equilibrium for the Infinite Horizon Game

To characterize the equilibrium, one uses the infinite horizon version of the maximum principle. Introduce the pre-Hamiltonian for Player j as

$$\begin{aligned}H_j(K_j, M_j, G_j, I_{K_j}, I_{M_j}, \lambda_{K_j}, \lambda_{M_j}, \lambda_{G_j}) = \\ L_j U_j(C_j/L_j) + \lambda_{K_j}(I_{K_j} - \alpha_j K_j) + \lambda_{M_j}(I_{M_j} - \beta_j M_j) \\ + \lambda_{G_j}(E_j(K_j, M_j) - \delta G_j),\end{aligned}$$

with C_j being given by

$$C_j = ELF(G) f_j(K_j) - (I_{K_j} + I_{M_J}). \tag{3.63}$$

Along an overtaking equilibrium trajectory, the following conditions will hold:

$$\begin{aligned}
0 &= \frac{\partial}{\partial I_{K_j}} H_j(K_j(t), M_j(t), G(t), I_{K_j}(t), I_{M_j}(t), \lambda_{K_j}(t), \lambda_{M_j}(t), \lambda_{G_j}(t)),\\
0 &= \frac{\partial}{\partial I_{M_j}} H_j(K_j(t), M_j(t), G(t), I_{K_j}(t), I_{M_j}(t), \lambda_{K_j}(t), \lambda_{M_j}(t), \lambda_{G_j}(t)),\\
\dot{\lambda}_{K_j}(t) &= -\frac{\partial}{\partial K_j} H_j(K_j(t), M_j(t), G(t), I_{K_j}(t), I_{M_j}(t), \lambda_{K_j}(t), \lambda_{M_j}(t), \lambda_{G_j}(t)),\\
\dot{\lambda}_{M_j}(t) &= -\frac{\partial}{\partial M_j} H_j(K_j(t), M_j(t), G(t), I_{K_j}(t), I_{M_j}(t), \lambda_{K_j}(t), \lambda_{M_j}(t), \lambda_{G_j}(t)),\\
\dot{\lambda}_{G_j}(t) &= -\frac{\partial}{\partial G} H_j(K_j(t), M_j(t), G(t), I_{K_j}(t), I_{M_j}(t), \lambda_{K_j}(t), \lambda_{M_j}(t), \lambda_{G_j}(t)),\\
\dot{K}_j(t) &= I_{K_j}(t) - \alpha_j K_j(t),\\
\dot{M}_j(t) &= I_{M_j}(t) - \beta_j M_j(t),\\
\dot{G}(t) &= \sum_{j=1}^{m} E_j(K_j(t), M_j(t)) - \delta G(t).
\end{aligned}$$

The equilibrium steady state satisfies the following algebraic equations:

$$\begin{aligned}
0 &= \frac{\partial}{\partial I_{K_j}} H_j(\bar{K}_j, \bar{M}_j, \bar{G}, \bar{I}_{K_j}, \bar{I}_{M_j}, \bar{\lambda}_{K_j}, \bar{\lambda}_{M_j}, \bar{\lambda}_{G_j})\\
0 &= \frac{\partial}{\partial I_{M_j}} H_j(\bar{K}_j, \bar{M}_j, \bar{G}, \bar{I}_{K_j}, \bar{I}_{M_j}, \bar{\lambda}_{K_j}, \bar{\lambda}_{M_j}, \bar{\lambda}_{G_j})\\
0 &= -\frac{\partial}{\partial K_j} H_j(\bar{K}_j, \bar{M}_j, \bar{G}, \bar{I}_{K_j}, \bar{I}_{M_j}, \bar{\lambda}_{K_j}, \bar{\lambda}_{M_j}, \bar{\lambda}_{G_j})
\end{aligned}$$

$$0 = -\frac{\partial}{\partial M_j} H_j(\bar{K}_j, \bar{M}_j, \bar{G}, \bar{I}_{K_j}, \bar{I}_{M_j}, \bar{\lambda}_{K_j}, \bar{\lambda}_{M_j}, \bar{\lambda}_{G_j})$$

$$0 = -\frac{\partial}{\partial G} H_j(\bar{K}_j, \bar{M}_j, \bar{G}, \bar{I}_{K_j}, \bar{I}_{M_j}, \bar{\lambda}_{K_j}, \bar{\lambda}_{M_j}, \bar{\lambda}_{G_j})$$

$$0 = \bar{I}_{K_j} - \alpha_j \bar{K}_j$$

$$0 = \bar{I}_{M_j} - \beta_j \bar{M}_j$$

$$0 = \sum_{j=1}^{m} E_j(\bar{K}_j, \bar{M}_j) - \delta \bar{G}.$$

The uniqueness of this steady-state equilibrium is guaranteed if the combined Hamiltonian $\sum_{j=1}^{m} H_j(K_j, M_j, G_j, I_{K_j}, I_{M_j})$ is diagonally strictly concave. This is the case, for this example, if the function $\sum_{j=1}^{m} ELF_j(G(t)) f_j(K_j)$ is strictly concave and the functions $E_j(K_j, M_j)$, $j = 1, \dots, m$, are convex.

Then, the steady-state equilibrium is an attractor for all overtaking equilibrium trajectories, emanating from different initial points.

6.2 Introduction of a Coupled State Constraint

Suppose that the cumulative GHG stock $G(t)$ is constrained to remain below a given threshold $\widetilde{G}$, at any time, that is,

$$G(t) \le \widetilde{G}, \quad \forall t. \tag{3.64}$$

This is a coupled state constraint, and the extended Hamiltonian is defined as follows:

$$\widetilde{H}_j(K_j, M_j, G, I_{K_j}, I_{M_j}, \lambda_{K_j}, \lambda_{M_j}, \lambda_{G_j}, \mu) = L_j U_j(C_j/L_j) + \lambda_{K_j}(I_{K_j} - \alpha_j K_j) + \lambda_{M_j}(I_{M_j} - \beta_j M_j) + \lambda_{G_j}(E_j(K_j, M_j) - \delta G_j) + \frac{\mu}{r_j}(\widetilde{G} - G).$$

The necessary conditions for a normalized equilibrium with weights $r_j \ge 0$ are as follows:

$$0 = \frac{\partial}{\partial I_{K_j}} \widetilde{H}_j(K_j(t), M_j(t), G(t), I_{K_j}(t), I_{M_j}(t), \lambda_{K_j}(t), \lambda_{M_j}(t), \lambda_{G_j}(t), \mu(t))$$

$$0 = \frac{\partial}{\partial I_{M_j}} \widetilde{H}_j(K_j(t), M_j(t), G(t), I_{K_j}(t), I_{M_j}(t), \lambda_{K_j}(t), \lambda_{M_j}(t), \lambda_{G_j}(t), \mu(t))$$

$$
\begin{aligned}
\dot{\lambda}_{K_j}(t) &= -\frac{\partial}{\partial K_j}\widetilde{H}_j(K_j(t), M_j(t), G(t), I_{K_j}(t), I_{M_j}(t), \lambda_{K_j}(t), \lambda_{M_j}(t), \lambda_{G_j}(t), \mu(t)) \\
\dot{\lambda}_{M_j}(t) &= -\frac{\partial}{\partial M_j}\widetilde{H}_j(K_j(t), M_j(t), G(t), I_{K_j}(t), I_{M_j}(t), \lambda_{K_j}(t), \lambda_{M_j}(t), \lambda_{G_j}(t), \mu(t)) \\
\dot{\lambda}_{G_j}(t) &= -\frac{\partial}{\partial G}\widetilde{H}_j(K_j(t), M_j(t), G(t), I_{K_j}(t), I_{M_j}(t), \lambda_{K_j}(t), \lambda_{M_j}(t), \lambda_{G_j}(t), \mu(t)) \\
\dot{K}_j(t) &= I_{K_j}(t) - \alpha_j K_j(t) \\
\dot{M}_j(t) &= I_{M_j}(t) - \beta_j M_j(t) \\
\dot{G}(t) &= \sum_{j=1}^{m} E_j(K_j(t), M_j(t)) - \delta G(t)
\end{aligned}
$$

and the complementary slackness condition for the multiplier $\mu(t)$ and the constraint $0 \leq \widetilde{G} - G(t)$ are given by

$$
\begin{aligned}
&0 = \mu(t)(\widetilde{G} - G(t)), \\
&0 \leq \widetilde{G} - G(t), \quad 0 \leq \mu(t).
\end{aligned}
$$

As these complementary conditions introduce a discontinuity in the right-hand side of some of the differential equations above, they will complicate the search for the overtaking normalized equilibrium trajectory in our control or calculus of variations formalisms. This difficulty will disappear if the multiplier is constant over time. Recalling that the steady-state normalized equilibrium is an attractor for the overtaking normalized equilibrium trajectory with coupled state constraint, it becomes attractive to use the constant multiplier $\mu(t) \equiv \bar{\mu}$, where $\bar{\mu}$ is the multiplier's value in the steady-state normalized equilibrium.

To define the steady-state normalized equilibrium associated with weights $r_j \geq 0$, one writes the necessary conditions, which are a set of algebraic equations,

$$0 = \bar{\lambda}_{\bar{K}_j} - U_j'(\bar{C}_j), \tag{3.65}$$

$$0 = \bar{\lambda}_{\bar{M}_j} - U_j'(\bar{C}_j), \tag{3.66}$$

$$0 = -U_j'(\bar{C}_j) ELF_j(\bar{G}) f_j'(\bar{K}_j) + \alpha_j \bar{\lambda}_{K_j}, \tag{3.67}$$

$$0 = -\bar{\lambda}_{\bar{G}_j} \partial_{M_j}(\bar{E}_j(\bar{K}_j, \bar{M}_j)) + \beta_j \bar{\lambda}_{M_j}, \tag{3.68}$$

$$0 = -ELF(\bar{G})' f_j(\bar{K}_j) - \bar{\lambda}_{\bar{G}_j} + \delta \bar{\lambda}_{\bar{G}_j} + \frac{\bar{\mu}}{r_j}, \tag{3.69}$$

$$0 = \bar{\mu}(\widetilde{G} - \bar{G}), \quad 0 \leq (\widetilde{G} - \bar{G}), \quad 0 \leq \bar{\mu}. \tag{3.70}$$

Under the diagonal strict concavity assumption, this normalized equilibrium is unique and is an attractor for the uncoupled game, where the constant $\bar{\mu}/r_j$ is used to penalize the non-satisfaction of the constraint by player j.

Condition (3.65) states that in equilibrium, the marginal utility of consumption is equal to its marginal cost, which is measured here by the shadow price of production capacity. Conditions (3.65) and (3.66) imply that the shadow prices of both capitals, that is, production and mitigation, are equal. Conditions (3.65), (3.66), and (3.67) imply $ELF(G)f_j'(K_j) = \alpha_j$, that is, the marginal output value is equal to the depreciation rate of production capital. Equation (3.68) stipulates that the marginal reduction in emissions due to mitigation times the shadow price of pollution stock is equal to the shadow price of mitigation capacity times the depreciation rate. Substituting for $ELF(G)f_j'(K_j) = \alpha_j$ in condition (3.69) leads to

$$\lambda_{G_j}(1-\delta) = \frac{\bar{\mu}}{r_j} - \alpha_j.$$

It is easy to see that the shadow cost value of the pollution stock depends on the status of the coupling constraint, which in turn would affect the other equilibrium conditions.

7 The Discounted Case

As in most economic models formulated over an infinite time horizon, the costs or profits are discounted, there is a need to also develop the theory in this case. For the most part, the focus here is to describe the differences between the theory in the undiscounted case discussed above and the discounted case.

7.1 The Discounted Model and Optimality Conditions

In the discounted case, the accumulated reward for player j, $j = 1, 2, \ldots, m$ up to time T is given by the integral functional

$$\tilde{\phi}_j^T(\mathbf{x}(\cdot)) = \int_0^T e^{-\rho_j t} L_j(\mathbf{x}(t), \dot{x}_j(t))\, dt, \tag{3.71}$$

where, for $j = 1, 2, \ldots, m$, $\rho_j > 0$ is the constant discount rate for player j and $L_j : \mathbb{R}^n \times \mathbb{R}^{n_j} \to \mathbb{R}$ satisfies the concavity and smoothness assumptions given for the undiscounted problem. All of the other constraints, both coupled and uncoupled, remain the same. The advantage of introducing the discount rate is that as $T \to \infty$ mere boundedness conditions ensures that the accumulated rewards of the players are convergent improper integrals. As a result, the analysis does not require the overtaking equilibria concept, and so we take the usual definition of a Nash equilibrium.

Definition 12. An M-program $\mathbf{x}^*(\cdot) : [0, +\infty) \to \mathbb{R}^n$ is a Nash equilibrium for the discounted problem at $\mathbf{x}^o \in \mathbb{R}$ if the following conditions are met:

(a) $\mathbf{x}^*(0) = \mathbf{x}^o$ and the improper integrals

$$\lim_{T\to\infty} \tilde{\phi}_j^T(\mathbf{x}^*(\cdot)) = \int_0^\infty e^{-\rho_j t} L_j(\mathbf{x}^*, \dot{x}_j^*(t))\, dt, \tag{3.72}$$

are finite for each $j = 1, 2, \ldots, m$.

(b) For each M-program $\mathbf{x}(\cdot)$ satisfying $\mathbf{x}(0) = \mathbf{x}^o$, the condition

$$\lim_{T\to\infty} \tilde{\phi}_j^T([\mathbf{x}^{*-j}(\cdot), x_j(\cdot)]) \le \int_0^\infty e^{-\rho_j t} L_j(\mathbf{x}^*, \dot{x}_j^*(t))\, dt, \tag{3.73}$$

is satisfied.

The introduction of the discount factors $e^{-\rho_j t}$ changes the problem from an autonomous problem into a nonautonomous problem, which can complicate the dynamics of the resulting optimality system (3.34) and (3.35) in that now the Hamiltonian depends on time.[18] By factoring out the discount factor, it is possible to again return to an autonomous dynamical system, which is similar to the autonomous pseudo-Hamiltonian system with a slight perturbation. This *modified pseudo-Hamiltonian system* takes the form

$$\dot{x}_j(t) \in \partial_{p_j} H_j(\mathbf{x}(t), p_j(t)), \tag{3.74}$$

$$\dot{p}_j(t) + \rho_j\, p_j(t) \in -\partial_{x_j} H_j(\mathbf{x}(t), p_j(t)), \tag{3.75}$$

for all $j \in M$, and the corresponding steady-state equilibrium $(\bar{\mathbf{x}}, \bar{\mathbf{p}})$ satisfies

$$0 \in \partial_{p_j} H_j(\bar{\mathbf{x}}, \bar{p}_j), \tag{3.76}$$

$$\rho_j\, \bar{p}_j \in -\partial_{x_j} H_j(\bar{\mathbf{x}}, \bar{p}_j), \tag{3.77}$$

for all $j \in M$. With this notation the discounted turnpike property can be considered.

7.2 The Turnpike Property with Discounting

The introduction of the discount factors requires a modification to the support property given earlier by introducing a so-called "curvature condition" into the support property. This terminology has been introduced by Rockafellar in (1976), and it indicates the "amount" of strict concavity-convexity needed to obtain GAS when discounting is introduced. A way to check this property, when the functions L_j are autonomous is proposed below.

[18] That is, $H_j(t, \mathbf{x}, p_j) = \sup_{z_j}\{e^{-\rho_j t} L_j(\mathbf{x}, z_j) + p_j' z_j\}$.

Definition 13. Let $(\bar{\mathbf{x}}, \bar{\mathbf{p}})$ be a steady-state equilibrium. The strong diagonal support property for $(\bar{\mathbf{x}}, \bar{\mathbf{p}})$ holds relative to the discount rates ρ_j, $j \in M$, if for each $\epsilon > 0$ there exists $\delta > 0$ so that whenever $\|\mathbf{x} - \bar{\mathbf{x}}\| + \|\mathbf{p} - \bar{\mathbf{p}}\| > \epsilon$, one has

$$\sum_{j \in M} \left[\left(p_j - \bar{p}_j\right)' \pi_j + \left(x_j - \bar{x}_j\right)' \left(\xi_j - \rho_j \bar{p}_j\right) \right] > \delta + \sum_{j \in M} \rho_j \left(x_j - \bar{x}_j\right)' \left(p_j - \bar{p}_j\right), \tag{3.78}$$

for all $j \in M$ and pairs $\left(\pi_j, \xi_j\right)$ satisfying

$$\pi_j \in \partial_{p_j} H_j \left(\mathbf{x}, p_j\right) \text{ and } \xi_j \in -\partial_{x_j} H_j \left(\mathbf{x}, p_j\right).$$

With this notation, the following analogue of Theorem 6 holds.[19]

Theorem 12. *Assume that $(\bar{\mathbf{x}}, \bar{\mathbf{p}})$ is a unique steady-state equilibrium that has the strong diagonal support property given by (3.78). Then, for any M-program $\mathbf{x}(\cdot)$ with an associated M-costate trajectory $\mathbf{p}(\cdot)$ that satisfies*

$$\limsup_{t \to \infty} \| (\mathbf{x}(t), \mathbf{p}(t)) \| < \infty,$$

one has

$$\lim_{t \to \infty} \| (\mathbf{x}(t) - \bar{\mathbf{x}}, \mathbf{p}(t) - \bar{\mathbf{p}}) \| = 0.$$

In the case of an infinite horizon optimal control problem with discount rate $\rho > 0$, Rockafellar (1976) and Brock and Scheinkman (1976) have given easy conditions to verify curvature. Namely, a steady-state equilibrium $\bar{\mathbf{x}}$ is assured to be an attractor of overtaking trajectories by requiring the Hamiltonian of the optimally controlled system to be a-*concave* in x and b-*convex* in p for values of $a > 0$ and $b > 0$ for which the inequality

$$(\rho_j)^2 < 4ab,$$

holds. This idea is extended to the dynamic game case beginning with the following definition.

Definition 14. Let $a = (a_1, a_2, \ldots a_m)$ and $b = (b_1, b_2, \ldots b_m)$ be two vectors in $\mathbb{R}^m$ with $a_j > 0$ and $b_j > 0$ for all $j \in M$. The combined Hamiltonian $\sum_{j \in M} H_j \left(\mathbf{x}, p_j\right)$ is strictly diagonally a-concave in $\mathbf{x}$, b-convex in $\mathbf{p}$ if

$$\sum_{j \in M} \left[H_j(\mathbf{x}, p_j) + \frac{1}{2} \left(a_j \|x_j\|^2 - b_j \|p_j\|^2\right) \right],$$

is strictly diagonally concave in $\mathbf{x}$, convex in $\mathbf{p}$.

[19] For a proof, see Carlson and Haurie (1995).

Theorem 13. *Assume that there exists a unique steady-state equilibrium,* $(\bar{\mathbf{x}}, \bar{\mathbf{p}})$. *Let* $a = (a_1, a_2, \ldots a_m)$ *and* $b = (b_1, b_2, \ldots b_m)$ *be two vectors in* $\mathbb{R}^m$ *with* $a_j > 0$ *and* $b_j > 0$ *for all* $j \in M$, *and assume that the combined Hamiltonian is strictly diagonally* a*-concave in* $\mathbf{x}$, b*-convex in* $\mathbf{p}$. *Further, let* $\mathbf{x}(\cdot)$ *be a bounded equilibrium* M*-program with an associated* M*-costate trajectory* $\mathbf{p}(\cdot)$ *that also remains bounded. Then, if the discount rates* ρ_j, $j \in M$ *satisfy the inequalities*

$$(\rho_j)^2 < 4a_j b_j, \tag{3.79}$$

the M*-program* $\mathbf{x}(\cdot)$ *converges to* $\bar{x}$.

Proof. See Carlson and Haurie (1995).

Remark 16. These results extend the classical discounted turnpike theorem to a dynamic game framework with separated dynamics. As mentioned earlier, the fact that the players interact only through the state variables and not the control ones is essential.

7.3 The Common Discounting Case and Coupled Constraint

Now introduce into the model considered above the coupled constraints given by (3.47), (3.48), and (3.49) under the assumption that all of the discount rates are the same (i.e., $\rho_j = \rho > 0$ for all $j \in M$). The introduction of a common discount rate modifies the approach to the above arguments in several ways.

- The steady-state equilibrium game must be replaced with an *implicit steady-state equilibrium game*.
- Since the discount rate introduces a tendency to delay tough decisions to the future, the means in which the pointwise state constraint is to be satisfied must be modified.
- It is not necessary to consider the weaker notions of overtaking and averaging Nash equilibria since the discount factor forces all of the improper integrals to converge.

The first two points will be discussed first.

7.3.1 The Implicit Steady-State Equilibrium Problem

Introduce the steady-state dynamic equilibrium conditions with positive discounting as follows: assume there exists a steady state $\hat{\mathbf{x}}$, a Lagrange multiplier $\mu_0 \in \mathbb{R}^k$, and an associated costate $\hat{\mathbf{p}} \in \mathbb{R}^n$ satisfying

$$0 \in \partial_{p_j} \widetilde{H}_j(\hat{\mathbf{x}}, \hat{p}_j; \mu_0), \tag{3.80}$$

$$\rho \hat{p}_j \in \partial_{x_j} \widetilde{H}_j(\hat{\mathbf{x}}, \hat{p}_j; \mu_0), \tag{3.81}$$

$$\mu_0' h(\hat{\mathbf{x}}) = 0, \tag{3.82}$$

$$h(\hat{\mathbf{x}}) \geq 0, \tag{3.83}$$

$$\mu_0 \geq 0, \tag{3.84}$$

where $\widetilde{H}_j(\hat{\mathbf{x}}, \hat{p}_j; \mu_0)$ is defined as in (3.57). As indicated in Feinstein and Luenberger (1981) for the control case and by Haurie and Roche (1994) for the differential game case, these steady-state conditions with a positive discount rate can be interpreted as an *implicit equilibrium solution*. More precisely, the vector $\hat{\mathbf{x}}$ is a normalized equilibrium solution for the static game with payoffs

$$L_j(\mathbf{x}, \rho(x_j - \hat{x_j})), \quad \text{where } (x_j, \rho(x_j - \hat{x_j})) \in X_j, \quad j = 1, \ldots, p, \tag{3.85}$$

and coupled constraint $h(\mathbf{x}) \geq 0$. The conditions insuring the uniqueness of such a steady-state implicit equilibrium are not easy to obtain. Indeed, the fixed-point argument that is inherent in the definition is at the origin of this difficulty. One shall, however, assume the following:

Assumption 7. *There exists a fixed vector $\mathbf{r} = (r_1, r_2, \ldots, r_p) \in \mathbb{R}^n$, with $r_j > 0$ so that a unique implicit normalized Nash equilibrium exists, denoted by $\hat{\mathbf{x}}$, with multipliers $\mu_j = (1/r_j)\mu_0$ for the game defined in* (3.85).

With this assumption, as in the undiscounted case, one introduces the associated discounted game with uncoupled constraints by considering the perturbed reward functionals

$$\tilde{\phi}_j^T(\mathbf{x}(\cdot)) = \int_0^T e^{-\rho t}\left[L_j(\mathbf{x}(t), \dot{x}_j(t)) + \frac{1}{r_j}\sum_{l=1}^{k} \mu_{0l} h_l(\mathbf{x}(t))\right] dt, \tag{3.86}$$

with uncoupled constraints

$$(x_j(t), \dot{x}_j(t)) \in X_j \quad j = 1, 2, \ldots, p. \tag{3.87}$$

Theorem 14. *Under Assumption 7 there exists a Nash equilibrium, say $\mathbf{x}^*(\cdot)$, for the associated discounted uncoupled game. In addition, if the combined Hamiltonian*

$$\sum_{j=1}^{m} \widetilde{H}_j(\mathbf{x}, p_j; \mu_0),$$

is sufficiently strictly diagonally concave in $\mathbf{x}$ and convex in $\mathbf{p}$ relative to the discount rate ρ, then this Nash equilibrium, $\mathbf{x}^*(\cdot)$, enjoys the turnpike property. That is,

$$\lim_{t \to +\infty} \mathbf{x}^*(t) = \hat{\mathbf{x}},$$

in which $\hat{\mathbf{x}}$ *is the unique solution of the implicit steady-state game.*

Proof. This result follows directly from Theorems 2.5 and Theorem 3.2 in Carlson and Haurie (1996).

Remark 17. The precise conditions under which the turnpike property in the above result is valid are given in Carlson and Haurie (1996). One such condition would be to have the combined Hamiltonian $\sum_{j=1}^{m} \widetilde{H}_j(\mathbf{x}, p_j; \mu_0)$ be strictly diagonally a-concave in $\mathbf{x}$ b-convex in $\mathbf{p}$ with a and b chosen so that $a_j b_j > \rho^2$ for $j = 1, 2, \ldots m$ (see Definition 14).

The natural question to ask now, as before, is what type of optimality is implied by the above theorem for the original dynamic discounted game. The perturbed cost functionals (3.86) seem to indicate that, with the introduction of the discount rate, it is appropriate to consider the coupled isoperimetric constraint

$$\int_0^{+\infty} e^{-\rho t} h_l(\mathbf{x}(t))\, dt \leq 0 \quad l = 1, 2, \ldots, k. \tag{3.88}$$

However, the multipliers μ_{0l} are not those associated with the coupled isoperimetric constraints (3.88) but those defined by the normalized steady-state implicit equilibrium problem. In the next section, the solution of the auxiliary game with decoupled controls and reward functionals (3.86) is shown to enjoy a weaker dynamic equilibrium property which was called an *implicit Nash equilibrium* in Carlson and Haurie (1996).

7.4 Existence of an Implicit Nash Equilibrium

The introduction of the positively discounted isoperimetric constraints (3.88) enables each of the players to delay costly decisions to the future. This of course would not lead the players to meet the coupled state constraint for most of the time in the future. To address this issue, the following definition of admissibility relative to a trajectory that asymptotically satisfies the constraints as in Definition 8.

Definition 15. Let $\bar{\mathbf{x}}(\cdot)$ be a fixed admissible trajectory for the discounted uncoupled dynamic game and let $\bar{\mathbf{x}}^\infty \in \mathbb{R}^n$ be a constant vector such that:

$$\lim_{t \to +\infty} \bar{\mathbf{x}}(t) = \bar{\mathbf{x}}^\infty,$$
$$h_l(\bar{\mathbf{x}}^\infty) \leq 0 \quad \text{for } l = 1, 2, \ldots, k, \tag{3.89}$$

$$(\bar{x}_j^{\infty}, 0) \in X_j \quad \text{for } j = 1, 2, \dots, p.$$

A trajectory $\mathbf{x}(\cdot)$ is said to asymptotically satisfy the coupled state constraints (3.49) for the discounted dynamic game relative to $\bar{\mathbf{x}}(\cdot)$ if

$$\int_0^{+\infty} e^{-\rho t} h_l(\mathbf{x}(t))\, dt \le \int_0^{+\infty} e^{-\rho t} h_l(\bar{\mathbf{x}}(t))\, dt \quad \text{for } l = 1, 2, \dots, k. \tag{3.90}$$

Remark 18. In the above definition, the coupled constraint is viewed as an isoperimetric-type constraint, where the expression

$$\rho \int_0^{+\infty} e^{-\rho t} h_l(\mathbf{x}(t))\, dt,$$

may be viewed as a "discounted average value" of the pointwise constraint (3.49) over the infinite time interval $[0, +\infty)$. With these terms, a unilateral change from a given asymptotically admissible strategy by any of the players is admissible if it does not exceed the "discounted average value" of the constraint.

Definition 16. An admissible trajectory $\bar{\mathbf{x}}(\cdot)$ is an implicit Nash equilibrium for the discounted dynamic game if there exists a constant vector $\bar{\mathbf{x}}^{\infty}$ so that (3.89) holds, and if for all $y_j(\cdot)$, for which the trajectories $[\bar{\mathbf{x}}^{(j)}(\cdot), y_j(\cdot)]$ asymptotically satisfy the coupled state constraint (3.49) relative to $\bar{\mathbf{x}}(\cdot)$, the following holds:

$$\int_0^{+\infty} e^{-\rho t} L_j(\bar{\mathbf{x}}(t), \dot{\bar{x}}_j(t))\, dt \ge \int_0^{+\infty} e^{-\rho t} L_j([\bar{\mathbf{x}}^{(j)}(t), y_j(t)], \dot{y}_j(t))\, dt.$$

With this definition in hand, the discounted version of Theorem 10 is obtained.[20]

Theorem 15. *Under the assumptions given above, there exists an implicit Nash equilibrium for the discounted coupled dynamic game.*

7.5 Oligopoly Example

To illustrate the theory developed above, consider a simple dynamic economic model of oligopolistic competition under a global environmental constraint. In this game, the quantity produced by a firm corresponds to its production capacity. Each firm emits pollution as a by-product of its production activities, with the level of emissions depending on the installed production capacity and on the available abatement (or pollution control) capacity. Both capacities, that is, production and

[20] See Carlson and Haurie (1995).

abatement, can be increased by costly investment. The product is homogeneous and the firms compete à la Cournot, that is, the market price is a decreasing function of total output (measured by total capacity) put on the market. Each player maximizes his discounted stream of profits. More specifically, the game is described by the following elements:

- The state variable $x_j(t)$ denotes the production capacity of firm j at time t and evolves according to the differential equation

$$\dot{x}_j(t) = I_{xj}(t) - \mu_j x_j(t), \quad x_j(0) \text{ given},$$

  where $I_{xj}(t)$ is the investment rate and μ_j is the depreciation rate.
- The state variable $y_j(t)$ denotes the stock of pollution control capital at time t and evolves over time according to the differential equation

$$\dot{y}_j(t) = I_{yj}(t) - \nu_j y_j(t), \quad y_j(0) \text{ given},$$

  where $I_{yj}(t)$ is the investment rate and μ_j is the depreciation rate.
- The inverse demand function that defines a market clearing price is denoted $D\left(\sum_{j=1,\ldots,m} x_j(t)\right)$.
- The emissions level at time t of firm j is a function $e_j(x_j(t), y_j(t))$, which is increasing in x_j and decreasing in y_j. To illustrate, a possible functional form is $e_j(x, y) = a_j(x/y)^{\alpha_j}$.
- The unit production cost for firm j is constant and denoted c_j.
- The investment cost function (production adjustment cost) is denoted $\gamma_j(\cdot)$, and the cost of investment in pollution control capital is denoted $\delta_j(\cdot)$.
- The instantaneous profit of firm j at time t is given by

$$L_j(\cdot) = \left(D\left(\sum_{j=1}^{m} x_j(t)\right) - c_j\right)x_j(t) - \gamma_j(\dot{x}_j(t) + \mu_j x_j(t)) - \delta_j(\dot{y}_j(t) + \nu_j y_j(t)).$$

The dynamic oligopoly model is then formulated as a differential game with a reward for player j defined over any time interval $[0, T]$ by

$$\tilde{\phi}_j^T(\mathbf{x}(\cdot), \mathbf{y}(\cdot)) = \int_0^T e^{-\rho t} L_j(\mathbf{x}(t), \mathbf{y}(t), \dot{x}_j(t), \dot{y}_j(t))\, dt, \tag{3.91}$$

where ρ is a positive discount rate and with a global environmental constraint formulated as the coupled pointwise state constraint

$$\sum_{j=1}^{m} e_j(x_j(t), y_j(t))) \leq \bar{E}, \quad t \geq 0, \tag{3.92}$$

where $\bar{E}$ denotes some global limit on the pollutant emissions level.

From the theory presented above, one knows that a unique normalized equilibrium exists for each choice of weights $\mathbf{r} \geq 0$, if the total revenue function $D\left(\sum_j x_j\right)\sum_j x_j$ is concave and the functions γ_j and δ_j are convex since this would imply SDCCA. The equilibrium conditions, in a regular case, are obtained as follows: one introduces a common multiplier $\pi(t)$, expressed in current value, for the constraint (3.92) and modifies the payoff of player j as follows:

$$\psi_j^T(\mathbf{x}(\cdot), \mathbf{y}(\cdot), \mathbf{r}) = \int_0^T e^{-\rho t}\Bigg(L_j(\mathbf{x}(t), \mathbf{y}(t), \dot{x}_j(t), \dot{y}_j(t)) + \frac{\pi(t)}{r_j}\left(\bar{E} - \sum_{j=1}^m e_j(x_j(t), y_j(t))\right)\Bigg)\, dt.$$

Then, the corresponding optimality conditions for this constrained game, in the Hamiltonian formulation, become

$$\begin{aligned}
\dot{x}_j(t) &= \partial_{p_j} H_j(\mathbf{x}(t), \mathbf{y}(t), p_j(t), q_j(t)),\\
\dot{y}_j(t) &= \partial_{q_j} H_j(\mathbf{x}(t), \mathbf{y}(t), p_j(t), q_j(t)),\\
\dot{p}_j(t) + \rho p_j(t) &= -H_j(\mathbf{x}(t), \mathbf{y}(t), p_j(t), q_j(t)) + \frac{\pi(t)}{r_j}\partial_{x_j} e_j(x_j(t), y_j(t)),\\
\dot{q}_j(t) + \rho q_j(t) &= -\partial_{y_j} H_j(\mathbf{x}(t), \mathbf{y}(t), p_j(t), q_j(t)) + \frac{\pi(t)}{r_j}\partial_{y_j} e_j(x_j(t), y_j(t)),\\
0 &= \pi(t)\left(\bar{E} - \sum_{j=1}^m e_j(x_j(t), y_j(t))\right),\\
0 &\leq \bar{E} - \sum_{j=1}^m e_j(x_j(t), y_j(t))),\\
0 &\leq \pi(t),
\end{aligned}$$

where $p_j(t)$ and $q_j(t)$ are the costate variables appended to the production capacity and pollution control state equations, respectively, and

$$H_j(\mathbf{x}, \mathbf{y}, p_j, q_j) = \sup_{z_j, w_j} \left\{L_j(\mathbf{x}, \mathbf{y}, z_j, w_j) + p_j z_j + q_j w_j\right\}.$$

This means that the investment controls must satisfy

$$\begin{aligned}
0 &= -\gamma_j'(I_{xj}) + p_j(t),\\
0 &= -\delta_j'(I_{yj}) + q_j(t),
\end{aligned}$$

that is, the familiar rule of marginal cost of investment being equal to the shadow price of the corresponding state variable.

Observe that the above optimality conditions state that if the coupled state constraint is not active (i.e., the inequality in (3.92) becomes strict equality), then the multiplier $\pi(t)$ equals zero; otherwise, the multiplier $\pi(t)$ can assume any positive value. If the coupled constraint is not active at all instants of time, then the multiplier $\pi(t)$ could have discontinuities, which is difficult to deal with. In the approach developed in this chapter, one can escape these possible discontinuities by using the discounted turnpike property and replacing the time-varying multiplier $\pi(t)$ by a constant multiplier obtained by solving for the steady-state equilibrium problem with coupled constraints. That is, to determine the turnpike $(\hat{\mathbf{x}}, \hat{\mathbf{y}})$ and the constant multiplier $\bar{\mu}_0$, one uses conditions (3.80), (3.80), (3.80), and (3.84), which here take the following form:

$$
\begin{aligned}
0 &= \partial_{p_j} H_j(\mathbf{x}, \mathbf{y}, p_j, q_j), \\
0 &= \partial_{q_j} H_j(\mathbf{x}, \mathbf{y}, p_j, q_j), \\
\rho p_j &= -\partial_{x_j} H_j(\mathbf{x}, \mathbf{y}, p_j, q_j), \\
\rho q_j &= -\partial_{y_j} H_j(\mathbf{x}, \mathbf{y}, p_j, q_j), \\
0 &= \mu_0 \left(\bar{E} - \sum_{j=1}^{m} e_j(x_j, y_j)) \right), \\
0 &\leq \bar{E} - \sum_{j=1}^{m} e_j(x_j, y_j)), \\
0 &\leq \mu_0.
\end{aligned}
$$

Then $\bar{\mu}_0$ could be interpreted as a constant emissions tax that would lead the oligopolistic firm j, playing a Nash equilibrium with payoffs including a tax $\frac{\bar{\mu}_0}{r_j}$, to satisfy the emissions constraint in the sense of a " discounted average value" as discussed in Remark 18.

8 Conclusion

In this chapter, a rather complete theory of infinite horizon concave open-loop differential games has been presented, including the case where the players have to satisfy jointly, in equilibrium, a coupled state constraint. It happens that the theory initiated by Rosen (1965) for static concave games, which concerned existence, uniqueness of normalized equilibrium, and stability for a pseudo-gradient path algorithm, can be extended nicely to the context of open-loop differential games played over an infinite time horizon. The strict diagonal concavity assumption, introduced by Rosen to prove his theorems, which is closely related to the monotonicity assumption in the theory of variational inequalities (see Haurie and Marcotte 1985) allows one, when extended to an associated pseudo-Hamiltonian system, to prove existence and uniqueness of an overtaking equilibrium for a given weighting of the players' payoffs in infinite horizon open-loop differential games. These proofs come along with the establishment of a global asymptotic stability property, called the " turnpike" property in economic applications. The turnpike is a steady state for the Hamiltonian system, which becomes an attractor for all equilibrium trajectories, emanating from different possible initial states.

The consideration of an infinite horizon with the turnpike property and the motivating example of an economic growth model with an environmental coupled constraint suggest the introduction of a relaxed asymptotic form of the coupled constraint, permitting a simplification of the computation of extremal trajectories. Under such a reformulation of the game, one can show that the use of the asymptotic steady-state game for the definition of a time invariant multiplier (to be interpreted in the environmental management context as a constant emission tax) permits the construction of an auxiliary decoupled differential game whose Nash equilibrium is also an equilibrium (in a weaker sense) of the original game under the asymptotic coupled state constraint.

References

Arrow KJ, Kurz M (1970) Public investment, the rate of return, and optimal fiscal policy. The Johns Hopkins Press, Baltimore

Brock WA (1977) Differential games with active and passive variables. In: Henn R, Moeschlin O (eds) Mathematical economics and game theory: essays in honor of Oskar Morgenstern. Lecture notes in economics and mathematical systems. Springer, Berlin, pp 34–52

Brock WA, Haurie A (1976) On existence of overtaking optimal trajectories over an infinite time horizon. Math Oper Res 1:337–346

Brock WA, Scheinkman JA (1976) Global asymptotic stability of optimal control systems with application to the theory of economic growth. J Econ Theory 12:164–190

Carlson DA, Haurie A (1995) A turnpike theory for infinite horizon open-loop differential games with decoupled dynamics. In: Olsder GJ (ed) New trends in dynamic games and applications. Annals of the international society of dynamic games, vol 3. Birkhäuser, Boston, pp 353–376

Carlson DA, Haurie A (1996) A turnpike theory for infinite horizon competitive processes. SIAM J Optim Control 34(4):1405–1419

Carlson DA, Haurie A (2000) Infinite horizon dynamic games with coupled state constraints. In: Filar JA, Gaitsgory V, Misukami K (eds) Advances in dynamic games and applications. Annals of the international society of dynamic games, vol 5. Birkhäuser, Boston, pp 195–212

Carlson DA, Haurie A, Leizarowitz A (1991) Infinite horizon optimal control: deterministic and stochastic systems, vol 332. Springer, Berlin/New York

Cass D, Shell K (1976) The Hamiltonian approach to dynamic economics. Academic Press, New York

Contreras J, Klusch M, Krawczyk JB (2004) Numerical solutions to Nash-Cournot equilibria in coupled constraint electricity markets. IEEE Trans Power Syst 19(1):195–206

Facchinei F, Fischer A, Piccialli V (2007) On generalized Nash games and variational inequalities. Oper Res Lett 35:159–164

Facchinei F, Fischer A, Piccialli V (2009) Generalized Nash equilibrium problems and newton methods. Math Program Ser B 117:163–194

Facchinei F, Kanzow C (2007) Generalized Nash equilibrium problems. 4OR: Q J Oper Res 53:173–210
Feinstein CD, Luenberger DG (1981) Analysis of the asymptotic behavior of optimal control trajectories: the implicit programming problem. SIAM J Control Optim 19:561–585
Fukushima M (2009) Restricted generalized Nash equilibria and controlled penalty algorithm. Comput Manag Sci 315:1–18
Halkin H (1974) Necessary conditions for optimal control problems with infinite horizon. Econometrica 42(2):267–273
Hämäläinen RP, Haurie A, Kaitala V (1985) Equilibria and threats in a fishery management game. Optim Control Appl Methods 6:315–333
Harker PT (1991) Generalized Nash games and quasivariational inequalities. Eur J Oper Res4: 81–94
Hartl RF, Sethi SP, Vickson RG (1995) A survey of the maximum principles for optimal control problems with state constraints. SIAM Rev 37:181–218
Haurie A (1995) Environmental coordination in dynamic oligopolistic markets. Gr Decis Negot 4:39–57
Haurie A, Krawczyk J, Zaccour G (2012) Games and dynamic games. Series in business, vol 1. World Scientific-Now Publishers, Singapore/Hackensack
Haurie A, Krawczyk BJ (1997) Optimal charges on river effluent from lumped and distributed sources. Environ Model Assess 2:177–199
Haurie A, Leitmann G (1984) On the global stability of equilibrium solutions for open-loop differential games. Larg Scale Syst 6:107–122
Haurie A, Marcotte P (1985) On the relationship between nash-cournot and wardrop equilibria. Networks 15:295–308
Haurie A, Roche M (1994) Turnpikes and computation of piecewsie open-loop equilibria in stochastic differential games. J Econ Dyn Control 18:317–344
Haurie A, Tolwinski B (1985) Definition and properties of cooperative equilibria in a two-player game of infinite duration. J Optim Theory Appl 46(4):525–534
Haurie A, Zaccour G (1995) Differential game models of global environmental management. In: Carraro C, Filar JA (eds) Control and game-theoretic models of the environment. Annals of the international society of dynamic games, vol 2. Birkhäuser, Boston pp 3–23
Hobbs BE, Pang JS (2007) Nash-Cournot equilibria in electric power markets with piecewise linear demand functions and joint constraints. Oper Res 55(1):113–127
Kesselman A, Leonardi S, Bonifaci V (2005) Game-theoretic analysis of internet switching with selfish users. In: Proceedings of the first international workshop on internet and network economics, WINE. Lectures notes in computer science, vol 3828, Springer, Berlin, Heildberg, pp 236–245
Krawczyk JB (2007) Numerical solutions to coupled-constraint (or generalised) Nash equilibrium problems. Comput Manag Sci 4(2):183–204
Mangasarian OL (1969) Nonlinear programming. McGraw-Hill, New York
McKenzie L (1976) Turnpike theory. Econometrica 44:841–865
Pang JS, Fukushima M (2005) Quasi-variational inequalities, generalized Nash equilibria and multi-leader-follower games. Comput Manag Sci 1:21–56
Rockafellar RT (1973) Saddle points of Hamiltonian systems in convex problems of Lagrange. Optim J Theory Appl 12:367–390. MR0358516
Rockafellar RT (1976) Saddle points of Hamiltonian systems in convex Lagrange problems having nonzero discount rate. J Econ Theory 12:71–113
Rosen JB (1965) Existence and uniqueness of equilibrium points for concave n-person games. Econometrica 33(3):520–534
Rubinstein A (1979) Equilibrium in supergames with the overtaking criterion. J Econ Theory 21:1–9
Tidball M, Zaccour G (2005) An environmental game with coupling constraints. Environ Model Assess 10(2):153–158

Tidball M, Zaccour G (2009) A differential environemental game with coupling constraints. Optim Control Appl Methods 30:197–207.65

von Heusingen C, Kanzow A (2006) Optimization reformulations of the generalized Nash equilibrium problem using Nikaido-Isoda-type functions, Preprint, vol 269. Institute of Mathematics, University of Würzburg, Wu´rzburg

von Weizsäcker CC (1965) Existence of optimal programs of accumulation for an infinite time horizon. Rev Econ Stud 32(2):85–104

Multistage Games

4

Jacek B. Krawczyk and Vladimir Petkov

Contents

J. B. Krawczyk (✉) · V. Petkov
Victoria University of Wellington, Wellington, New Zealand
e-mail: Jacek.Krawczyk@vuw.ac.nz; J.Krawczyk@vuw.ac.nz; Vladimir.Petkov@vuw.ac.nz

T. Başar, G. Zaccour (eds.), *Handbook of Dynamic Game Theory*,
https://doi.org/10.1007/978-3-319-44374-4_3

Abstract

In this chapter, we build on the concept of a *repeated game* and introduce the notion of a *multistage game*. In both types of games, several antagonistic agents interact with each other over time. The difference is that, in a multistage game, there is a dynamic system whose state keeps changing: the controls chosen by the agents in the current period affect the system's future. In contrast with repeated games, the agents' payoffs in multistage games depend directly on the state of this system. Examples of such settings range from a microeconomic dynamic model of a fish biomass exploited by several agents to a macroeconomic interaction between the government and the business sector. In some multistage games, physically different decision-makers engage in simultaneous-move competition. In others, agents execute their actions sequentially rather than simultaneously. We also study hierarchical games, where a leader moves ahead of a follower. The chapter concludes with an example of memory-based strategies that can support Pareto-efficient outcomes.

Keywords

Collusive Equilibrium · Discrete-Time Games · Feedback (Markovian) Equilibrium · Information Patterns · Open-Loop Nash Equilibrium · Sequential Games · Stackelberg Solutions

1 Introduction

In this chapter, we present a brief accessible overview of established research concerning multistage games. For further details, we refer the readers to several books on the topic, such as Başar and Olsder (1999) or Haurie et al. (2012).

Games that are *not* one-shot, or "static," but develop over time are called *dynamic*. There are several possible ways to classify such games. In this chapter, we will investigate a subclass of dynamic games in which the agents (also called players or actors) obtain rewards that depend on the state of a dynamic system jointly controlled by these agents. Our focus will be on games that are played in discrete time (i.e., over many stages).

We will define and study multistage games in a *state space*. A state space contains *state variables* which provide an exhaustive summary of how different input variables (known as *controls*) have impacted the system in the past. A system is called *dynamic* when its current state depends on its past states and inputs. If the current *state* and the future time profiles of the input variables are known, we are able to predict the subsequent behavior of this dynamic system.

We focus on games that are played in discrete time. The evolution of such dynamic systems is mathematically described by *difference* equations. These equations are called *state equations* because they determine how the state changes from one stage to the next. Differential games, on the other hand, are played in continuous time and are described by *differential* equations (see ▸ Chap. 2 "Nonzero-Sum Differential Games").

In the next section, we will briefly recall some fundamental properties of dynamic systems defined in state space. Then we will discuss various equilibrium solutions under different information structures. We will show how to characterize these solutions through either the coupled maximum principle or dynamic programming. We will describe a procedure for determining the equilibrium using an infinite-horizon fishery's management problem. This chapter will also cover multistage games where agents choose their actions sequentially rather than simultaneously. Examples of such games include alternating-move Cournot competition between two firms, intergenerational bequest games, and intrapersonal games of decision-makers with *hyperbolic discounting*. Furthermore, we will discuss solutions to a simple dynamic Stackelberg game where the system's dynamics are described by discrete transitions, rather than by difference equations. This chapter will conclude with an analysis of equilibria in infinite-horizon systems that incorporate the use of threats in the players' strategies, allowing the enforcement of cooperative solutions.

2 Description of Multistage Games in a State Space

Optimal control theory deals with problems of selecting the best control input (called optimal control) in a dynamic system to optimize a single performance criterion. If the controls are distributed among a set of independent actors who strive to optimize their individual performance criteria, we then have a *multistage game in a state space*.

In discrete time, $t = 0, 1, 2, \ldots, T$, the system's dynamics of a game played in a state space by m players can be represented by the following state equation:

$$\mathbf{x}(t+1) = \mathbf{f}(\mathbf{x}(t); \mathbf{u}_1(t), \ldots, \mathbf{u}_m(t), t) \qquad \mathbf{x}(0) = \mathbf{x}^0. \tag{4.1}$$

Here, $\mathbf{x}^0 \in \mathbb{R}^n$ is a given initial condition, $\mathbf{x}(t) = (x_1(t), \ldots, x_n(t)) \in \mathbb{R}^n$ is the state vector at t, $\mathbf{u}_j(t) \in \mathbb{R}^{p_j}$, $j = 1, 2 \ldots m$ is the control of Player j at t, and

$$\mathbf{f}(\cdot,\cdot,\cdot) = (f_i(\cdot,\cdot,\cdot))_{j=1,\ldots,n} : \mathbb{R}^n \times \mathbb{R}^{p_1+\cdots+p_m} \times \mathbb{R}_+ \mapsto \mathbb{R}^n$$

is the transition function from t to $t+1$.[1] Hence, equation (4.1) determines the values of the state vector at time $t+1$, for a given $\mathbf{x}(t) = (x_1(t), \ldots, x_n(t))$ and controls[2] $\mathbf{u}_1(t), \ldots, \mathbf{u}_m(t)$ of the players. We will use the following notation regarding the players' combined controls and state trajectory:

[1] In the sections where we deal with dynamic systems described by multiple state equations, we adopt a notation where vectors and matrices are in boldface style to distinguish them from scalars that are in regular style.

[2] In stochastic systems, some "controls" may come from nature and are thus independent of other players' actions.

$$\underline{\mathbf{u}}(t) = \{\mathbf{u}_1(t), \ldots, \mathbf{u}_m(t)\}, \tag{4.2}$$

$$\tilde{\mathbf{u}}_j^t = \{\mathbf{u}_j(0), \mathbf{u}_j(1), \mathbf{u}_j(2), \ldots, \mathbf{u}_j(t-1)\}, \tag{4.3}$$

$$\tilde{\underline{\mathbf{u}}}^t = \{\underline{\mathbf{u}}(0),\ \underline{\mathbf{u}}(1), \underline{\mathbf{u}}(2), \ldots, \underline{\mathbf{u}}(t-1)\}, \tag{4.4}$$

$$\tilde{\mathbf{x}}^t = \{\mathbf{x}(0), \mathbf{x}(1), \mathbf{x}(2), \ldots, \mathbf{x}(t)\}. \tag{4.5}$$

When there is no final period T, we will drop the time superscript t.

The vector $h^t = \{t, \tilde{\underline{\mathbf{u}}}^t, \tilde{\mathbf{x}}^t\}$ is called the *history of the game* at time t. In other words, the history is the sequence of values of the control and state variables that have driven the system up to period t. The information available to the players when choosing their controls in period t is either the entire history or part of it.

In summary, when defining a multistage state-space game, we must specify the following elements.

- The set of players $M = \{1, 2, \ldots m\}$.
- The state equation:

$$\mathbf{x}(t+1) = \mathbf{f}(\mathbf{x}(t), \underline{\mathbf{u}}(t), t), \quad \mathbf{x}(t_0) = \mathbf{x}^0,$$

  where $\mathbf{x}(t) \in \mathbb{R}^n$ (possibly $\mathbb{X} \subset \mathbb{R}^n$) is the state variable vector; $\underline{\mathbf{u}}(t) = (\mathbf{u}_1(t), \ldots, \mathbf{u}_m(t))$, $\mathbf{u}_j(t) \in \mathbf{U}_j \subset \mathbb{R}^{p_j}$, are players' controls, all of appropriate dimensions; and $\mathbf{f}(\cdot,\cdot,\cdot)$ is the vector of functions in (4.1) that define the evolution of the state variables.
- The information structure, i.e., the part of the history vectors $h^t, t = 0, 1, \ldots, T$, that is utilized by the players when computing their controls at time t.
- Any other relevant restrictions on the system's variables.
- The payoff functions (also called utility functions or performance criteria) optimized by the players. Depending on whether the time horizon of the game is finite or infinite, these performance criteria could take different forms:

  – In finite-horizon settings ($T < \infty$), the payoff of Player j is typically defined as

$$J_j \triangleq \sum_{t=t_0}^{T-1} g_j(\mathbf{x}(t), \underline{\mathbf{u}}(t), t) + S_j(\mathbf{x}(T)), \tag{4.6}$$

  where $g_j(\mathbf{x}(t), \underline{\mathbf{u}}(t), t) \in \mathbb{R}$ is the *transition reward* of Player j and $S_j(\mathbf{x}(T)) \in \mathbb{R}$ is his *terminal reward* (or the *bequest function*).

  – In infinite-horizon settings (when $T \to \infty$), the infinite sum of transition rewards may tend to infinity, hence it is not obvious at the outset how a performance criterion should be defined or how performances could be compared if the criterion is not convergent. For such settings, the literature proposes several approaches. Here we list some commonly used criteria.[3]

(a) *Discounted sum of rewards*

$$J_j \triangleq \sum_{t=0}^{\infty} \beta_j^t g_j(\mathbf{x}(t), \underline{\mathbf{u}}(t)), \tag{4.7}$$

where $0 \le \beta_j < 1$ is the discount factor of Player j. This approach to evaluating performance is known as the *discounted* criterion. If the transition reward is a uniformly bounded function, this infinite discounted sum converges to a finite value, so payoffs computed in this manner can be compared. Note that this criterion assigns a diminishing weight to rewards that occur in the future. Thus, performance is mostly influenced by what happens early on. The discounted criterion discriminates against generations in the far future.

(b) Another criterion which puts more weight on rewards obtained at a distant future is the *limit of average reward*.[4] It offers another way of dealing with a non-convergent series,

$$J_j \triangleq \liminf_{T \to \infty} \frac{1}{T} \sum_{t=0}^{T-1} g_j(\mathbf{x}(t), \underline{\mathbf{u}}(t)). \tag{4.8}$$

This criterion is based on the limit of the average reward per period. Note that, as T tends to infinity, the rewards gained over a finite number of periods will tend to have a negligible influence on performance; only what happens in the long term matters. Unlike the previous criterion, this one does not assign a higher weight on payoffs in early periods.

(c) There are other methods for comparing infinite streams of rewards even when their sums do not converge. For example, an alternative approach is to use the *overtaking optimality criteria*. We do not discuss these criteria here. Instead, we refer the readers to Haurie et al. (2012) and the bibliography provided there. When using one of the overtaking optimality criteria, we cannot talk about payoff maximization; however, such a criterion can still help us determine whether one stream of rewards is better than another.

We conclude this section by presenting a stylized model of a plausible conflict situation in fishery management. This situation has all the ingredients of a dynamic game played in a state space. This was recognized by several authors, e.g., Clark

[3] We should also note that, when the time horizon is infinite, it is usually assumed that the system is *stationary*. That is, the reward and state transition functions do not depend explicitly on time t.

[4] This limit is known as *Cesaro limit*.

(1976), Long (1977), and Levhari and Mirman (1980). Later we will use this model to illustrate a procedure for computing equilibria.

Example 1. **The Great Fish War.** The name for this fishery-management model was coined by Levhari and Mirman (1980). Suppose that two players, $j = 1, 2$, exploit a fishery. Let $x(t)$ be a measure of the fish biomass at time t, and let $u_j(t)$ denote Player j's catch in that period (also measured in normalized units). Player j strives to maximize a performance criterion with form

$$J_j \triangleq \sum_{t=t_0}^{T-1} \beta_j^t \sqrt{u_j(t)} + K_j \beta_j^T \sqrt{x(T)} \qquad j = 1, 2, \tag{4.9}$$

where $\beta_j \in [0, 1)$ is a discount factor, and $K_j > 0$ is a scaling parameter of the fishery's scrap value. The square roots of harvest and of the fishery's scrap value in (4.9) reflect diminishing marginal utility from the catch and the bequest. When the fishery is used in perpetuity, these payoff functions can be modified as follows:

$$J_j \triangleq \sum_{t=t_0}^{\infty} \beta_j^t \sqrt{u_j(t)} \qquad j = 1, 2 \tag{4.10}$$

where $0 < \beta_j < 1$ is the discount factor of Player j.

The interdependence between the two players is due to their joint exploitation of the common resource. This interdependence is represented by the state equation for the evolution of the quantity of fish when there is exploitation:

$$x(t+1) = f(x(t) - u_1(t) - u_2(t)), \tag{4.11}$$

where f is an increasing function.

Equation (4.11) provides a state-space description for the fishing process at hand. The state variable is $x(t)$, and the controls are $u_1(t)$, $u_2(t)$. Expressions (4.9) and (4.10) suggest possible utility functions that the players may want to maximize.

The dynamics of the model will crucially depend on the transition function f. One specification considered by Levhari and Mirman (1980) is $f(y) = y^\alpha$, where $0 < \alpha < 1$. In this case, without human intervention (i.e., $u_1(t) = u_2(t) = 0$), the long-run fish biomass will converge to a steady state of 1. An alternative possibility is $f(y) = ay$. Given this law of motion, setting $u_1(t) = u_2(t) = 0$ would imply that the fish biomass will either grow monotonically (if $a > 1$) or will approach extinction (if $0 < a < 1$). In a later section of this chapter, we will derive explicit solutions for the game with a linear law of motion.

3 Multistage Game Information Structures, Strategies, and Equilibria

To simplify the exposition, we will restrict the analysis to a two-player multistage game with a state equation

$$\mathbf{x}(t+1) = \mathbf{f}(\mathbf{x}(t), \mathbf{u}_1(t), \mathbf{u}_2(t), t), \quad \mathbf{x}(t_0) = \mathbf{x}^0, \tag{4.12}$$

where $\mathbf{x}(t) \in \mathbb{R}^n$, $\mathbf{u}_j \in U_j \subset \mathbb{R}^{m_j}$ and payoff functions $J_j(t_0, \mathbf{x}^0; \tilde{\mathbf{u}}_1, \tilde{\mathbf{u}}_2)$, $j = 1, 2$, which may have forms (4.6) or (4.7). Here U_j denotes Player j's constraint set.[5] A control is *admissible* if it satisfies the agents' constraints.

3.1 Information Structures

During game play, information is mapped into actions by the players' strategies. Depending on what information is available to the players in a multistage game, the control of the dynamic system can be designed in different ways (e.g., *open loop* or *closed loop*).

In general, the most complete information that an agent can have is the entire game history. Since the total amount of information tends to increase with t, the use of the entire history for decision-making may be impractical. In most cases of interest, not all information accumulated up to time t turns out to be relevant for decisions at that point. Thus, we may only need to consider information that is reduced to a vector of a fixed and finite dimension. The possibility of utilizing the game history to generate strategies is a topic routinely discussed in the context of repeated games. Here we will consider the following information structures:

1. *Closed-loop information structure.* In each period t, both players have access to the information about the history of the game h^t. This history can include only the state trajectory, i.e., for all t

$$h^t = \{\mathbf{x}(0), \mathbf{x}(1), \ldots, \mathbf{x}(t)\}, \tag{4.13}$$

   or include the history of players' actions, i.e., for all t

$$h^t = \{\mathbf{x}(0), \mathbf{u}_1(0), \mathbf{u}_2(0), \ldots, \mathbf{u}_1(t-1), \mathbf{u}_2(t-1)\}. \tag{4.14}$$

[5] In the most general formulation of the problem, the control constraint sets may depend on time and the current state, i.e., $U_j(t, x) \subset \mathbb{R}^{m_j}$. Moreover, sometimes the state may be constrained to remain in a subset $\mathbf{X} \subset \mathbb{R}^n$. We avoid these complications here.

In a deterministic context, the knowledge of $\mathbf{x}(t)$ for $t > 0$ is implicit, since the state trajectory can be reconstructed from the control inputs and the initial value of the state.

Let us call $\mathcal{H}^t$ the set of all possible game histories at time t. The strategy of Player j is defined by a sequence of mappings $\gamma_{jt}(\cdot) : \mathcal{H}^t \to U_j$, which associate a control value at time t with the observed history h^t,

$$\mathbf{u}_j(t) = \gamma_{jt}(h^t).$$

2. *Feedback information structure*, frequently referred to as *Markovian information structure*, is a special case of the closed-loop information structure. Both players know the *current* state of the system and "forget," or do not retain, information from the previous stages. A strategy for Player j is thus a function $\sigma_j(\cdot,\cdot) : \mathrm{N} \times \mathbf{R}^n \to U_j$ that maps the current state $\mathbf{x}(t)$ and, in finite-horizon games, calendar time t, into Player j's control variable. That is, *feedback* or *Markovian* strategies have forms

$$\mathbf{u}_j(t) = \sigma_j(t, \mathbf{x}(t)).$$

3. *Open-loop information structure.* Both players use only the knowledge of the initial state x^0 and the time t to determine their controls throughout the play. A *strategy* for Player j is thus defined by a mapping $\mu_j(\cdot,\cdot) : \mathbf{R}^n \times \mathrm{N} \to U_j$, where

$$\mathbf{u}_j(t) = \mu_j(\mathbf{x}^0, t) \quad \text{for all } t.$$

Given an initial state $\mathbf{x}^0$, an open-loop strategy assigns values to a player's control for all periods t. The players may use this information structure if they are able to commit to actions in all forthcoming periods. They may also resort to such strategies because they simply do not have access to any information other than the initial condition.

Open-loop and closed-loop control structures are common in real life. The trajectory of interplanetary spacecraft is often designed via an open-loop control law. In airplanes and modern cars, controls are implemented through servomechanisms, i.e., in feedback form.[6] Many economic dynamics models, e.g., those related to the theory of economic growth (such as the Ramsey model) are formulated as open-loop control systems. In industrial organization models of market competition,

[6]For example, the anti-block braking systems (ABS) used in cars, the automatic landing systems of aircrafts, etc.

"strategic" agents define their actions based on the available information; hence, they are assumed to use closed-loop controls.[7]

The following example illustrates the difference between open-loop and feedback controls in a simple nonstrategic setting.

Example 2. If you choose what you wear according to the calendar: "if-it-is-summer-I-wear-a-teeshirt; if-it-is-winter-I-wear-a-sweater," you are using an open-loop control. If, however, you check the actual temperature before you choose a piece of garment, you use a feedback control: "if-it-is-warm-I-wear-a-teeshirt; if-it-is-cold-I-wear-a-sweater." Feedback control is "better" for these decisions because it adapts to weather uncertainties.

As in the above example, dynamic systems may be subjected to random disturbances; we call such systems *stochastic*. In these systems, the control must be adapted to the changing conditions and that would require a closed loop, feedback, or some sort of adaptive control. Stochastic systems and adapting controls to the history of random disturbances will be studied in other chapters of this handbook.

The information structure has important repercussions for the *solution* to a dynamic game.

Remark 1.

- The open-loop structure defined in "3" is often considered implausible in the context of dynamic games, mainly because it does not lead to **subgame perfect** equilibrium solutions. This point will be discussed later in this section.
- The difference between information structures "2" and "1" can profoundly influence the outcome of play if agents have some ability to utilize the game history, or "memory," to formulate strategies based on threats (e.g., to collude). It also matters if the game context is stochastic. In the presence of stochasticity, the game history (4.13) is not equivalent to (4.14).
- Under the information structure "2," players "forget" the opponents' actions and are unable to implement an adequate punishment. If, on the other hand, the structure is as specified by "1," it is obvious "who did what," and players may impose penalties on their opponents for undesirable behavior.

We also make two observations regarding the computation of equilibria.

1. The procedure for computing equilibrium solutions under structures "1" and "2" can be substantially more difficult relative to that for structure "3."

[7]Commitments (agreements, treaties, schedules, planning processes, etc.) may force the agents to use the open-loop control even if state observations are available. On the other hand, some state variables (like the quantity of fish biomass in a management model for an ocean fishery) cannot be easily observable. In such cases, the agents may try to establish feedback controls using proxy variables, e.g., fish prices on a particular market.

2. Unless otherwise stated, we always assume that the rules of the games, i.e., the dynamics, the control sets, and the information structure, are common knowledge.

Example 3 below illustrates the notion of information structure and its implications for game play.

Example 3. In Example 1, if the fishermen do not know the fish biomass, they are bound to use open-loop strategies, based on some initial x^0 assessed or measured at some point in time. If the fishermen could measure the fish biomass reasonably frequently and precisely (e.g., by using some advanced process to assess the catchability coefficient or a breakthrough procedure involving satellite photos), then, most certainly, they would strive to compute feedback-equilibrium strategies. However, measuring the biomass is usually an expensive process, and, thus, an intermediate control structure may be practical: update the "initial" value x^0 from time to time and use an open-loop control in between.[8]

As discussed earlier, a multistage game can admit different solutions depending on the information available to players.

3.2 Open-Loop Nash Equilibrium Strategies

In a finite-horizon setting, a multistage two-player game is defined by the following utility functions (or performance criteria) and state equations[9]:

$$J_j \triangleq \sum_{t=0}^{T-1} g_j(\mathbf{x}(t), \mathbf{u}_1(t), \mathbf{u}_2(t), t) + S_j(\mathbf{x}(T)), \text{ for } j = 1, 2 \quad (4.15)$$

$$\mathbf{u}_j(t) \in U_j \quad (4.16)$$

$$\mathbf{x}(t+1) = \mathbf{f}(\mathbf{x}(t), \mathbf{u}_1(t), \mathbf{u}_2(t), t)\,, \ t = 0, 1, \dots T-1 \quad (4.17)$$

$$\mathbf{x}(0) = \mathbf{x}^0. \quad (4.18)$$

If players use open-loop strategies, each observes the initial state $\mathbf{x}^0$ and chooses an admissible control sequence $\tilde{\mathbf{u}}_j^T = (\mathbf{u}_j(0), \dots, \mathbf{u}_j(T-1))$, $j = 1, 2$. From the initial position $(0, \mathbf{x}^0)$, these choices generate a state trajectory $\tilde{\mathbf{x}}^T$ that solves (4.17)–(4.18),

$$\mathbf{x}(1) = \mathbf{f}(\mathbf{x}^0, \mathbf{u}_1(0), \mathbf{u}_2(0), 0)$$

[8] An information structure of this type is known as *piecewise open-loop control*.

[9] Also see equations (4.35)–(4.36) later in the chapter.

$$\mathbf{x}(2) = \mathbf{f}(\mathbf{x}(1), \mathbf{u}_1(1), \mathbf{u}_2(1), 1)$$
$$\dots$$
$$\mathbf{x}(T) = \mathbf{f}(\mathbf{x}(T-1), \mathbf{u}_1(T-1), \mathbf{u}_2(T-1), T-1)$$

as well as payoffs according to (4.15). The performance criteria (4.15), together with (4.16), (4.17), and (4.18), define the *normal form* of the open-loop multistage game at the initial point $(0, \mathbf{x}^0)$. We will use $J_j(0, \mathbf{x}^0; \tilde{\mathbf{u}}_1^T, \tilde{\mathbf{u}}_2^T)$, $j = 1, 2$, to denote the players' payoffs as functions of the open-loop strategies.[10]

Definition 1. A pair of admissible control sequences[11] $\underline{\tilde{\mathbf{u}}}^* = (\tilde{\mathbf{u}}_1^*, \tilde{\mathbf{u}}_2^*)$ is an **open-loop Nash equilibrium** at $(0, \mathbf{x}^0)$ if it satisfies the following equilibrium conditions

$$J_1(0, \mathbf{x}^0; \underline{\tilde{\mathbf{u}}}^*) \geq J_1(0, \mathbf{x}^0; \tilde{\mathbf{u}}_1, \tilde{\mathbf{u}}_2^*) \quad \forall \text{ admissible } \tilde{\mathbf{u}}_1,$$
$$J_2(0, \mathbf{x}^0; \underline{\tilde{\mathbf{u}}}^*) \geq J_2(0, \mathbf{x}^0; \tilde{\mathbf{u}}_1^*, \tilde{\mathbf{u}}_2) \quad \forall \text{ admissible } \tilde{\mathbf{u}}_2.$$

The standard approach to characterizing an open-loop Nash equilibrium solution is to apply the so-called coupled maximum principle. For this purpose, we define the Hamiltonian for each Player j as

$$H_j(\mathbf{p}_j(t+1), \mathbf{x}(t), \mathbf{u}_1(t), \mathbf{u}_2(t), t) \equiv$$
$$g_j(\mathbf{x}(t+1), \mathbf{u}_1(t), \mathbf{u}_2(t), t) + \mathbf{p}_j(t+1)'\mathbf{f}(\mathbf{x}(t), \mathbf{u}_1(t), \mathbf{u}_2(t), t), \qquad (4.19)$$

where $\mathbf{p}_j(t)$ is a *costate* vector[12] in $\mathbb{R}^n$ and $'$ indicates the transposition of the vector $\mathbf{p}_j(t+1)$ in a scalar product.

Assumption 1. *Assume that* $\mathbf{f}(\mathbf{x}, \underline{\mathbf{u}}, t)$ *and* $g_j(\mathbf{x}, \underline{\mathbf{u}}, t)$ *are continuously differentiable in the state* $\mathbf{x}$ *and continuous in the controls* $\underline{\mathbf{u}}$ *for each* $t = 0, \dots, T-1$, *and that* $S_j(\mathbf{x})$ *is continuously differentiable in* $\mathbf{x}$. *Moreover, assume that, for each* j, U_j *is compact and convex. Finally, suppose that, for each* t, $\mathbf{x}$, *the function* $H_j(\mathbf{p}, \mathbf{x}, \mathbf{u}_j, \mathbf{u}_{-j}, t)$ *is concave in* $\mathbf{u}_j$.

We can now state the following lemma that provides the necessary conditions for the open-loop equilibrium strategies:

[10] We note that expressing payoffs as functions of players' strategies is necessary for a game definition in normal form.

[11] To simplify notations, from now on we will omit the superscript ^{T} and refer to $\tilde{\mathbf{u}}_j$ instead of $\tilde{\mathbf{u}}_j^T$ or $\tilde{\mathbf{x}}$ instead of $\tilde{\mathbf{x}}^T$.

[12] Also called adjoint vector. This terminology is borrowed from optimal control theory.

Lemma 1. *Suppose that Assumption 1 is satisfied, and let* $\tilde{\underline{\mathbf{u}}}^*$ *be an open-loop Nash equilibrium pair of controls, generating the trajectory* $\tilde{\mathbf{x}}^*$ *from the initial state* $\mathbf{x}^0$ *for the game (4.15), (4.17). Then there exist functions of time* $\mathbf{p}_j(\cdot)$*, with values in* $\mathbb{R}^n$*, such that the following relations hold*

$$\mathbf{u}_j^*(t) = \arg \max_{u_j(t) \in U_j} H_j(\mathbf{p}_j(t+1), \mathbf{x}^*(t), \mathbf{u}_j(t), \mathbf{u}_{-j}^*(t), t), \tag{4.20}$$

$$\mathbf{p}_j(t)' = \frac{\partial}{\partial \mathbf{x}} H_j(\mathbf{p}_j(t+1), \mathbf{x}^*(t), \mathbf{u}_1^*(t), \mathbf{u}_2^*(t), t), \tag{4.21}$$

$$\mathbf{p}_j(T)' = \frac{\partial}{\partial \mathbf{x}(T)} S_j(\mathbf{x}^*(T)), \;\; j = 1, 2. \tag{4.22}$$

The terminal conditions for the functions $\mathbf{p}_j(T)$ are called **transversality conditions**.

For a simple proof see Haurie et al. (2012). Başar and Olsder (1982) or Fan and Wang (1964) present a more complete proof.

To compute open-loop Nash equilibria, we can use a *mathematical-programming* approach, provided that the following assumption holds.

Assumption 2. *Assume that* $\mathbf{f}(\mathbf{x}, \underline{\mathbf{u}}, t)$ *and* $g_j(\mathbf{x}, \underline{\mathbf{u}}, t)$ *are continuously differentiable in the state* $\mathbf{x}$ *and the controls* $\underline{\mathbf{u}}$ *for each* $t = 0, \ldots, T-1$*, and that* $S_j(\mathbf{x})$ *is continuously differentiable in* $\mathbf{x}$*. Moreover, assume that, for each* j*,* U_j *is defined by inequalities* $h_j(\mathbf{u}_j) \leq 0$*, where* $h_j : \mathbb{R}^{m_j} \to \mathbb{R}^{p_j}$*,* $p_j < m_j$*, are given continuously differentiable mappings.*

Now we consider the Lagrangians

$$\begin{aligned} L_j(\mathbf{p}_j(t+1), \nu_j(t), \mathbf{x}(t), \mathbf{u}_1(t), \mathbf{u}_2(t), t) \equiv\; & g_j(\mathbf{x}(t+1), \mathbf{u}_1(t), \mathbf{u}_2(t), t) \\ & + \mathbf{p}_j(t+1)'\mathbf{f}(\mathbf{x}(t), \mathbf{u}_1(t), \mathbf{u}_2(t), t) + \nu_j(t)' h_j(u_j(t)). \end{aligned} \tag{4.23}$$

Lemma 2. *Suppose that Assumption 2 is satisfied. Let* $\tilde{\underline{\mathbf{u}}}^*$ *be an open-loop Nash equilibrium pair of controls, generating the trajectory* $\tilde{\mathbf{x}}^*$ *from the initial state* $\mathbf{x}^0$ *for the game (4.15), (4.17), and let the constraint qualification conditions of Karush-Kuhn-Tucker hold. Then there exist functions of time* $\mathbf{p}_j(t)$*, with values in* $\mathbb{R}^n$ *and functions of time* $\nu_j(t)$*, with values in* $\mathbb{R}^{p_j}$*, such that, when forming the Lagrangians (4.23), the following holds true:*

$$0 = \frac{\partial}{\partial u_j} L_j(\mathbf{p}_j(t+1), \nu_j(t), \mathbf{x}^*(t), \mathbf{u}_j^*(t), \mathbf{u}_{-j}^*(t), t) \tag{4.24}$$

$$\mathbf{p}_j(t)' = \frac{\partial}{\partial \mathbf{x}} L_j(\mathbf{p}_j(t+1), \nu_j(t), \mathbf{x}^*(t), \mathbf{u}_1^*(t), \mathbf{u}_2^*(t), t), \tag{4.25}$$

$$0 = \nu_j(t)' h_j(\mathbf{u}_j^*(t))\rangle, t = 0, \ldots, T-1 \tag{4.26}$$

$$0 \le \nu_j(t), t = 0, \ldots, T-1 \tag{4.27}$$

$$\mathbf{p}_j(T)' = \frac{\partial}{\partial \mathbf{x}(T)} S_j(\mathbf{x}^*(T)), \ \ j = 1, 2. \tag{4.28}$$

For a simple proof see Haurie et al. (2012). Luenberger (1969) presents a more complete proof.

An open-loop equilibrium of an *infinite*-horizon multistage game can, in principle, be characterized by the same apparatus as a game played over a finite horizon. The only difference is that the transversality condition (4.22) needs to be modified. When $T \to \infty$, there may be many functions that satisfy this condition. A general rule is that the function $\mathbf{p}_j(\infty)$ cannot "explode." We refer the readers to Michel (1982) and the publications cited there (Halkin 1974 in particular) as they contain a thorough discussion on this issue.

3.3 Feedback-Nash (Markovian) Equilibrium Strategies

In this section we characterize the Nash equilibrium solutions for the class of feedback (or Markovian) strategies. Our approach is based on the dynamic programming method introduced by Bellman for control systems in Bellman (1957).[13]

Consider a game defined in normal form[14] at the initial data $(\tau, \mathbf{x}_\tau)$ by the payoff functions

$$J_j(\tau, \mathbf{x}^\tau; \sigma_1, \sigma_2) \triangleq \sum_{t=\tau}^{T-1} g_j(\mathbf{x}(t), \sigma_1(t, \mathbf{x}(t)), \sigma_2(t, \mathbf{x}(t)), t) + S_j(\mathbf{x}(T)), \ j = 1, 2, \tag{4.29}$$

where the state evolves according to

$$\mathbf{x}(t+1) = \mathbf{f}(\mathbf{x}(t), \sigma_1(t, \mathbf{x}(t)), \sigma_2(t, \mathbf{x}(t)), t), \ \ t = \tau, \ldots T-1, \tag{4.30}$$

$$\mathbf{x}(\tau) = \mathbf{x}^\tau, \quad \tau \in \{0, \ldots, T-1\}. \tag{4.31}$$

Assume that this game is played in feedback (or Markovian) strategies. That is, players use admissible feedback rules $\sigma_j(t, \mathbf{x})$, $j = 1, 2$. These rules generate a state trajectory from any initial point $(\tau, \mathbf{x}^\tau)$ according to (4.30)–(4.31) and payoffs according to (4.29). Let Σ_j denote the set of all possible feedback strategies of Player j.

[13] If a feedback strategy pair $\underline{\boldsymbol{\sigma}}(t, \mathbf{x})$ is continuous in t and its partial derivatives $\frac{\partial}{\partial \mathbf{x}} \underline{\boldsymbol{\sigma}}(t, \mathbf{x})$ exist and are continuous, then it is possible to characterize a feedback-Nash equilibrium through a coupled maximum principle (see Haurie et al. 2012).

[14] For notational simplicity, we still use J_j to designate this game's normal form payoffs.

Assuming that players strive to maximize (4.29), expressions (4.29), (4.30), and (4.31) define the normal form of the feedback multistage game.

When the time horizon is infinite ($T = \infty$), the normal form of these games is as follows:

$$J_j(\tau, \mathbf{x}^\tau; \sigma_1, \sigma_2) \triangleq \sum_{t=\tau}^{\infty} \beta_j^t g_j(\mathbf{x}(t), \sigma_1(\mathbf{x}(t)), \sigma_2(\mathbf{x}(t))), \quad j = 1, 2. \quad (4.32)$$

$$\mathbf{x}(t+1) = \mathbf{f}(\mathbf{x}(t), \sigma_1(\mathbf{x}(t)), \sigma_2(\mathbf{x}(t))), \quad t = 0, 1, \ldots, \infty, \quad (4.33)$$

$$\mathbf{x}(\tau) = \mathbf{x}^\tau, \quad \tau \in \{0, \ldots, \infty\}. \quad (4.34)$$

As usual, j is the index of a generic player, and β_j^t is Player j's discount factor, $0 < \beta_j < 1$, raised to the power t.

Definition 2. A pair of admissible feedback strategies $\underline{\sigma}^* = (\sigma_1^*, \sigma_2^*)$ is a **feedback-Nash equilibrium** if it satisfies the following equilibrium conditions:

$$J_1(\tau, \mathbf{x}^\tau; \underline{\sigma}^*) \geq J_1(\tau, \mathbf{x}^\tau; \sigma_1, \sigma_2^*) \quad \forall \sigma_1 \in \Sigma_1$$
$$J_2(\tau, \mathbf{x}^\tau; \underline{\sigma}^*) \geq J_2(\tau, \mathbf{x}^\tau; \sigma_1^*, \sigma_2) \quad \forall \sigma_2 \in \Sigma_2,$$

at any admissible initial point $(\tau, \mathbf{x}^\tau)$.

It is important to notice that, unlike the open-loop Nash equilibrium, the above definition must hold at *any* admissible initial point and not solely at the initial data $(0, \mathbf{x}^0)$. This is also why the *subgame perfection* property, which we shall discuss later in this chapter, is built into this solution concept.

Even though the definitions of equilibrium are relatively easy to formulate for multistage feedback or closed-loop games, we cannot be certain of the *existence* of these solutions. It is often difficult to find conditions that guarantee the existence of feedback-Nash equilibria. We do not encounter this difficulty in the case of open-loop Nash equilibria, which are amenable to existence proofs that are similar to those used for static concave games. Unfortunately, these methods are not applicable to feedback strategies.

Because of these existence issues, *verification theorems* are used to confirm a proposed equilibrium. A verification theorem shows that *if* we can find a solution to the dynamic programming equations, *then* this solution constitutes a feedback-Nash (Markovian) equilibrium. The existence of a feedback-Nash equilibrium can be established only in specific cases for which an explicit solution of the *dynamic programming* equations is obtainable. This also means that the dynamic programming technique is crucial for the characterization of feedback-Nash (Markovian) equilibrium. We discuss the application of this technique in Sect. 4 (for finite-horizon games) and in Sect. 5 (for infinite-horizon games).[15]

3.4 Subgame Perfection and Other Equilibrium Aspects

Strategies are called *time consistent* if they remain optimal throughout the entire equilibrium trajectory. The following lemma argues that open-loop Nash equilibrium strategies are time consistent.

Lemma 3. *Let* $(\tilde{\mathbf{u}}_1^*, \tilde{\mathbf{u}}_2^*)$ *be an open-loop Nash equilibrium at* $(0, \mathbf{x}^0)$ *and let* $\tilde{\mathbf{x}}^* = (\mathbf{x}^*(0), \mathbf{x}^*(1), \ldots, \mathbf{x}^*(T))$ *be the equilibrium trajectory generated by* $(\tilde{\mathbf{u}}_1^*, \tilde{\mathbf{u}}_2^*)$ *from* $(0, \mathbf{x}^0)$*. Then the truncation* $(\tilde{\mathbf{u}}_1^*, \tilde{\mathbf{u}}_2^*)_{[\tau, T-1]}$ *of these strategies to the periods* $\tau, \ldots, T-1$ *is an open-loop Nash equilibrium at the point* $(\tau, \mathbf{x}^*(\tau))$*, where* $\tau = 0, 1, \ldots, T-1$*, and* $\mathbf{x}^*(\tau)$ *is an intermediate state along the equilibrium trajectory.*

For a proof see Haurie et al. (2012).

However, an open-loop Nash equilibrium is not *subgame perfect* in the sense of Selten (see Selten 1975). If a player temporarily deviates from the equilibrium control and then resumes following an open-loop strategy, the remainder of the original control sequence will no longer be optimal. Subgame perfection requires that, from any state, the strategy always generates a control sequence that is optimal.

The lack of subgame perfection of open-loop equilibria has been viewed as a grave drawback of this solution. Games often involve randomness, like the "trembling hand" of an agent. This would trigger a deviation from the equilibrium trajectory, and therefore a collapse of the equilibrium under the open-loop information structure. If, however, we are in a non-stochastic environment, there is no reason to consider such deviations, since they would be detrimental to the deviating player.

Despite the above positive conclusion, we need to remind the reader that an open-loop behavior relies on the assumption that, at the beginning of time, the players can commit to a complete list of future actions without any possibility of update or revision during the course of the game. So, from that point of view, a feedback (or Markovian) strategy appears more suited to model the behavior of *strategic* agents, i.e., those who react to available information.

Notwithstanding the above reservation, some *real-life* situations can be modeled as open-loop equilibria. The oligopolistic R&D problem analyzed in Spencer and Brander (1983b) (also, see Spencer and Brander 1983a) is such an example. The so-called *patent races* are also naturally viewed as open-loop games. In those models, the use of open-loop strategies is justified because it approximates players'

[15] In a stochastic context, perhaps counterintuitively, certain multistage (*supermodular*) games defined on lattices admit feedback equilibria which can be established via a fixed-point theorem due to Tarski. See Haurie et al. (2012) and the references provided there.

real-world behavior: here the open-loop information structure adequately reflects the fact that a firm *cannot* observe its rivals' new technology before choosing its own output level.

On the other hand, a feedback-Nash equilibrium strategy is subgame perfect by construction (see Theorem 1). It is easy to see that subgame perfection is a stronger concept than time consistency,[16] implying that a feedback Nash equilibrium must be time consistent.

4 Dynamic Programming for Finite-Horizon Multistage Games

It is not difficult to extend the dynamic programming algorithm that is used to establish subgame perfection of equilibria in *repeated* games to *multistage* games.

Consider a finite-horizon dynamic game played by two agents. We will denote them by j and "not j," i.e., $-j$.[17] When the initial point is $(\tau, \mathbf{x}^\tau)$ and the control sequences are $\tilde{\mathbf{u}}_j$ and $\tilde{\mathbf{u}}_{-j}$, the payoff of Player j is defined by

$$J_j \triangleq \sum_{t=\tau}^{T-1} g_j(\mathbf{x}(t), \mathbf{u}_j(t), \mathbf{u}_{-j}(t), t) + S_j(\mathbf{x}(T)), \quad j = 1, 2, \tag{4.35}$$

where the state trajectory follows

$$\mathbf{x}(t+1) = \mathbf{f}(\mathbf{x}(t), \mathbf{u}_j(t), \mathbf{u}_{-j}(t), t), \ t = \tau, \tau+1, \ldots T-1 \tag{4.36}$$

$$\mathbf{x}(\tau) = \mathbf{x}^\tau. \tag{4.37}$$

Because we deal with feedback information structure, each player knows the current state. Players j and $-j$ will therefore choose their controls in period t using feedback strategies $\mathbf{u}_j(t) = \sigma_j(t, \mathbf{x})$ and $\mathbf{u}_{-j}(t) = \sigma_{-j}(t, \mathbf{x})$.

Let $(\sigma_j^*(t, \mathbf{x}), \sigma_{-j}^*(t, \mathbf{x}))$ be a feedback-equilibrium solution to the multistage game, and let $\tilde{\mathbf{x}}^* = (\mathbf{x}^*(\tau), \ldots, \mathbf{x}^*(T))$ be the associated trajectory resulting from $(\tau, \mathbf{x}_\tau)$. The *value function* for Player j is defined as

$$W_j^*(\tau, \mathbf{x}_\tau) = \sum_{t=\tau}^{T-1} g_j(\mathbf{x}^*(t), \sigma_j^*(t, \mathbf{x}^*(t)), \sigma_{-j}^*(t, \mathbf{x}^*(t)), t) + S_j(\mathbf{x}^*(T)), \tag{4.38}$$

$$W_j^*(T, \mathbf{x}) = S_j(\mathbf{x}) \tag{4.39}$$

$$\mathbf{x}^*(t+1) = \mathbf{f}(\mathbf{x}^*(t), \sigma_j^*(t, \mathbf{x}^*(t)), \sigma_{-j}^*(t, \mathbf{x}^*(t)), t)\,, \ t = \tau, \ldots T-1 \tag{4.40}$$

$$\mathbf{x}^*(\tau) = \mathbf{x}^\tau. \tag{4.41}$$

[16] For this reason, it is also sometimes referred to as strong time consistency; see Başar and Olsder (1999) and Başar (1989).

[17] This notation helps generalize our results. They would formally be unchanged if there were $m > 2$ players. In that case, $-j$ would refer to the $m-1$ opponents of Player j.

This value function represents the payoff that Player j will receive if the feedback-equilibrium strategy is played from an initial point (τ, x^τ) until the end of the horizon. The following result provides a decomposition of the equilibrium conditions over time.

Lemma 4. *The value functions $W_j^*(t, \mathbf{x})$ and $W_{-j}^*(t, \mathbf{x})$ satisfy the following recurrent equations, backward in time, also known as the* ***Bellman equations****:*

$$W_j^*(t, \mathbf{x}^*(t)) = \max_{\mathbf{u}_j} g_j(\mathbf{x}^*(t), \mathbf{u}_j, \sigma_{-j}^*(t, \mathbf{x}^*(t)), t) + W_j^*(t+1, \mathbf{f}(\mathbf{x}^*(t), \mathbf{u}_j, \sigma_{-j}^*(t, \mathbf{x}^*(t)), t))), \tag{4.42}$$

$$W_{-j}^*(t, \mathbf{x}^*(t)) = \max_{\mathbf{u}_{-j}} g_{-j}(\mathbf{x}^*(t), \sigma_j^*(t, \mathbf{x}^*(t)), \mathbf{u}_{-j}, t) + W_{-j}^*(t+1, \mathbf{f}(\mathbf{x}^*(t), \sigma_j^*(t, \mathbf{x}^*(t)), \mathbf{u}_{-j}, t))), \tag{4.43}$$

$$t = T-1, T-2, \ldots 0, \tag{4.44}$$

with the boundary condition (4.39).[18]

For a proof see Haurie et al. (2012).

The above lemma can be used to obtain necessary and sufficient conditions for a feedback-Nash equilibrium. It also underscores the fact that the equilibrium condition must hold in a set of local games defined at each possible initial point $(\tau, \boldsymbol{x}^\tau)$.

Consider the local game defined at $(\tau, \mathbf{x}^\tau)$, in which the players' actions are $(\mathbf{u}_j, \mathbf{u}_{-j})$ and their payoffs are given by

$$\mathfrak{h}_j(\tau, \mathbf{x}^\tau; \mathbf{u}_j, \mathbf{u}_{-j}) \equiv g_j(\mathbf{x}^\tau, \mathbf{u}_j, \mathbf{u}_{-j}, \tau,) + W_j^*(t+1, \mathbf{f}(\mathbf{x}^\tau, \mathbf{u}_j, \mathbf{u}_{-j}, \tau)). \tag{4.45}$$

Then the value of the feedback-equilibrium pair $(\sigma_j^*(\tau, \mathbf{x}^\tau), \sigma_{-j}^*(\tau, \mathbf{x}^\tau))$ at $(\tau, \mathbf{x}^\tau)$ is a Nash equilibrium for this local game.

The lemma suggests the following recursive approach for computing a feedback-Nash equilibrium.

- **At time** $T-1$, for any initial point $(T-1, \mathbf{x}(T-1))$, solve the local game with payoffs

$$\mathfrak{h}_j(T-1, \mathbf{x}(T-1); \mathbf{u}_j, \mathbf{u}_{-j}) \equiv g_j(\mathbf{x}(T-1), \mathbf{u}_j, \mathbf{u}_{-j}, T-1) + S_j(\mathbf{f}(\mathbf{x}(T-1), \mathbf{u}_j, \mathbf{u}_{-j}, T-1)), \tag{4.46}$$

[18]We note that the functions $W_j^*(\cdots)$ and $W_{-j}^*(\cdots)$ are *continuation payoffs*. Compare Sect. 6.1.

$$\mathfrak{h}_{-j}(T-1, \mathbf{x}(T-1); \mathbf{u}_j, \mathbf{u}_{-j})$$
$$\equiv g_{-j}(\mathbf{x}(T-1), \mathbf{u}_j, \mathbf{u}_{-j}, T-1) + S_{-j}(\mathbf{f}(\mathbf{x}(T-1), \mathbf{u}_j, \mathbf{u}_{-j}, T-1)). \tag{4.47}$$

Assume that a Nash equilibrium exists for each of these games, and let

$$(\sigma_j^*(T-1, \mathbf{x}(T-1)), \sigma_{-j}^*(T-1, \mathbf{x}(T-1)))$$

be the equilibrium strategy vector. That is,

$$\sigma_j^*(T-1, \mathbf{x}(T-1)) = \arg\max_{\mathbf{u}_j} \Big\{ g_j(\mathbf{x}(T-1), \mathbf{u}_j, \sigma_{-j}^*(T-1, \mathbf{x}(T-1)), T-1) + S_j(\mathbf{f}(\mathbf{x}(T-1), \mathbf{u}_j, \sigma_{-j}^*(T-1, \mathbf{x}(T-1)), T-1)) \Big\}$$

and similarly for Player $-j$. Then define

$$W_j^*(T-1, \mathbf{x}(T-1)) \equiv g_j(\mathbf{x}(T-1), \sigma_j^*(T-1, \mathbf{x}(T-1)), \sigma_{-j}^*(T-1, \mathbf{x}(T-1)), T-1) + S_j(\mathbf{f}(\mathbf{x}(T-1), \sigma_j^*(T-1, \mathbf{x}(T-1)), \sigma_{-j}^*(T-1, \mathbf{x}(T-1)), T-1)).$$

- **At time** $T-2$, for any initial point $(T-2, x(T-2))$, solve the local game with payoffs

$$\mathfrak{h}_j(T-2, \mathbf{x}(T-2); \mathbf{u}_j, \mathbf{u}_{-j})$$
$$\equiv g_j(\mathbf{x}(T-2), \mathbf{u}_j, \mathbf{u}_{-j}, T-2) + W_j^*(T-1, \mathbf{f}(\mathbf{x}(T-2), \mathbf{u}_j, \mathbf{u}_{-j}, T-2)),$$
$$\mathfrak{h}_{-j}(T-2, \mathbf{x}(T-2); \mathbf{u}_j, \mathbf{u}_{-j})$$
$$\equiv g_{-j}(\mathbf{x}(T-2), \mathbf{u}_j, \mathbf{u}_{-j}, T-2) + W_{-j}^*(T-1, \mathbf{f}(\mathbf{x}(T-2), \mathbf{u}_j, \mathbf{u}_{-j}, T-2)).$$

To ensure the existence of an equilibrium, we also need to assume that the functions $W_j^*(T-1, \cdot)$ and $W_{-j}^*(T-1, \cdot)$ identified in the first step of the procedure are sufficiently smooth.[19] Suppose that an equilibrium exists everywhere, and define

$$(\sigma_j^*(T-2, \mathbf{x}(T-2)), \sigma_{-j}^*(T-2, \mathbf{x}(T-2)))$$

[19] Recall that an equilibrium is a fixed point of a best-reply function, and that a fixed point requires some regularity to exist.

to be the equilibrium strategies in the above game. In other words,

$$\sigma_j^*(T-2,\mathbf{x}(T-2)) = \arg\max_{\mathbf{u}_j} \Big\{ g_j(\mathbf{x}(T-2),\mathbf{u}_j,\sigma_{-j}^*(T-2,\mathbf{x}(T-2)),T-2) + W_j^*(T-1,\mathbf{f}(\mathbf{x}(T-2),\mathbf{u}_j,\sigma_{-j}^*(T-2,\mathbf{x}(T-2)),T-2)) \Big\}$$

and similarly for Player $-j$. Then define

$$W_j^*(T-2,\mathbf{x}(T-2)) \equiv g_j(\mathbf{x}(T-2),\ \sigma_j^*(T-2,\mathbf{x}(T-2)),\sigma_{-j}^*(T-2,\mathbf{x}(T-2)), T-2) + W_j^*(T-1,\mathbf{f}(\mathbf{x}(T-2),\sigma_j^*(T-2,\mathbf{x}(T-2)), \sigma_{-j}^*(T-2,\mathbf{x}(T-2)),T-2)).$$

- **At time** $T-3$, etc. proceed recurrently, defining local games and their solutions for all preceding periods, until period 0 is reached.

  In the following *verification theorem*, we show that this procedure generates a feedback-Nash equilibrium strategy vector.

Theorem 1. *Suppose that there are value functions* $W_j^*(t,\mathbf{x})$ *and feedback strategies* $(\sigma_j^*,\sigma_{-j}^*)$ *which satisfy the local-game equilibrium conditions defined in equations* (4.39), (4.43) *for* $t = 0,1,2,\dots,T-1$, $\mathbf{x}\in\mathbb{R}^n$*. Then the feedback pair* $(\sigma_j^*,\sigma_{-j}^*)$ *constitutes a (subgame) perfect equilibrium of the dynamic game with the feedback information structure.*[20] *Moreover, the value function* $W_j^*(\tau,\mathbf{x}^\tau)$ *represents the equilibrium payoff of Player* j *in the game starting at point* $(\tau,\mathbf{x}^\tau)$.

For a proof see Haurie et al. (2012).

Remark 2. The above result, together with Lemma 4, shows that dynamic programming and the Bellman equations are both **necessary** and **sufficient** conditions for a feedback-Nash equilibrium to exist. Thus, we need to solve the system of Bellman equations (4.44) for Players j and $-j$ to verify that a pair of feedback-equilibrium strategies exists.[21] This justifies the term **verification theorem**.[22]

[20]The satisfaction of these conditions guarantees that such an equilibrium is feedback-Nash, or Markovian, equilibrium.

[21]We observe that feedback-Nash equilibrium is also a Nash equilibrium for the normal (strategic) form of the game. In that case it is not defined recursively, but using state-feedback control laws as strategies, see Başar and Olsder (1999) and Başar (1989).

[22] If it were possible to show that, at every stage, the local games are *diagonally strictly concave* (see e.g., Krawczyk and Tidball 2006), then one can guarantee that a unique equilibrium exists $\underline{\sigma}(t,\mathbf{x}) \equiv (\sigma_j(t,\mathbf{x}(t)),\sigma_{-j}(t,\mathbf{x}(t)))$. However, it turns out that diagonal strict concavity for a game at t does not generally imply that the game at $t-1$ possesses this feature.

5 Infinite-Horizon Feedback-Nash (Markovian) Equilibrium in Games with Discounted Payoffs

5.1 The Verification Theorem

We can extend the method for computing equilibria of the finite-horizon multistage games from Sect. 4 to infinite-horizon stationary discounted games. Let $(\sigma_j^*, \sigma_{-j}^*)$ be a pair of stationary feedback-Nash equilibrium strategies. The value function for Player j is defined by

$$\mathbf{x}^*(\tau) = \mathbf{x}^\tau \tag{4.48}$$

$$\mathbf{x}^*(t+1) = \mathbf{f}(\mathbf{x}^*(t), \sigma_j^*(\mathbf{x}^*(t)), \sigma_{-j}^*(\mathbf{x}^*(t))), \quad t = 0, 1, \ldots, \infty \tag{4.49}$$

$$W_j^*(\tau, \mathbf{x}^\tau) = \sum_{t=\tau}^{\infty} \beta_j^t g_j(\mathbf{x}^*(t), \sigma_j^*(\mathbf{x}^*(t)), \sigma_{-j}^*(\mathbf{x}^*(t))), \quad j = 1, 2\,. \tag{4.50}$$

Note that discounting of payoffs implies

$$W_j^*(\tau, \mathbf{x}) = \beta_j^\tau W_j^*(0, \mathbf{x}). \tag{4.51}$$

We will call $V_j^*(\mathbf{x}) \equiv W_j^*(0, \mathbf{x})$ the current-valued value function. A value function for Player $-j$ is defined in a similar way.

An analog to Lemma 4 for this infinite-horizon discounted game is provided below.

Lemma 5. *The current-valued value functions $V_j^*(\mathbf{x})$ and $V_{-j}^*(\mathbf{x})$, defined in (4.51), satisfy the following recurrence equations, backward in time, also known as **Bellman equations**:*

$$\begin{aligned} V_j^*(\mathbf{x(t)}^*(t)) = \max_{\mathbf{u}_j(t)} & \, g_j(\mathbf{x}^*(t), \mathbf{u}_j(t), \sigma_{-j}^*(\mathbf{x}^*(t))) + \\ & \beta_j \, V_j^*(\mathbf{f}(\mathbf{x(t)}^*, \mathbf{u}_j(t), \sigma_{-j}^*(\mathbf{x}^*(t)))), \end{aligned} \tag{4.52}$$

$$\begin{aligned} V_{-j}^*(\mathbf{x}^*(t)) = \max_{\mathbf{u}_{-j}(t)} & \, g_{-j}(\mathbf{x}^*(t), \sigma_j^*(\mathbf{x}^*(t)), \mathbf{u}_{-j}(t)) + \\ & \beta_{-j} \, V_{-j}^*(\mathbf{f}(\mathbf{x}^*(t), \sigma_j^*(\tilde{\mathbf{x}}^*(t)), \mathbf{u}_{-j}(t))), \\ & t = 0, 1, \ldots \infty\,. \end{aligned} \tag{4.53}$$

The proof is identical to that proposed for Lemma 4 (see Haurie et al. 2012), except for obvious changes in the definition of current-valued value functions.

Remark 3. There is, however, an important difference between Lemmas 4 and 5. The boundary conditions (4.39) which determine the value function in the final period are absent in infinite-horizon discounted games.

This remark suggests that we can focus on the local current-valued games, at any initial point $\mathbf{x}$, with payoffs defined by

$$\mathfrak{h}_j(\mathbf{x};\mathbf{u}_j,\mathbf{u}_{-j}) \equiv g_j(\mathbf{x},\mathbf{u}_j,\mathbf{u}_{-j}) + \beta_j V_j^*(\mathbf{f}(\mathbf{x},\mathbf{u}_j,\mathbf{u}_{-j})) \tag{4.54}$$

$$\mathfrak{h}_{-j}(\mathbf{x};\mathbf{u}_j,\mathbf{u}_{-j}) \equiv g_{-j}(\mathbf{x},\mathbf{u}_j,\mathbf{u}_{-j}) + \beta_{-j} V_{-j}^*(\mathbf{f}(\mathbf{x},\mathbf{u}_j,\mathbf{u}_{-j})). \tag{4.55}$$

Theorem 2. *Consider value functions $V_j^*(\mathbf{x})$ and $V_{-j}^*(\mathbf{x})$ and a stationary feedback strategy vector $(\sigma_j^*,\sigma_{-j}^*)$, such that the following holds true:*

$$V_j^*(\mathbf{x}) = \max_{\mathbf{u}_j} \mathfrak{h}_j(\mathbf{x};\mathbf{u}_j,\sigma_{-j}^*(\mathbf{x})) = \mathfrak{h}_j(\mathbf{x};\sigma_j^*(\mathbf{x}),\sigma_{-j}^*(\mathbf{x})) \tag{4.56}$$

$$V_{-j}^*(\mathbf{x}) = \max_{\mathbf{u}_{-j}} \mathfrak{h}_{-j}(\mathbf{x};\sigma_j^*(\mathbf{x}),\mathbf{u}_{-j}) = \mathfrak{h}_{-j}(\mathbf{x};\sigma_j^*(\mathbf{x}),\sigma_{-j}^*(\mathbf{x})), \tag{4.57}$$

where $\mathfrak{h}_j$ and $\mathfrak{h}_{-j}$ are defined as in (4.54) and (4.55). Then $(\sigma_j^*,\sigma_{-j}^*)$ is a pair of stationary feedback-Nash equilibrium strategies and $V_j^*(\mathbf{x})$ (resp. $V_{-j}^*(\mathbf{x})$) is the current-valued equilibrium value function for Player j (resp. Player $-j$).

For a proof see Haurie et al. (2012).

Remark 4. A feedback equilibrium is, by construction, subgame perfect irrespective of whether the game is played over finite or infinite horizon.

Remark 5. Theorems 1 and 2 (i.e., the verification theorems) provide *sufficient* conditions for feedback-Nash (Markovian) equilibria. This means that if we manage to solve the corresponding Bellman equations, then we can claim that an equilibrium exists. A consequence of this fact is that the methods for solving Bellman equations are of *prime* interest to economists and managers who wish to characterize or implement feedback equilibria.

Remark 6. The dynamic programming algorithm requires that we determine the value functions $W_j(\mathbf{x})$ (or $W_j(t,\mathbf{x})$ for finite-horizon games), for every $\mathbf{x} \in \mathbb{X} \subset \mathbb{R}^n$. This can be achieved in practice if $\mathbb{X}$ is a finite set, or the system of Bellman equations has an analytical solution that could be obtained through the **method of undetermined coefficients**. If $W_j(\cdot,\mathbf{x})$ is affine or quadratic in $\mathbf{x}$, then that method is easily applicable.

Remark 7. When the dynamic game has linear dynamics and quadratic stage payoffs, value functions are usually quadratic. Such problems are called **linear-quadratic games**. The feedback equilibria of these games can be expressed using coupled **Riccati** equations. Thus, linear-quadratic games with many players (i.e., $m > 2$) can be solved as long as we can solve the corresponding Riccati equations. This can be done numerically for reasonably large numbers of players. For a detailed description of the method for solving linear-quadratic games see Haurie et al. (2012); for the complete set of sufficient conditions for discrete-time linear-quadratic games see Başar and Olsder (1999).

Remark 8. In general, however, when $\mathbb{X}$ is a subset of $\mathbb{R}^n$ and W_j is not linear or quadratic (or of some other predetermined form as in Sect. 5.2 below), the only way to tackle the problem of computing W_j is to approximate $\mathbb{X}$ by a finite set, say $\mathbb{X}_d$, and to compute an equilibrium for the new game with a discrete (grid) state space $\mathbb{X}_d$, where the index d corresponds to the grid width. This equilibrium may or may not converge to the original game equilibrium as $d \to 0$. The practical limitation of this approach is that the cardinality of $\mathbb{X}_d$ tends to be high. This is the well-known **curse of dimensionality** mentioned by Bellman in his original work on dynamic programming.

5.2 An Infinite-Horizon Feedback-Nash Equilibrium Solution to the Fishery's Problem

Next we show how a dynamic game formulated in general in Example 1 can be solved when the players use the fishery in perpetuity. For a solution to the finite-horizon version of this game, we refer the readers to Haurie et al. (2012).[23] Our analysis is based on the dynamic programming method explained in the previous section. This method enables us to find the Markovian strategies in any infinite-horizon dynamic games where the value functions are available in analytical form. Although the model we use is stylized, it highlights several important economic issues arising in competition over a long time horizon.

This game is a special case of Example 1. In particular, we assume a linear state equation:

$$x(t+1) = a(x(t) - u_1(t) - u_2(t)). \tag{4.58}$$

The players' payoffs are as in Example 1:

$$J_j \triangleq \sum_{t=0}^{\infty} \beta^t \sqrt{u_j(t)} \qquad j = 1, 2. \tag{4.59}$$

We endeavor to compute a pair of feedback stationary equilibrium strategies (σ_1^*, σ_2^*) that satisfy (4.56)–(4.57). The strategies must also be feasible, i.e., $x - \sigma_1^*(x) - \sigma_2^*(x) \geq 0$.

We now apply Theorem 2 to this game. The Bellman equations of the players are

[23] The system's dynamics with a linear law of motion is an example of a BIDE model (Birth, Immigration, Death, Emigration model, see, e.g., Fahrig 2002 and Pulliam 1988).

$$W_1^*(x) = \max_{u_1} \left\{ \sqrt{u_1} + \beta W_1^*(a(x - u_1 - \sigma_2^*(x)) \right\} \tag{4.60}$$

$$W_2^*(x) = \max_{u_2} \left\{ \sqrt{u_2} + \beta W_2^*(a(x - \sigma_1^*(x) - u_2) \right\}, \tag{4.61}$$

where $W_1^*(x), W_2^*(x)$ are the (Bellman) value functions and $(\sigma_1^*(x), \sigma_2^*(x))$ is the pair of Markovian equilibrium strategies.

Assuming sufficient regularity of the right-hand sides of (4.60) and (4.61) (specifically differentiability), we derive the first order conditions for the feedback-equilibrium Nash harvest strategies:

$$\begin{cases} \dfrac{1}{2\sqrt{u_1}} = \beta \frac{\partial}{\partial u_1} W_1^*(a(x - u_1 - \sigma_2^*(x)) \\ \\ \dfrac{1}{2\sqrt{u_2}} = \beta \frac{\partial}{\partial u_2} W_2^*(a(x - u_2 - \sigma_1^*(x)). \end{cases} \tag{4.62}$$

Because of symmetry, we will restrict attention to value functions which have the same form for each player. Specifically, we conjecture the following value functions[24]: $W_1^*(x) = C\sqrt{x}$ and $W_2^*(x) = C\sqrt{x}$. Expand the necessary conditions (4.62) and solve them simultaneously to obtain

$$\sigma_1^*(x) = \sigma_2^*(x) = \frac{x}{2 + \beta^2 C^2 a}. \tag{4.63}$$

We will demonstrate later that these strategies are feasible. Substituting these forms for u_1 and u_2 in the Bellman equations (4.60), (4.61) yields the following identity:

$$C\sqrt{x} \equiv \frac{1 + \beta^2 C^2 a}{\sqrt{2 + \beta^2 C^2 a}} \sqrt{x}. \tag{4.64}$$

Hence,

$$C = \frac{1 + \beta^2 C^2 a}{\sqrt{2 + \beta^2 C^2 a}}. \tag{4.65}$$

If the value of C solves (4.65), then $W_1^*(x), W_2^*(x)$ will satisfy the Bellman equations.

Determining C requires solving a double-quadratic equation. The only positive root of that equation is

[24]To solve a functional equation like (4.60), we use the *method of undetermined coefficients*; see Haurie et al. (2012) or any textbook on difference equations.

$$\bar{C} = \frac{1}{\beta}\sqrt{\frac{\frac{1}{\sqrt{1-a\,\beta^2}} - 1}{a}}. \tag{4.66}$$

This root will be real if

$$a\beta^2 < 1\,. \tag{4.67}$$

Remark 9. Equation (4.67) is a necessary condition for the existence of a feedback equilibrium in the fishery's game (4.58)–(4.59). We note that this condition will **not** be satisfied by economic systems with fast growth and very patient (or "forward-looking") players. Consequently, games played in such economies will not have an equilibrium. Intuitively, the players in these economies are willing to wait infinitely many periods to catch an infinite amount of fish. (Compare Lemma 6 and also Lemma 8.)

Substituting C in the strategies and value functions yields

$$\sigma_1^*(x) = \sigma_2^*(x) = \frac{x}{\frac{1}{\sqrt{1-a\,\beta^2}} + 1} \tag{4.68}$$

$$W_1^*(x) = W_2^*(x) = \frac{1}{\beta}\sqrt{\frac{\frac{1}{\sqrt{1-a\,\beta^2}} - 1}{a}}\sqrt{x}\,. \tag{4.69}$$

In macroeconomics, a growth model described by equation (4.58) is called an "AK" model.[25] A known feature of AK models is that they have a unique steady state $x = 0$. A zero steady state would suggest extinction in our fishery's game. Let us examine the conditions under which the equilibrium catch strategies (4.68) will lead to extinction.

First we compute the steady state. Substituting the equilibrium strategies in the state equation (4.58) yields

$$x(t+1) = a\left(x(t) - \frac{x(t)}{\frac{1}{\sqrt{1-a\,\beta^2}} + 1} - \frac{x(t)}{\frac{1}{\sqrt{1-a\,\beta^2}} + 1}\right) =$$
$$= x(t)\cdot a\left(1 - \frac{2}{\frac{1}{\sqrt{1-a\,\beta^2}} + 1}\right) = x(t)\cdot a\,\frac{1-\sqrt{1-a\,\beta^2}}{1+\sqrt{1-a\,\beta^2}} \tag{4.70}$$

[25] Any linear growth model in which capital expands proportionally to the growth coefficient a is called an *AK model*.

We can rewrite (4.70) in the following form:

$$x(t+1) = \lambda x(t), \tag{4.71}$$

where

$$\lambda \equiv a\,\frac{1-\sqrt{1-a\,\beta^2}}{1+\sqrt{1-a\,\beta^2}}$$

is the state equation's *eigenvalue*. Note that if (4.67) is satisfied, we will have $\lambda > 0$.

Difference-equations analysis (see, e.g., Luenberger 1979) tells us that (4.71) has a unique steady state at zero, which is asymptotically stable if $\lambda < 1$ and unstable if $\lambda > 1$. We can now formulate the following lemmas concerning the long-term exploitation of the fishery in game (4.58)–(4.59).

Lemma 6. *Feedback-equilibrium strategies* (4.68) *lead to the fishery's extinction when the steady state of* (4.71) *is asymptotically stable, i.e., when*

$$a < \frac{2}{\beta} - 1.$$

Proof. We need to show that

$$\lambda = a\,\frac{1-\sqrt{1-a\,\beta^2}}{1+\sqrt{1-a\,\beta^2}} < 1\,. \tag{4.72}$$

This inequality implies that

$$\begin{aligned} a\left(1-\sqrt{1-a\,\beta^2}\right) &< 1+\sqrt{1-a\,\beta^2} \\ a-1 &< (a+1)\sqrt{1-a\,\beta^2} \\ \left(-\frac{2}{\beta}-1\right) < a &< \left(\frac{2}{\beta}-1\right). \end{aligned}$$

However, by assumption, the growth coefficient a is bigger than 1. Thus

$$1 < a < \left(\frac{2}{\beta}-1\right). \tag{4.73}$$

◇

We also notice that

$$\left(\frac{2}{\beta} - 1\right) < \frac{1}{\beta^2} \quad \text{for} \quad \beta < 1, \tag{4.74}$$

so (4.73) is compatible with condition (4.67) for existence of equilibrium.

Lemma 7. *Feedback-equilibrium strategies* (4.68) *do* ***not*** *lead to the fishery's extinction if*

$$\frac{2}{\beta} - 1 < a < \frac{1}{\beta^2}\,.$$

When the above condition is satisfied, the steady state of (4.71) *is unstable.*

Proof. We require

$$\lambda = a\,\frac{1 - \sqrt{1 - a\,\beta^2}}{1 + \sqrt{1 - a\,\beta^2}} > 1\,. \tag{4.75}$$

The previous lemma established that the (zero) unique steady state is unstable if

$$a > \frac{2}{\beta} - 1\,.$$

On the other hand, (4.67) has to be satisfied in order to ensure existence of equilibrium. We know from (4.74) that there are values of a which satisfy both. $\diamond$

The above value functions and strategies can also be characterized graphically with Figs. 4.1 and 4.2. In each of these figures, we see two lines in the plane (a, β). The solid line represents $a = \dfrac{1}{\beta^2}$, while the dashed line shows where $a = \dfrac{2}{\beta} - 1$. A simple conclusion drawn from Fig. 4.2 is that when β is higher (i.e., the players are more "forward-looking"), equilibrium catch will be lower.

We can also use these figures to make some interesting observations about fishery's management. From the above lemma, we know that the fishery will become extinct if a, β are below the dashed line. This would happen when the players are *not* sufficiently patient i.e., $\beta < \dfrac{2}{a+1}$. On the other hand, if a, β are above the solid line, the players are "over-patient," i.e., $\beta > \dfrac{1}{\sqrt{a}}$. Then fishing is postponed indefinitely, and so the biomass will grow to infinity. We see that fishing can be profitable and sustainable only if a, β are between the dashed and solid lines (see both figures). Notwithstanding the simplicity of this model (linear dynamics (4.58) and concave utility (4.59)), the narrowness of the region between the lines can

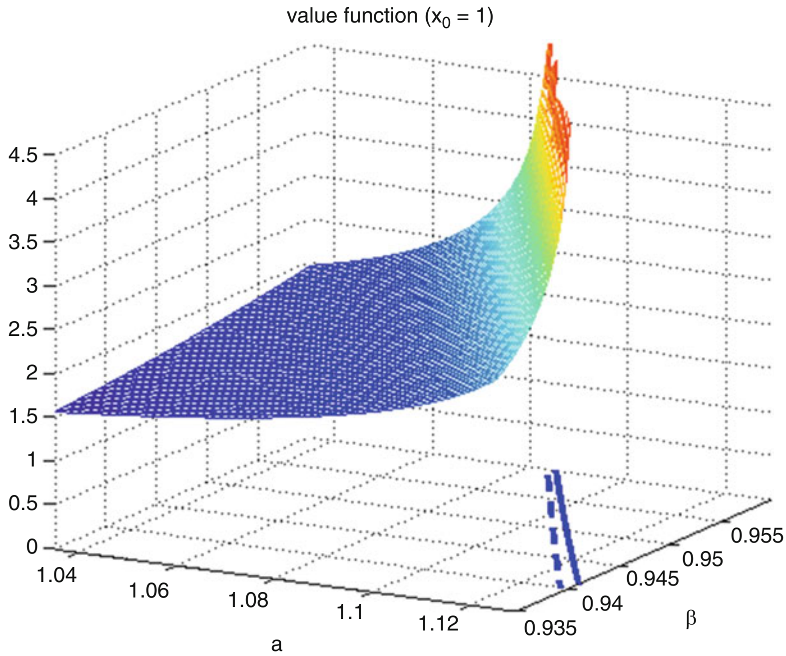


Fig. 4.1 Dependence of value function on β and a

explain why real-life fisheries management can be difficult. Specifically, fishermen's time preferences should be related to biomass productivity in a particular way. But the parameters a and β are exogenous to this model. Thus, their values may well be outside the region between the two lines in Figs. 4.1 and 4.2.

6 Sequential-Move Games

Next we consider several examples of multistage games in which players move sequentially rather than simultaneously as in the previous sections. Again, agents receive payoffs over multiple periods. However, in any given stage, only a single player is able to choose his control, and so only he can affect the subsequent value of the state. We will study infinite-horizon versions of these games and focus on their feedback (or Markovian) equilibria. The sequential-move structure makes these equilibria analytically tractable: they can often be solved in closed form. Also, they provide a good illustration of the strategic motives underlying agents' choices. Specifically, the Markovian equilibria of these games highlight the players' concern regarding the effect of their actions on the future behavior of their opponents.

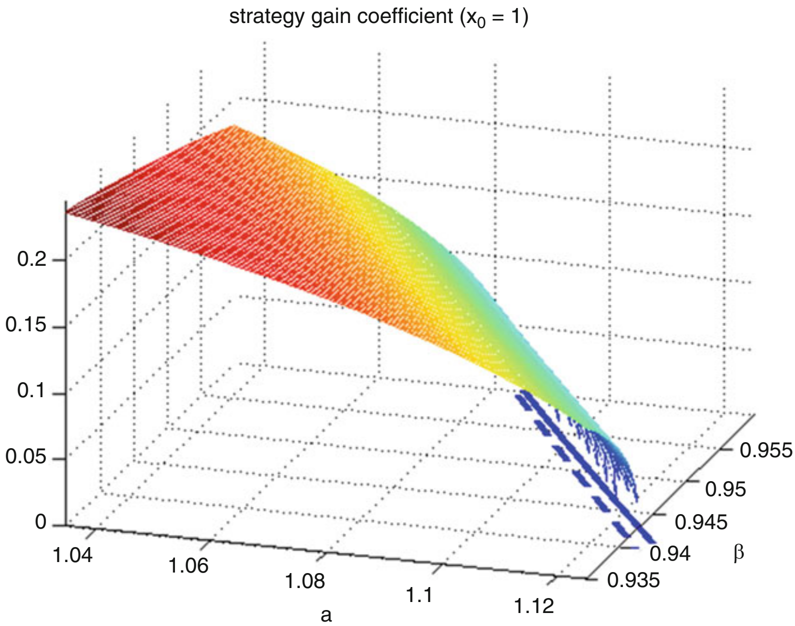


Fig. 4.2 Dependence of strategy on β and a

6.1 Alternating-Move Games

We begin by briefly studying a class of dynamic settings known as alternating-move games. These games have gained popularity in the industrial organization literature. Our discussion will draw primarily from Maskin and Tirole (1987). In their model, two firms, j and $-j$, engage in output competition. The players set their quantities in alternating periods. Thus, only one firm can (costlessly) change its output in any given stage. During the period following an adjustment, a player's output remains fixed, and can be observed by his opponent. This assumption is meant to capture the idea of *short-term commitment*.[26]

Without loss of generality, assume that j makes his choices u_j in odd periods, while his opponent makes his choices u_{-j} in even periods. Thus, the state equation is

$$x(t) = \begin{cases} u_{-j}(t-1) \text{ when } t = 2k+1 \\ u_j(t-1) \text{ when } t = 2k. \end{cases}$$

Firm j's profit is $g_j(u_j(t), x(t)) \equiv g_j(u_j(t), u_{-j}(t-1))$ in odd periods, and $g_j(x(t), u_{-j}(t)) \equiv g_j(u_j(t-1), u_{-j}(t))$ in even periods. The other player's stage

[26] Alternatively, we could postulate that the cost of adjusting output in the subsequent period is infinite.

payoff is symmetrical. Agents have a common discount factor δ. Each maximizes his infinite discounted stream of profits.

Note that this setting satisfies the description of a multistage game from Sect. 2. The state of the dynamic system evolves according to a well-defined law of motion. Furthermore, each player receives instantaneous rewards that depend on the controls and the state. However, in contrast with the previous examples, only a subset of players are able to influence the state variable in any given stage.

We will characterize the Markovian equilibrium of this game, in which player j uses a strategy $\sigma_j(x(t))$ in odd periods, and player $-j$ uses a strategy $\sigma_{-j}(x(t))$ in even periods. Maskin and Tirole interpret these Markovian strategies as *dynamic reaction functions*.

Let us consider the problem of firm j. When $t = 2k + 1$, $k = 0, 1, 2, \ldots$, this player's Bellman equation is

$$V_j(u_{-j}(t-1)) = \max_{u_j(t)}\{g_j(u_j(t), u_{-j}(t-1)) + \delta W_j(u_j(t))\}, \tag{4.76}$$

where V_j is his value function when he can choose his output, and W_j is his value function when he cannot. In a Markovian equilibrium, W_j satisfies

$$W_j(u_j(t)) = g_j(u_j(t), \sigma_{-j}(u_j(t))) + \delta V_j(\sigma_{-j}(u_j(t))). \tag{4.77}$$

After substituting W_j in (4.76), the Bellman equation can be rewritten as

$$\begin{aligned} V_j(u_{-j}(t-1)) = &\max_{u_j(t)}\{g_j(u_j(t), u_{-j}(t-1)) + \delta g_j(u_j(t), \sigma_{-j}(u_j(t))) \\ &+ \delta^2 V_j(\sigma_{-j}(u_j(t)))\}. \end{aligned}$$

The first-order condition of the maximization problem is

$$\begin{aligned} &\frac{\partial}{\partial u_j(t)} g_j(u_j(t), u_{-j}(t-1)) + \delta \frac{\partial}{\partial u_j(t)} g_j(u_j(t), u_{-j}(t+1)) + \delta \frac{\partial}{\partial u_j(t)} \sigma_{-j}(u_j(t)) \\ &\left[\frac{\partial}{\partial u_{-j}(t+1)} g_j(u_j(t), u_{-j}(t+1)) + \delta \frac{\partial}{\partial u_{-j}(t+1)} V_j(u_{-j}(t+1))\right] = 0. \end{aligned} \tag{4.78}$$

Differentiating both sides of the Bellman equation with respect to the state variable $x(t) = u_{-j}(t-1)$ yields the envelope condition

$$\frac{\partial}{\partial u_{-j}(t-1)} V_j(u_{-j}(t-1)) = \frac{\partial}{\partial u_{-j}(t-1)} g_j(u_j(t), u_{-j}(t-1)). \tag{4.79}$$

Substituting (4.79) into (4.78) gives us firm j's equilibrium condition:

$$\frac{\partial}{\partial u_j(t)} g_j(u_j(t), u_{-j}(t-1)) + \delta \frac{\partial}{\partial u_j(t)} g_j(u_j(t), u_{-j}(t+1)) + \delta \frac{\partial}{\partial u_j(t)} \sigma_{-j}(u_j(t))$$
$$\left[\frac{\partial}{\partial u_{-j}(t+1)} g_j(u_j(t), u_{-j}(t+1)) + \delta \frac{\partial}{\partial u_{-j}(t+1)} g_j(u_j(t+2), u_{-j}(t+1)) \right] = 0. \tag{4.80}$$

The term on the second line of (4.80) represents the strategic effect of j's choice on his lifetime profit. It accounts for the payoff consequences of $-j$'s reaction if j were to marginally deviate from his equilibrium strategy in the current period.

Suppose that stage profits are quadratic:

$$g_j = (1 - u_j - u_{-j})u_j, \quad g_j = (1 - u_j - u_{-j})u_{-j}.$$

We conjecture a symmetric Markovian equilibrium involving linear strategies:

$$\sigma_j = a - bx, \quad \sigma_{-j} = a - bx.$$

Substituting these forms in (4.80) and applying the method of undetermined coefficients yield the following equations for a, b:

$$\delta^2 b^4 + 2\delta b^2 - 2(1-\delta)b + 1 = 0, \quad a = \frac{1+b}{3-\delta b}. \tag{4.81}$$

The steady-state output per firm is $1/(3 - \delta b)$. It exceeds $1/3$, which is the Nash equilibrium output in the one-shot simultaneous-move Cournot game. This result makes intuitive sense. In an alternating-move setting, agents' choices have commitment power. Given that firms compete in strategic substitutes, they tend to commit to higher output levels, exceeding those that would arise in the Nash equilibrium of the one-shot game.

6.2 Bequest Games

Next we analyze dynamic games in which agents make choices only once in their lifetimes. However, these agents continue to obtain payoffs over many periods. Consequently, their payoffs will depend on the actions of their successors. Discrepancies between the objectives of the different generations of agents create a conflict between them, so these agents will behave strategically.

A simple example of such a setting is the bequest game presented in Fudenberg and Tirole (1991). In this model, generations have to decide how much of their capital to consume and how much to bequeath to their successors. It is assumed that each generation lives for two periods. Hence, it derives utility only from its own consumption and from that of the subsequent generation, but not from the consumption of other generations. Specifically, let the period-t consumption be $u(t) \geq 0$. The payoff of the period-t generation is $g(u(t), u(t+1))$. Suppose that the stock of capital available to this generation is $x(t)$. Capital evolves according to the following state equation:

$$x(t+1) = f(x(t) - u(t)). \tag{4.82}$$

The function f can be interpreted as a production function. Assume that $f(0) = 0$.

The above payoff specification gives rise to a strategic conflict between the current generation and its successor. If the period-t generation could choose $u(t+1)$, it would not leave any bequests to the period-$t+2$ generation. However, the period-$t+1$ agent will save a positive fraction of his capital, as he cares about $u(t+2)$. Thus, from the period-t viewpoint, period-$t+1$ consumption would be too low in equilibrium.

Again, we study the Markovian equilibrium of this game, in which the period-t consumption strategy is $\sigma(x(t))$. This strategy must be optimal for the current generation, so long as future generations adhere to the same strategy. That is, σ must satisfy

$$\sigma(x(t)) = \arg\max_{u(t)} g(u(t), \sigma(f(x(t) - u(t)))).$$

Let us consider a parametric example. Suppose that

$$g(u(t), u(t+1)) = \ln u(t) + \delta \ln u(t+1). \tag{4.83}$$

Moreover, we assume the following production function:

$$f(x(t) - u(t)) = (x(t) - u(t))^{\alpha}.$$

The parameters δ, α are in the interval $(0, 1)$. We conjecture a Markovian strategy with the form $\sigma(x(t)) = \lambda x(t)$. If future generations follow such a strategy, the payoff of the current generation can be written as

$$\ln u(t) + \delta \ln \lambda (x(t) - u(t))^{\alpha}.$$

The period-t optimal consumption satisfies the first-order condition

$$\frac{1}{u(t)} - \frac{\alpha\delta}{x(t) - u(t)} = 0. \tag{4.84}$$

Solving the above equation for $u(t)$ gives us $u(t) = x(t)/(1 + \alpha\delta)$. But, by conjecture, the equilibrium strategy is $\lambda x(t)$. Hence $\lambda = 1/(1 + \alpha\delta)$. Since $0 < \alpha\delta < 1$, the marginal propensity to consume capital will be between 0.5 and 1.

6.3 Intrapersonal Games and Quasi-hyperbolic Discounting

In most of the economic literature, the standard framework for analyzing intertemporal problems is that of discounted utility theory, see e.g., Samuelson (1937). This framework has also been used extensively throughout this chapter: in all infinite-horizon settings, we assumed a constant (sometimes agent-specific) discount factor that is independent of the agent's time perspective. This type of discounting is also known as *exponential*. Its popularity is due to its analytical convenience. As demonstrated earlier, exponential discounting allows us to use dynamic programming to characterize the decision-maker's optimal policy. However, there are many situations in which human behavior cannot be explained with this type of discounting. In fact, experimental and empirical studies agree that, in the real world, agents' time preferences exhibit "present bias" or "increasing patience." In particular, people are more patient regarding intertemporal trade-offs that will happen further in the future. As in bequest games, such preferences would cause a discrepancy between the objective of the current decision-maker and those of his successors.

Discounting that exhibits increasing patience over the entire planning horizon is generally known as "hyperbolic." However, models based on true hyperbolic preferences are intractable: they are not amenable to recursive formulation and often give rise to infinite payoffs. To avoid these difficulties, the existing literature usually approximates them with an alternative specification referred to as "quasi-hyperbolic" (or beta-delta) discounting. It was first proposed by Phelps and Pollak (1968) in the context of an intergenerational bequest game. Later it was used by Laibson (1997) to study the savings behavior of a consumer with self-control problems who has access to imperfect commitment devices (e.g., illiquid assets). Quasi-hyperbolic preferences have also been applied to the problems of procrastination, see O'Donoghue and Rabin (1999); retirement decisions, see Diamond and Köszegi (2003); asset pricing, see Kocherlakota (2001); job search, see Paserman (2008); growth, see Barro (1999); and addiction, see Gruber and Koszegi (2001).

Quasi-hyperbolic (or beta-delta) preferences capture the spirit of increasing patience by specifying a lower discount factor only for the trade-off between the current period and the one that follows immediately. For any other two consecutive future periods, the discount factor is higher and invariant to the agent's perspective. Formally, the performance criterion payoff of a period-t agent with quasi-hyperbolic preferences is defined as

$$J \triangleq g(x(t), u(t)) + \beta \sum_{\tau=t+1}^{\infty} \delta^{\tau} g(x(\tau), u(\tau)), \tag{4.85}$$

where $\beta, \delta \in (0, 1)$. Note that when $\beta = 1$ this specification is reduced to standard exponential discounting.

As Strotz (1955) has demonstrated, unless agents are exponential discounters, in the future they will be tempted to deviate from the plans that are currently considered optimal. There are several ways of modeling the behavior of quasi-hyperbolic decision-makers. In this section, we assume that they are sophisticated. That is, they are aware of future temptations to deviate from the optimal precommitment plan. We will model their choices as an intrapersonal dynamic game. In particular, each self (or agent) of the decision-maker associated with a given time period is treated as a separate player. Our focus will be on the feedback (Markovian) equilibrium of this game. In other words, we consider strategies that are functions of the current state: $u(t) = \sigma(x(t))$. If all selves adhere to this strategy, the resulting plan will be subgame perfect.

To study the internal conflict between the period-t self and its successors, we need to impose some structure on stage payoffs. In particular, suppose that

$$\frac{\partial}{\partial u(t)} g(x(t), u(t)) > 0, \quad \frac{\partial}{\partial x(t)} g(x(t), u(t)) \leqslant 0.$$

Furthermore, we assume that the state variable $x(t)$ evolves according to a linear law of motion: $x(t+1) = f(x(t), u(t))$, where

$$f(x(t), u(t)) = ax(t) + bu(t).$$

Now we derive a difference-differential equation for the equilibrium Markovian strategy following the method of Laibson (1996), which makes use of two value functions. The strategy solves the Bellman equation

$$V(x(t)) = \max_{u(t)} \{g(x(t), u(t)) + \delta\beta W(ax(t) + bu(t))\}, \tag{4.86}$$

where V is the period-t agent's current value function. The function W represents the agent's continuation payoff, i.e., the net present value of all stage payoffs after t. As discounting effectively becomes exponential from period $t+1$ onward, W must satisfy

$$W(x(t)) = g(x(t), \sigma(x(t))) + \delta W(ax(t) + b\sigma(x(t))). \tag{4.87}$$

Differentiating the right-hand side of (4.86) with respect to $u(t)$ yields the first-order condition

$$\frac{\partial}{\partial u(t)} g(x(t), u(t)) + \beta\delta b \frac{\partial}{\partial x(t+1)} W(x(t+1)) = 0.$$

Thus,

$$\frac{\partial}{\partial x(t+1)} W(x(t+1)) = -\frac{1}{b\beta\delta} \frac{\partial}{\partial u(t)} g(x(t), u(t)). \tag{4.88}$$

Moreover, differentiating (4.87) with respect to $x(t)$ delivers

$$\frac{\partial}{\partial x(t)}W(x(t)) = \frac{\partial}{\partial x(t)}\sigma(x(t))\frac{\partial}{\partial u(t)}g(x(t),u(t)) + \frac{\partial}{\partial x(t)}g(x(t),u(t)) + \delta\left(a + b\frac{\partial}{\partial x(t)}\sigma(x(t))\right)\frac{\partial}{\partial x(t+1)}W(x(t+1)).$$

Substitute the derivatives of the continuation payoff function from (4.88) in the above condition to obtain the agent's equilibrium condition:

$$0 = \frac{\partial}{\partial u(t)}g(x(t),u(t)) + b\beta\delta\frac{\partial}{\partial x(t+1)}g(x(t+1),u(t+1)) - \delta\left[a + b(1-\beta)\frac{\partial}{\partial x(t+1)}\sigma(x(t+1))\right]\frac{\partial}{\partial u(t+1)}g(x(t+1),u(t+1)). \tag{4.89}$$

The term

$$-\delta b(1-\beta)\frac{\partial}{\partial x(t+1)}\sigma(x(t+1))\frac{\partial}{\partial u(t+1)}g(x(t+1),u(t+1))$$

in the above equation captures the agent's intrapersonal strategic considerations. Note that this term disappears when $\beta = 1$. It accounts for the discrepancy between the objectives of the period-t and period-$t+1$ selves. Because of this discrepancy, the current agent adjusts his action in order to strategically influence the behavior of his successor.

We now illustrate this intrapersonal game with a specific application. Our discussion follows a simplified version of the consumption-saving model in Laibson (1996). Consider a quasi-hyperbolic consumer whose period-t stock of savings is $x(t)$. Let his current consumption be $u(t)$. It generates utility

$$g(x(t),u(t)) = (u(t)^{1-\rho} - 1)/(1-\rho). \tag{4.90}$$

Savings evolve according to

$$x(t+1) = R(x(t) - u(t)), \tag{4.91}$$

where R is the gross interest rate. We conjecture a Markovian consumption strategy with form $\sigma(x(t)) = \lambda x(t)$. Substitution in (4.89) yields the following equation for the equilibrium value of λ:

$$1 - \frac{\delta R^{1-\rho}[1-\lambda(1-\beta)]}{(1-\lambda)^{\rho}} = 0.$$

In the limiting case when $g(x(t), u(t)) = \ln u(t)$, we can obtain a closed-form solution for λ:

$$\lambda = \frac{1-\delta}{1-\delta(1-\beta)}.$$

From the current viewpoint, future agents are expected to consume too much. The reasoning is that the period-$t+1$ consumer will discount the period-$t+2$ payoff by only $\beta\delta$. However, the period-t agent's discount factor for the trade-off between $t+1$ and $t+2$ is δ. In the example with logarithmic utility, it is easy to show that all agents would be better off if they followed a rule with a lower marginal propensity to consume: $u(t) = (1-\delta)x(t)$. But if no precommitment technologies are available, this rule is not an equilibrium strategy in the above intrapersonal game.

7 Stackelberg Solutions in Multistage Games

Another class of sequential-move games is known as Stackelberg games. In these games, when there are two players, one of them is a leader, and the other one is a follower. But, while in the settings from the previous section players with symmetric rewards use symmetric strategies in equilibrium, Stackelberg games are hierarchically structured: *in each stage*, the leader moves ahead of the follower.

7.1 Definitions

The Stackelberg solution is another approach that can be applied to some multistage games (see e.g., Başar and Olsder 1999). Specifically, it is suitable for dynamic conflict situations in which the *leader* announces his strategy and the *follower* optimizes his policy subject to the constraint of the leader's strategy. The leader is able to infer the follower's reaction to any strategy the leader may announce. Therefore, the leader announces[27] a strategy that optimizes his own payoff, given the predicted behavior of the follower. Generally, any problems in which there is *hierarchy* among the players (such as the *principal-agent* problem) provide examples of that structure.

In Sect. 3.1 we defined several information structures under which a dynamic game can be played. We established that the solution to a game is sensitive to the information structure underlying the players' strategies. Of course, this was also true when we were computing the Nash equilibria in other sections. Specifically, in the case of open-loop Nash equilibrium, strategies are not subgame perfect, while in the case of a feedback-Nash equilibrium, they do have the subgame perfection property.

[27]The announced strategy should be implemented. However, the leader could deviate from that strategy.

In the latter case, if a player decided to reoptimize his strategy for a *truncated* problem (i.e., after the game has started) the resulting control sequence would be the *truncated* control sequence of the full problem. If, on the other hand, a solution was time *inconsistent,* a truncated problem would yield a solution which differs from the truncated solution of the full problem, and thus the players will have an incentive to reoptimize their strategies in future periods.

We will now show how to compute the Stackelberg equilibrium strategies $(\boldsymbol{\gamma}_1^s, \boldsymbol{\gamma}_2^s) \in \Gamma_1 \times \Gamma_2$. Here, the index "1" refers to the leader and "2" denotes the follower; $\Gamma_j, j = 1, 2$ are the strategy sets. As usual, the strategies in dynamic problems are indexed by time $t \in \{0, 1, \dots T\}$, where T may be finite or infinite. The follower's reaction set is the set of his best answers to the leader's strategies:

$$\mathcal{M}(\boldsymbol{\gamma}_1), \quad \boldsymbol{\gamma}_1 \in \Gamma_1. \tag{4.92}$$

The symbol $\mathcal{M}$ denotes the follower's reaction mapping (for details refer to (4.95), (4.96), (4.97), and (4.98)).

The leader has to find $\boldsymbol{\gamma}_1^s$ such that

$$J_1(0, \mathbf{x}^0; \boldsymbol{\gamma}_1^s, \mathcal{M}(\boldsymbol{\gamma}_1^s)) \geq J_1(0, \mathbf{x}^0; \boldsymbol{\gamma}_1, \mathcal{M}(\boldsymbol{\gamma}_1)) \quad \text{for all } \boldsymbol{\gamma}_1 \in \Gamma_1. \tag{4.93}$$

where we have allowed for (4.92).

In general, this problem is rather difficult.

Firstly, the constraint (4.92) in the (variational) problem (4.93) is itself in the form of maximum of a functional.

Secondly, the solution to this optimization problem (if it exists), i.e., (4.93)–(4.92), does not admit a *recursive definition* (see Başar and Olsder 1999) and, therefore, is generically *not* subgame perfect.

We will demonstrate the lack of subgame perfection of the Stackelberg solution with a simple finite (matrix) game. This game will also be used to highlight the differences between three Stackelberg-game equilibria obtained under various information structures.

7.2 A Finite Stackelberg Game

We will illustrate the different Stackelberg-game solutions using a *finite* game, originally studied in Simaan and Cruz (1973). This game has a finite number of strategies, which substantially simplifies its analysis.

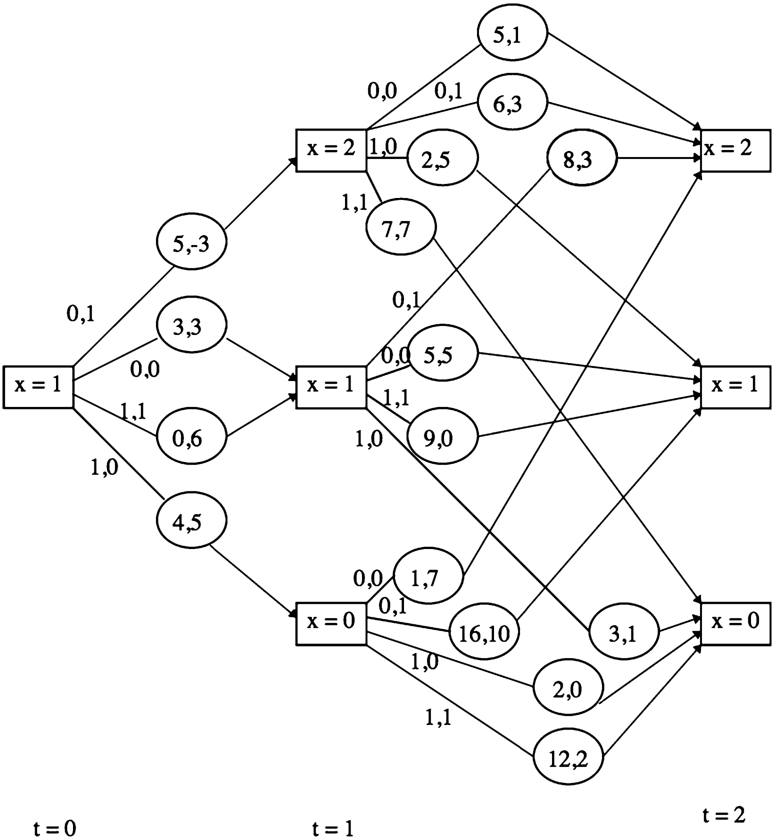


Fig. 4.3 A two-stage game

Example 4. Consider a two-stage finite[28] *minimization* game shown in Fig. 4.3.

This game has three states ($x = 0, x = 1, x = 2$) and is played over two stages ($t = 0$, $t = 1$). At every stage and from every state, each player has a choice between two controls, 0 and 1. The encircled quantities are the costs and the non-encircled ones are controls (first entry – Player 1, second entry – Player 2). Let us compute Stackelberg equilibria for different information structures in this game and investigate whether or not they are subgame perfect.

The game at hand will be played under three different information structures. We will compute equilibria for two of these information structures and, due to

[28]That is, with a finite number of states. While games described by a state equation are typically infinite, matrix games are always finite.

space constraints, sketch the computation process for the third one. In particular, we consider the following structures:

1. Open loop
2. Closed loop
3. Stackelberg feedback

7.3 Open-Loop Information

Assume that the players commit to controls $u_1(t)$ and $u_2(t), t = 0, 1$, which are functions of time only. Since the game has two stages, each player has four possible sequences to choose from. They give rise to a 4×4 bimatrix game shown in Table 4.1. In that table, the row player is the leader, and the column player is the follower.

When Player 1 is the leader, the solution to the game is entry $\{4, 1\}$ (i.e., fourth row, first column) in Table 4.1. The corresponding costs to the leader and follower are 6 and 5, respectively. This solution defines the following control and state sequence:

$$\left.\begin{array}{l} u_1(0) = 1,\ u_2(0) = 0 \Longrightarrow\ x(1) = 0 \\ u_1(1) = 1,\ u_2(1) = 0 \Longrightarrow\ x(2) = 0. \end{array}\right\} \tag{4.94}$$

To show that the solution is $\{4, 1\}$, consider the options available to the leader. If he announces $u_1(0) = 0,\ u_1(1) = 0$, the follower selects $u_2(0) = 1,\ u_2(1) = 0$, and the cost to the leader is 10; if he announces $u_1(0) = 0,\ u_1(1) = 1$, the follower selects $u_2(0) = 1,\ u_2(1) = 0$, and the cost to the leader is 7; if he announces $u_1(0) = 1,\ u_1(1) = 0$, the follower selects $u_2(0) = 1,\ u_2(1) = 1$, and the cost to the leader is 8; and if he announces $u_1(0) = 1,\ u_1(1) = 1$, the follower selects $u_2(0) = 0,\ u_2(1) = 0$, and the cost to the leader is 6, which is minimal.

Subgame perfection means that if the game was interrupted at time $t > 0$, Reoptimization would yield a continuation path which is identical to the remainder of the path calculated at $t = 0$ (see Sect. 3.4 and also Remark 4). If the continuation path deviates from the path that was optimal at $t = 0$, then the equilibrium is not subgame perfect. Therefore, to demonstrate non-subgame perfection of open-loop equilibrium of the game at hand, it is sufficient to show that, if the game was to

Table 4.1 Bimatrix game under open-loop information structure

	00	01	10	11
00	8,8	11,6	10,−2	11,0
01	6,4	12,3	7,2	12,4
10	5,12	20,15	5,11	8,9
11	6,5	16,7	3,7	9,6

Table 4.2 Final stage of the game

	0	1
0	1,7	16,10
1	2,0	12,2

restart at $t = 1$, the players would select actions that are different from (4.94), calculated using Table 4.1.

Let us now compute the players' controls when the game restarts from $x = 0$ at $t = 1$. Since each player would have only two controls, the truncated game (i.e., the stage game at $t = 1$) becomes a 2×2 bimatrix game. We see in Fig. 4.3 that if the players select $u_1(1) = 0,\ u_2(1) = 0$, their payoffs are (1,7); if they select $u_1(1) = 0,\ u_2(1) = 1$, they receive (16,10). The truncated game that is played at state $x = 0$ at time $t = 1$ is shown in Table 4.2.

It is easy to see that the Stackelberg equilibrium solution to this truncated game is entry $\{1, 1\}$. The resulting overall cost of the leader is $4 + 1 = 5$. Note that $\{2, 1\}$, which would have been played if the leader had been committed to the open-loop strategy, is *not* an equilibrium solution: the leader's cost would have been $6 > 5$. Thus, by not keeping his promise to play $u_1(0) = 1,\ u_1(1) = 1$, the leader can improve his payoff.

7.4 Closed-Loop Information

Next, we sketch how a closed-loop solution can be computed for Example 4. Although this game is relatively simple, the computational effort is quite big, so we will omit the solution details due to space constraints.

Assume that, before the start of the game in Example 4, players must announce controls based on where and when they are (and, in general, also where they were). So the sets of admissible controls will comprise of four-tuples ordered, for example, in the following manner:

$$\begin{aligned}
&\text{for } x = 1,\ \ t = 0 && \gamma_{i0}(1) = u_{i0}\\
&\text{for } x = 0,\ \ t = 1 && \gamma_{i1}(0) = u_{i1}\\
&\text{for } x = 1,\ \ t = 1 && \gamma_{i1}(1) = u_{i1}\\
&\text{for } x = 2,\ \ t = 1 && \gamma_{i1}(2) = u_{i1}
\end{aligned}$$

where $i = 1$ corresponds to leader and $i = 2$ corresponds to follower. Since each of u_{it} can be 0 or 1, each player will have 16 closed-loop strategies.

For every such strategy of Player 1, Player 2 has to solve an optimization problem. His strategies can easily be obtained via dynamic programming. For example, if Player 1 chooses $(0, 0, 0, 0)$, i.e., 0 for every state and stage, Player

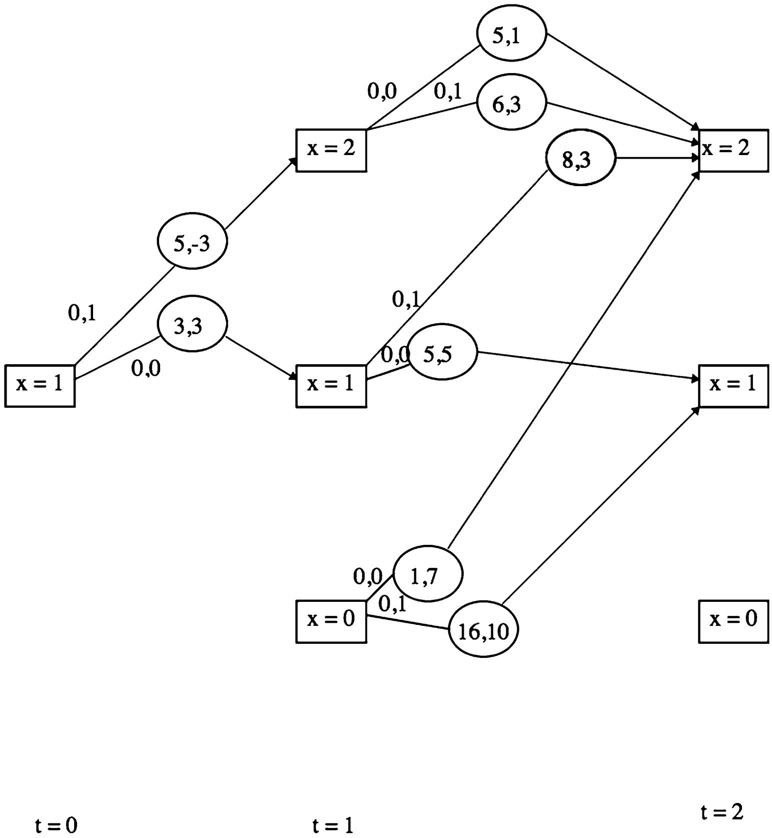


Fig. 4.4 Follower's optimization problem when leader chooses (0, 0, 0, 0)

2's optimization problem is illustrated in Fig. 4.4. Its solution is the four-tuple $(1, 0, 1, 0)$.

To find the leader's optimal closed-loop strategy, the above procedure would have to be repeated for the remaining 15 policies. This would give us the follower's *reaction set*, which is composed of 16 pairs of the four-tuples: $\{leader\ policy, follower\ reaction\}$. The optimal policy of the leader will minimize his cost function on that set. In general, the closed-loop solution differs from the one we obtained under the open-loop information structure . Typically neither solution is subgame perfect.

7.5 Stackelberg Feedback Equilibrium

In Sects. 4, 5, and 6, we computed feedback-Nash equilibria with the help of Bellman equations (i.e., by using dynamic programming). Because of that, these equilibria satisfied the definition of subgame perfection. Now we *define* a solution to hierarchical games that is obtained via dynamic programming, called a *Stackelberg feedback strategy*. This strategy has the property that, at any instant of the game and from any allowable state (at that instant), it delivers the leader's best control *irrespective* of previous decisions, under the assumption that the same strategy will be used for the remainder of the game. Thus, this strategy is subgame perfect *by construction* (see Fudenberg and Tirole 1991 and Başar and Olsder 1999).

Now we will show how to obtain a *feedback Stackelberg* (also called *stagewise* or *successive*) *equilibrium* for an abstract problem. Then we will revisit Example 4.

Consider a dynamic Stackelberg game defined by

$$J_j(\tau, x_\tau; u_1, u_2) = \sum_{t=\tau}^{T-1} g_j(x(t), u_1(t), u_2(t), t) + S_j(x(T)), \tag{4.95}$$

$$x(t+1) = f(x(t), u_1(t), u_2(t), t) \tag{4.96}$$

$$\text{for } i = 1, 2$$

where Player 1 is the leader and Player 2 is the follower. Let, at every t, $t = 0, 1, \ldots T-1$, the players' strategies be feedback. Suppose that the strategies belong to the sets of admissible strategies Γ_{1t} and Γ_{2t}, respectively. When players use Stackelberg feedback strategies, at each $t, x(t)$ we have

$$u_1(t) = \gamma_{1t}(x) \in \Gamma_{1t}(x) \tag{4.97}$$

$$\begin{aligned} u_2(t) &= \tilde{\gamma}_{2t}(x) \in \tilde{\Gamma}_{2t}(x) \equiv \\ \{\tilde{\gamma}_{2t}(x) \ &: \ H_2(t, x(t); \gamma_{1t}(x), \tilde{\gamma}_{2t}(x)) \\ &= \max_{u_{2t} \in \Gamma_{2t}(x)} H_2(t, x(t); \gamma_{1t}(x), u_{2t}(x))\} \end{aligned} \tag{4.98}$$

where

$$H_2(t, x; \gamma_{1t}(x), u_2) \equiv g_2(x, \gamma_{1t}(x), u_2, t) + W_2(t+1, f(x, \gamma_{1t}(x), u_2, t)), \tag{4.99}$$

and

$$W_2(t, x) = H_2(t, x; \gamma_{1t}(x), \tilde{\gamma}_{2t}(x)), \tag{4.100}$$

$$\text{each for } t = T-1, T-2, \ldots, 1, 0,$$

and $W_j(T, x) = S_j(x_T)$ for $i = 1, 2$. A strategy pair $\tilde{\gamma}_t(x) = (\tilde{\gamma}_{1t}(x), \tilde{\gamma}_{2t}(x))$ is called a *Stackelberg feedback equilibrium* at $(t, x(t))$ *if*, in addition to (4.97), (4.98), (4.99), and (4.100) the following relationship is satisfied:

$$H_1(t, x; \tilde{\gamma}_{1t}(x), \tilde{\gamma}_{2t}(x)) = \max_{u_1 \in \Gamma_{1t}(x)} H_1(t, x; u_1, \tilde{\gamma}_{2t}(x)), \tag{4.101}$$

where

$$H_1(t, x; u_1, \tilde{\gamma}_{2t}(x)) \equiv$$
$$g_1(x, u_1, \tilde{\gamma}_{2t}(x), t) + W_1(t+1, f(x, u_1, \tilde{\gamma}_{2t}(x), t)), \tag{4.102}$$

and

$$W_1(t, x) = H_1(t, x; \tilde{\gamma}_{1t}(x), \tilde{\gamma}_{2t}(x)). \tag{4.103}$$

each for $t = T-1, T-2, \ldots, 1, 0,$

Note that (4.97), (4.98), (4.99), (4.100), (4.101), (4.102), and (4.103) provide sufficient conditions for a strategy pair $(\tilde{\gamma}_{1t}(x), \tilde{\gamma}_{2t}(x))$ to be the Stackelberg feedback equilibrium solution to the game (4.95)–(4.96).

Example 5. Compute the Stackelberg feedback equilibrium for the game shown in Fig. 4.3.

At time $t = 1$, there are three possible states. From every state, the transition to the next stage ($t = 2$) is a 2×2 bimatrix game. In fact, Table 4.2 is one of them: it determines the last-stage Stackelberg equilibrium from state $x = 0$. The policies and the costs to the players are

$$(\gamma_{11}(0), \gamma_{21}(0)) = (0, 0), \quad (W_1(1, 0), W_2(1, 0)) = (1, 7).$$

In the same manner, we find that

$$(\gamma_{11}(1), \gamma_{21}(1)) = (0, 1), \quad (W_1(1, 0), W_2(1, 0)) = (8, 3)$$

and

$$(\gamma_{11}(2), \gamma_{21}(2)) = (1, 0), \quad (W_1(1, 2), W_2(1, 2)) = (2, 5).$$

At time $t = 0$, the game looks as in Fig. 4.5. It is easy to show that the Stackelberg feedback policy from $x = 1$ is

$$(\gamma_{10}(1), \gamma_{20}(1)) = (0, 1), \quad (W_1(0, 1), W_2(1, 0)) = (7, 2).$$

7.6 Final Remarks

Remark 10. The fact that the open-loop and closed-loop Stackelberg equilibrium solutions are not subgame perfect casts doubt on the applicability of these solution concepts. The leader may, or may not, prefer to maintain his *reputation* over the gains of a deviation from an announced strategy. If he does, the game will

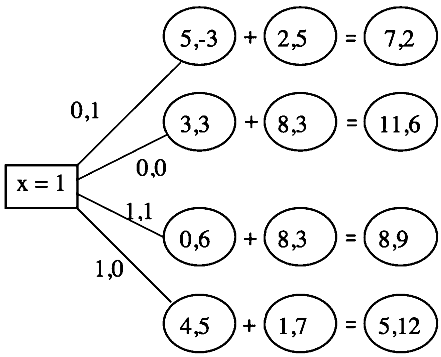


Fig. 4.5 Stackelberg feedback game in $t = 0$

evolve accordingly to these solution concepts. However, if he prefers the gains and actually deviates from the announced policy, his policies lose credibility under these concepts, and he should not expect the follower to respond "optimally." As a result of that, the leader will not be able to infer the constraint (4.92), and the game development may remain undetermined. On the other hand, it seems implausible that human players would consciously forget their previous decisions and the consequences of these decisions. This is what would have to happen for a game to admit a Stackelberg feedback equilibrium.

The most important conclusion one should draw from what we have learned about the various solution concepts and information structures is that their careful specification is crucial for making a game solution realistic.

8 Equilibria for a Class of Memory Strategies in Infinite-Horizon Games

The feedback equilibrium obtained through dynamic programming and illustrated with the fishery-management game in Sect. 5.2 is by no means the only subgame perfect equilibrium of a multistage game. In repeated games, it is well known that other equilibria can be enforced with the help of trigger strategies. Next we characterize a subgame perfect equilibrium of this type in a multistage game played in a state space. The method we will use is generic and can be applied to many dynamic games. Following Krawczyk and Shimomura (2003) (compare Haurie et al. 2012), we show how a trigger-strategy mechanism can be employed to deliver a Pareto-efficient subgame perfect equilibrium in a capital-labor conflict situation.

8.1 Feedback Threats and Trigger Strategies

Definition 3. A **trigger strategy** for Player j is defined by an agreed control sequence $\mathbf{u}^*(\cdot)$ and a feedback law σ_j, which are combined in the following sequence of policies:

$$\gamma_j^t(\mathbf{x}^0, \mathcal{U}_{[0,t-1]}) = \begin{cases} \mathbf{u}_j^*(t) & \text{if } \mathbf{u}(s) = \mathbf{u}^*(s), \quad s = 0, \ldots, t-1 \\ \boldsymbol{\sigma}_j(\mathbf{x}(t)) & \text{otherwise,} \end{cases}$$

where $\mathbf{x}(t)$ is the current state of the system and $\mathcal{U}_{[0,t-1]}$ is the set of admissible controls on $[0, 1, \ldots, t-1]$.

Thus, as long as the controls used by all players in the history of play agree with some nominal control sequence, Player j plays $\mathbf{u}_j^*(t)$ in stage t. If a deviation from the nominal control sequence has occurred in the past, Player j will set his control according to the feedback law σ_j in all subsequent periods. Effectively, this feedback law is a *threat* that can be used by Player j to motivate his opponents to adhere to the agreed (nominal) control sequence.

8.2 A Macroeconomic Problem

One of the main goals of modern growth theory is to examine why countries experience substantial divergences in their long-term growth rates. These divergences were first noted in Kaldor (1961). Since then, growth theorists have paid this issue a great deal of attention. In particular, Romer (1986) heavily criticized the so-called S-S (Solow-Swan) theory of economic growth, because it fails to give a satisfactory explanation for divergences in long-term growth rates.

There have been a few attempts to explain growth rates through dynamic strategic interactions between capitalists and workers. Here we will show that a range of different growth rates can occur in the equilibrium of a dynamic game played between these two social classes.

Industrial relations and the bargaining powers of workers and capitalists differ across countries, as do growth rates. Bargaining power can be thought of as the weight the planner places on the utility of a given social class in his utility function. More objectively, the bargaining power of the worker class can be linked to the percentage of the labor force that is unionized and, by extension, to how frequently a labor (or socio-democratic) government prevails.

The argument that unionization levels have an impact on a country's growth rate has been confirmed empirically. Figure 4.6 shows a cross-sectional graph of the average growth rate in the period 1980–1988 plotted against the corresponding

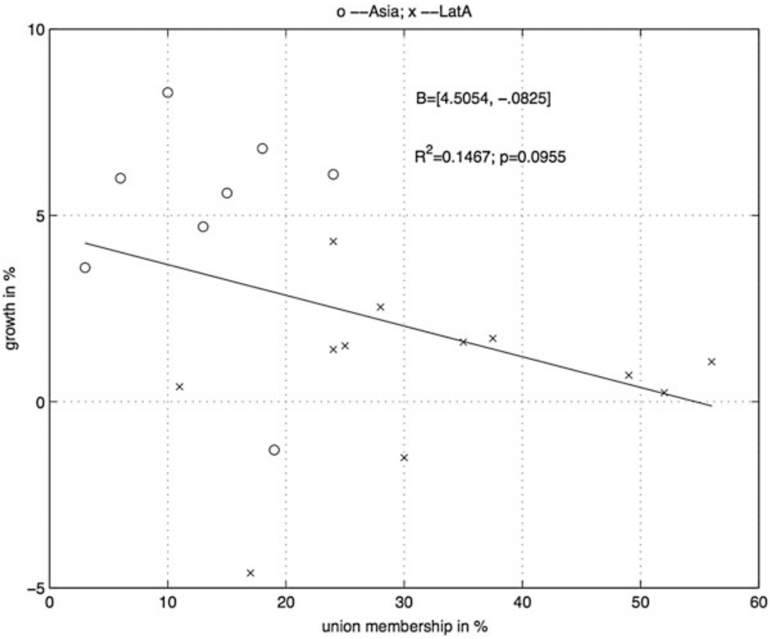


Fig. 4.6 Growth rates and unionization levels

percentage of union membership for 20 Asian and Latin American countries.[29] About 15 % of the growth rate variability can be explained by union membership.

The above observations motivate our claim that differences in countries' industrial relations are at least partly responsible for the disparities in growth performance.

To analyze these issues, we incorporate two novel features into a game-theoretic growth model. First, we employ a noncooperative (self-enforcing) equilibrium concept that we call a *collusive equilibrium*. Essentially, a collusive-equilibrium strategy consists of a pair of trigger policies which initiate punishments when the state deviates from the desired trajectory. Second, we assume that capitalists and workers have different utility functions. Specifically, workers derive utility from consumption, while capitalists derive utility from both consumption and possession of capital.

We use a simple model of endogenous growth proposed in Krawczyk and Shimomura (2003) which incorporates these two new features. It enables us to show that:

(i) There exist an uncountably large number of collusive-equilibrium paths associated with different long-term growth rates.

[29] Source DeFreitas and Marshall (1998).

(ii) The growth rate disparities may be driven by differences in the *bargaining power* of the workers and capitalists.

Thus, our multistage game demonstrates that economies can grow at different rates even when they have the same economic fundamentals (other than industrial relations).

8.3 A Multistage Game Over an Endogenous Growth Model

Consider an economy with two players (social classes): *capitalists* and *workers*.

The state variable is capital $x(t) \geq 0$. It evolves according to the state equation

$$x(t+1) = ax(t) - w(t) - c(t), \tag{4.104}$$

where a is a productivity parameter, $c(t) \geq 0$ is capitalists' consumption, and $w(t) \geq 0$ represents the workers' wages.

The utility functions of the capitalists and the workers are as follows:

$$J_c \triangleq \sum_{t=\tau}^{\infty} \beta^t \left(B \cdot (x(t))^\nu + D \cdot (c(t))^\nu\right) \qquad B, D > 0 \tag{4.105}$$

$$J_w \triangleq \sum_{t=\tau}^{\infty} \beta^t F \cdot (w(t))^\nu \qquad F > 0\,, \tag{4.106}$$

where $\nu \in (0, 1)$ and β is a common discount factor.

The capitalists and the workers are playing a dynamic game. They choose feasible strategies $(\tilde{c}, \tilde{w})$ such that (4.105) and (4.106) are jointly maximized for a historically given initial capital x^τ in period τ. We will use J_c and J_w to denote the *total utility measures,* i.e., the discounted utility sums in period t for the capitalists and the workers, respectively.

8.3.1 Solution Concepts

Suppose that the players can observe the state variable. Then the game solution can be described by a pair of feedback, or Markovian, stationary strategies for capitalists' consumption and workers' wages, as follows:

$$\sigma_c(x), \quad \sigma_w(x). \tag{4.107}$$

Depending on how the players organize themselves, we can expect the game to admit:

- A feedback-Nash equilibrium $\left(\sigma_c(x), \sigma_w(x)\right)$ (noncooperative)
- A Pareto-efficient solution $\left(c^\alpha(x), w^\alpha(x)\right)$ (cooperative), where α is the weight the planner assigns to the payoff of the capitalists (or their bargaining power)

or assuming that the players can observe the actions of their opponents

- A collusive equilibrium (noncooperative) enforced with a pair of stationary trigger strategies

$$\delta_c(x, \mathsf{y}), \ \ \delta_w(x, \mathsf{y}),$$

where y is an auxiliary state variable to be defined below. This solution concept can yield a solution that is both Pareto efficient and subgame perfect. In reality, such an equilibrium may require players to engage in negotiations (e.g., about wages) before the game starts.

8.4 Game Solutions

8.4.1 A Feedback-Nash Equilibrium Solution

Following Theorem 2, the strategy pair $(\sigma_c(x), \sigma_w(x))$ is a feedback-Nash equilibrium solution if and only if there exist value functions $V_c(x)$ and $V_w(x)$ that satisfy the Bellman equations:

$$V_c(x) = \max_c \left\{Bx^{\nu} + Dc^{\nu} + \beta V_c\big(ak - c - \sigma_w(x)\big)\right\} \tag{4.108}$$

$$V_w(x) = \max_w \left\{Fw^{\nu} + \beta V_w\big(ak - \sigma_c(x) - w\big)\right\}. \tag{4.109}$$

To determine the feedback-Nash equilibrium, we need to solve these Bellman equations.

As in Sect. 5.2, we will use the method of undetermined coefficients. Let us conjecture that the players' value functions are as follows:

$$V_c(x) = \Gamma_c x^{\nu} \qquad V_w(x) = \Gamma_w x^{\nu}. \tag{4.110}$$

We will show that these functions satisfy equations (4.108) and (4.109). Substituting the conjectured value functions in (4.108)–(4.109) and using the first-order condition for the maxima of the right-hand sides, we obtain the following Markovian-equilibrium strategies:

$$\sigma_c(x(t)) = \left(\frac{\beta}{D}\Gamma_c\right)^{\frac{1}{\nu-1}} x(t) \equiv \overline{c}x(t) \ \ \text{and} \ \ \sigma_w(x(t)) = \left(\frac{\beta}{F}\Gamma_w\right)^{\frac{1}{\nu-1}} x(t) \equiv \overline{w}x(t)\,. \tag{4.111}$$

If we find nonnegative values of $\overline{c}$, $\overline{w}$, Γ_c, and Γ_w, then we will know that a feedback-Nash (Markovian) equilibrium with strategies specified by (4.111) exists.

Allowing for (4.111), differentiation yields first-order conditions and the envelope conditions:

$$\nu D\overline{c}^{\nu-1}x^{\nu-1} = \beta\nu\Gamma_c(a-\overline{c}-\overline{w})^{\nu-1}x^{\nu-1} \tag{4.112}$$

$$\nu F\overline{w}^{\nu-1}x^{\nu-1} = \beta\nu\Gamma_w(a-\overline{c}-\overline{w})^{\nu-1}x^{\nu-1} \tag{4.113}$$

and

$$\nu\Gamma_c x^{\nu-1} = \nu B x^{\nu-1} + \beta\nu(a-\overline{w})\Gamma_c(a-\overline{c}-\overline{w})^{\nu-1}x^{\nu-1} \tag{4.114}$$

$$\nu\Gamma_w x^{\nu-1} = \beta\nu(a-\overline{c})\Gamma_w(a-\overline{c}-\overline{w})^{\nu-1}x^{\nu-1}, \tag{4.115}$$

respectively. Assuming that the undetermined coefficients are nonzero, from (4.112), (4.113), (4.114), and (4.115) we can derive a system of four equations for the four unknown coefficients:

$$D\overline{c}^{\nu-1} = \beta\Gamma_c(a-\overline{c}-\overline{w})^{\nu-1} \tag{4.116}$$

$$F\overline{w}^{\nu-1} = \beta\Gamma_w(a-\overline{c}-\overline{w})^{\nu-1} \tag{4.117}$$

$$\Gamma_c = B + \beta(a-\overline{w})\Gamma_c(a-\overline{c}-\overline{w})^{\nu-1} \tag{4.118}$$

$$1 = \beta(a-\overline{c})(a-\overline{c}-\overline{w})^{\nu-1}. \tag{4.119}$$

Repeating substitutions delivers the following equation for the capitalists' *gain coefficient* $\overline{c}$:

$$Q(\overline{c}) \equiv a - 2\overline{c} - \left(\frac{B}{D}\right)\overline{c}^{1-\nu} - \beta^{\frac{1}{1-\nu}}(a-\overline{c})^{\frac{1}{1-\nu}} = 0. \tag{4.120}$$

Notice that $Q(0)$ must be positive for (4.120) to have a solution in $[0, a]$. Imposing $Q(0) = a - (\beta a)^{\frac{1}{1-\nu}} > 0$ gives us

$$\beta a^{\nu} < 1. \tag{4.121}$$

To ensure a unique feedback-Nash equilibrium of the capitalist game, we require a unique solution to (4.120). For large $\overline{c}$ (e.g., $\overline{c} > \frac{a}{2}$), $Q(\overline{c}) < 0$, and the graph of $Q(\overline{c})$ crosses 0 at least once. The intersection will be unique if

$$\frac{dQ(\overline{c})}{d\overline{c}} = -2 - (1-\nu)\left(\frac{B}{D}\right)\overline{c}^{-\nu} + \frac{1}{1-\nu}\beta^{\frac{1}{1-\nu}}(a-\overline{c})^{\frac{1}{1-\nu}-1} < 0 \tag{4.122}$$

for all $\overline{c} \in (0, a)$. This inequality is difficult to solve. However, notice that the second term in the above expression is always negative. Thus, it only helps achieve the desired monotonic decrease of $Q(\overline{c})$; in fact, (4.122) will always be negative for sufficiently large B/D. We will drop this term from (4.122) and prove a stronger (sufficient) condition for a unique solution to (4.120). Using (4.121) we get

$$\frac{dQ(\overline{c})}{d\overline{c}} < -2+\frac{1}{1-\nu}\beta^{\frac{1}{1-\nu}}(a-\overline{c})^{\frac{1}{1-\nu}-1} < -2+\frac{1}{1-\nu}(\beta a^{\nu})^{\frac{1}{1-\nu}} < -2+\frac{1}{1-\nu} < 0. \tag{4.123}$$

The above inequality holds for

$$\nu < \frac{1}{2}. \tag{4.124}$$

Hence there is a unique positive $\overline{c} \in (0, a)$ such that $Q(\overline{c}) = 0$ when conditions (4.121) and (4.124) are satisfied.

Once $\overline{c}$ has been obtained, $\overline{w}$ is uniquely determined by (4.119). In turn, Γ_c and Γ_w are uniquely determined by (4.116) and (4.117), respectively. Note that all constants obtained above are positive under (4.121).

Lemma 8. *A pair of linear feedback-Nash equilibrium strategies* (4.111) *and a pair of value functions* (4.110) *which satisfy equations* (4.108) *and* (4.109) *uniquely exist if*

$$a^{\nu} < \frac{1}{\beta} < a \qquad \text{and} \qquad \nu < \frac{1}{2}.$$

Moreover, if B/D is sufficiently large, the growth rate $\mu = \dfrac{x(t+1)}{x(t)}$, $t = 0, 1, 2, \ldots$ corresponding to the equilibrium strategy pair is positive and time invariant.

For a proof see Krawczyk and Shimomura (2003) or Haurie et al. (2012).

The above lemma establishes a relationship between capital productivity, the discount factor, and the degree of agent risk aversion (i.e., ν) which ensures that an equilibrium will exist. For example, a feedback-Nash equilibrium in our capitalist game would be *unlikely* to exist in economies with very high productivity of capital ($a \gg 1$), a high discount factor (β close to 1), and almost risk-neutral agents ($\nu \approx 1$). Note that the conditions in the above lemma are sufficient and may be weakened for large values of B/D. The economic interpretation of a large B/D is that the capitalists have a high preference for capital in their utility function. Thus, the "greedier" the capitalists, the stronger the growth.

Equation (4.120), as well as the other conditions determining $\overline{c}, \overline{w}, \Gamma_c$, Γ_w, and the growth rate, can be solved for numerically. These coefficients are needed to define the collusive equilibrium, and so we have computed them for a number of different values of the discount factor. The remaining parameters of the economy are assumed to be:

$a = 1.07$	7 % capital productivity
$\nu = 0.2$	utility function exponent
$B = 1$	capitalists' weight for "passion for accumulation"
$D = 1$	capitalists' weight for "desire for enjoyment"
$F = 1$	workers' weight of utility from consumption
$0.98 \le \beta \le 0.984$	discount-factor range corresponding to "true" discount rate $\in [1.6, 2]$ %
$x^0 = 1$	capital stock

Evidently, each pair of strategies (4.111) depends on the discount rate. Furthermore, for each discount rate, the strategy pair (4.111) generates a distinct pair of utility measures (see Fig. 4.7).

8.4.2 Pareto-Efficient Solutions

Suppose that the players have agreed to act jointly to maximize

$$\alpha J_c^{\alpha} + (1 - \alpha) J_w^{\alpha} ,$$

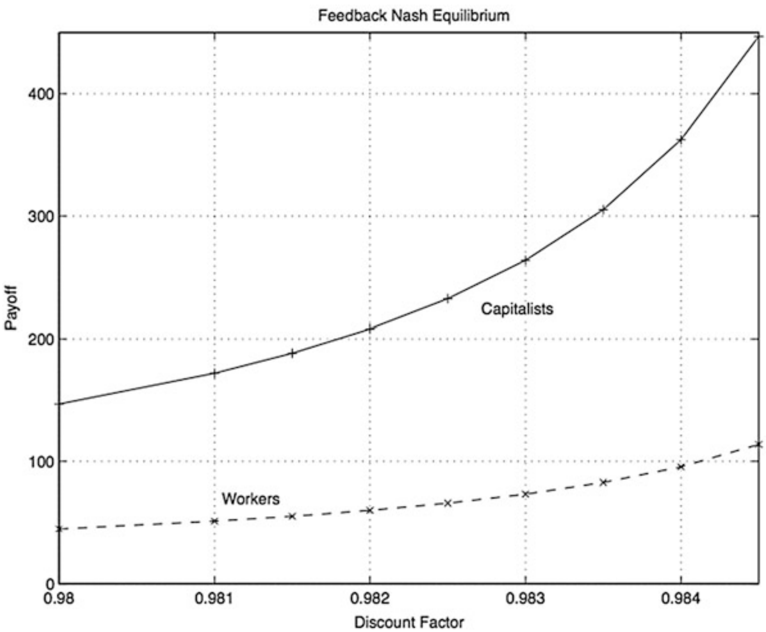


Fig. 4.7 Feedback-Nash equilibrium utility measures

where the parameter $\alpha \in [0, 1]$ is a "proxy" for the capitalists' bargaining power, and J_c^α, J_w^α are defined as in (4.105) and (4.106), with a superscript α added[30] to recognize that the objective values will depend on bargaining power.

Let us derive a pair of Pareto-optimal strategies. We compute them as a joint maximizer of the Bellman equation

$$U(x) = \max_{c,w} \{\alpha(Bx^\nu + Dc^\nu) + (1-\alpha)Fw^\nu + \beta W_1(ax - c - w)\}. \qquad (4.125)$$

We specify the Pareto-efficient strategies for c and w and the value function U as the following functions of state $x(t)$:

$$c(x(t)) = \underline{c}x(t) \quad w(x(t)) = \underline{w}x(t) \qquad (4.126)$$

$$U^\alpha(x) = \Gamma x^\nu. \qquad (4.127)$$

From duopoly theory, we expect $\underline{c} \leq \overline{c}$ and $\underline{w} \leq \overline{w}$.

The first-order and envelope conditions yield

$$\alpha D\underline{c}^{\nu-1} = \beta\Gamma(a - \underline{c} - \underline{w})^{\nu-1} \qquad (4.128)$$

$$\alpha D\underline{c}^{\nu-1} = (1-\alpha)F\underline{w}^{\nu-1} \qquad (4.129)$$

$$\Gamma = \alpha B + \beta a(a - \underline{c} - \underline{w})^{\nu-1}\Gamma. \qquad (4.130)$$

By solving equations (4.128), (4.129), and (4.130), we obtain the Pareto-efficient strategies (4.126) and the value function (4.127). Similar substitutions as on page 203 yield the following equation:

$$R(\underline{c}) \equiv D\left(a - \underline{c}\left(1 + \left(\frac{(1-\alpha)F}{\alpha D}\right)^{\frac{1}{1-\nu}}\right)\right)^{1-\nu} - \beta(B\underline{c}^{1-\nu} + aD) = 0. \qquad (4.131)$$

Reasoning analogous to the one we used to solve equation (4.120) leads us to the conclusion that there exists a unique positive $\underline{c}$ such that $R(\underline{c}) = 0$ if and only if

$$\beta a^\nu < 1 \qquad (4.132)$$

holds.

Lemma 9. *A pair of linear Pareto-optimal strategies uniquely exists if and only if (4.132) is satisfied. Moreover, the growth rate μ corresponding to the strategy pair is positive and time invariant, even if $B = 0$.*

---

[30]In fact, the coefficients $\underline{c}$, $\underline{w}$, and Γ will all depend on α. However, to simplify notation, we will keep these symbols nonindexed.

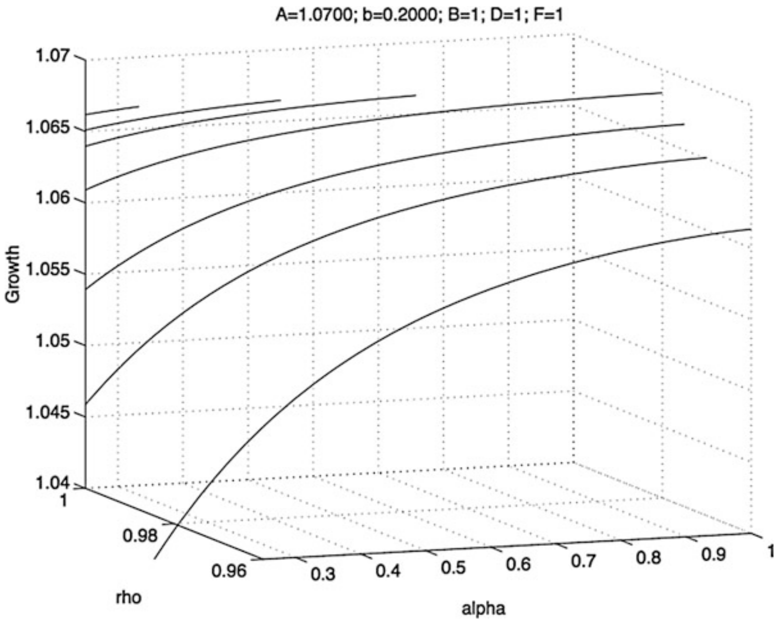


Fig. 4.8 Pareto-efficient growth rates

For a proof see Krawczyk and Shimomura (2003) or Haurie et al. (2012).

Notice that condition (4.132) is identical to the one needed for the existence of a feedback-Nash equilibrium. However, Pareto-efficient strategies can generate positive growth even if capitalists are not very "passionate" about accumulation.

Equation (4.131) and the remaining conditions determining $\underline{c}, \underline{w}, \Gamma$, and the growth rate can be solved numerically. Each pair of strategies will depend on α (the bargaining power) and β (the discount factor) and will result in a different growth rate as shown in Fig. 4.8. These growth rates generate different pairs of utility measures $(J_w^\alpha, J_c^\alpha) =$ (Workers, Capitalists); see Fig. 4.9, where the "+" represents the growth rate obtained for the feedback-Nash equilibrium.

8.4.3 Collusive Equilibria

Now we define a pair of strategies

$$\delta_c^\alpha(x, y), \quad \delta_w^\alpha(x, y)$$

which enforce the payoffs $\big(c^\alpha(x), w^\alpha(x)\big)$ associated with the Pareto-efficient strategies. We will show that $(\delta_c^\alpha, \delta_w^\alpha)$ can be subgame perfect.

Assume that the game is played under the *feedback information structure with observable actions* and define an auxiliary state variable y, called the *mode of the game*, as follows:

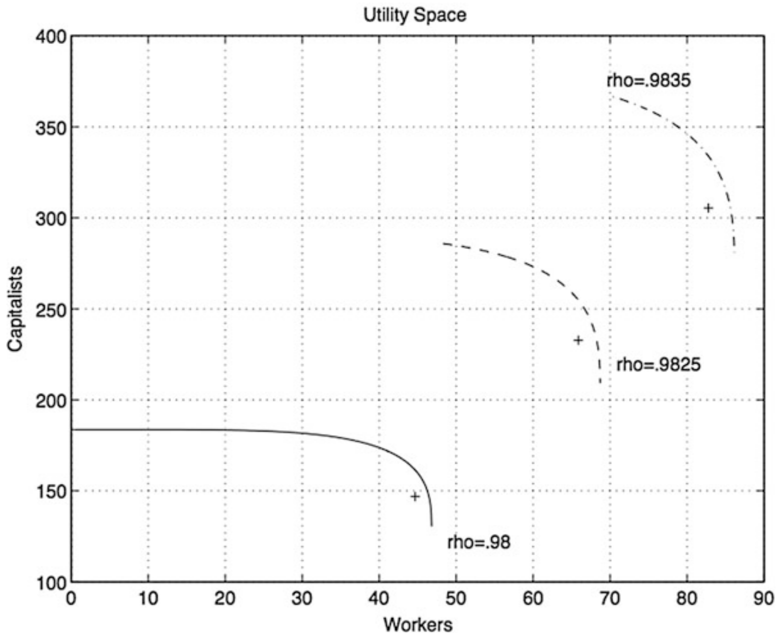


Fig. 4.9 Pareto-efficient utility frontiers

$$\begin{cases} y(0) &= 1 \\ y(t+1) &= \begin{cases} 1 & \text{if } \ y(t) = 1 \ \text{ and } \ c(t) = \underline{c}x(t) \text{ and } w(t) = \underline{w}x(t) \\ 0 & \text{otherwise.} \end{cases} \end{cases} \tag{4.133}$$

Once the *mode of the game* changes from 1 (cooperative) to 0 (noncooperative), the players remain in this mode indefinitely (i.e., there is no return to negotiations).

Define a new pair of strategies $(\delta_c^\alpha, \delta_w^\alpha)$ in the following way:

$$\begin{cases} \delta_c^\alpha(x(t), y(t)) = \begin{cases} \underline{c}x(t) & \text{if } \ y(t) = 1 \\ \overline{c}x(t) & \text{otherwise} \end{cases} \\ \delta_w^\alpha(x(t), y(t)) = \begin{cases} \underline{w}x(t) & \text{if } \ y(t) = 1 \\ \overline{w}x(t) & \text{otherwise.} \end{cases} \end{cases} \tag{4.134}$$

This definition implies that each player would use a Pareto-efficient strategy so long as his opponent does the same.

To argue that the payoffs obtained in Fig. 4.9 (and the strategies that generate them) are plausible, we need to show that (4.134) constitutes a subgame perfect equilibrium. We will prove this claim by demonstrating that $(\delta_c^\alpha, \delta_w^\alpha)$ is an equilibrium at every point (x, y).

It is easy to see that the pair is an equilibrium if $y(\tau) = 0$ for $\tau > 0$. Indeed, if one of the players has cheated before τ, the players would use feedback-Nash equilibrium strategies $\overline{c}k, \overline{w}k$, which are subgame perfect.

Moreover, $(\delta_c^\alpha, \delta_w^\alpha)$ will be an equilibrium for $(x, 1)$ if the maximum gain from "cheating," i.e., from breaching the agreement concerning consumption streams $\overline{c}k, \overline{w}k$, does not exceed the agreed-upon Pareto-efficient utility levels J_c^α, J_w^α for either player. To examine the range of values of the discount factor and the bargaining power that support such equilibria, we will solve the following system of inequalities:

$$\max_{c^+} \left\{Bx^\nu + D \cdot (c^+)^\nu + \beta V_c\big(ax - c^+ - \beta_w^\alpha(x)\big)\right\} < W_c^\alpha(x) \quad (4.135)$$

$$\max_{w^+} \left\{F \cdot (w^+)^\nu + \beta V_w\big(ax - \beta_c^\alpha(x) - w^+\big)\right\} < W_w^\alpha(x). \quad (4.136)$$

Here, c^+ and w^+ denote the best cheating policies, $V_c(\kappa), V_w(\kappa)$ is the pair of utilities resulting from the implementation of the punishment strategies $(\overline{c}\,\kappa, \overline{w}\,\kappa)$ from any state $\kappa > 0$, and $W_c^\alpha(\kappa), W_w^\alpha(\kappa)$ is the pair of utilities resulting from the Pareto-efficient strategies $(\underline{c}\,\kappa, \underline{w}\,\kappa)$.

The maxima of the left-hand sides of the above inequalities are achieved at

$$c^+ = \frac{(a - \underline{w}^\alpha)\, x}{1 + \left(\frac{D}{\beta \Gamma_c}\right)^{\frac{1}{\nu-1}}} \quad (4.137)$$

$$w^+ = \frac{(a - \underline{c}^\alpha)\, x}{1 + \left(\frac{F}{\beta \Gamma_w}\right)^{\frac{1}{\nu-1}}} \quad (4.138)$$

respectively.[31] After substituting c^+ and w^+ in the expressions, under the *max* operator in (4.135), (4.136) and simplifying the result, we obtain

$$\left(B + \frac{(a - \underline{w}^\alpha)^\nu D}{\left(1 + \left(\frac{D}{\beta \Gamma_c}\right)^{\frac{1}{\nu-1}}\right)^{\nu-1}}\right) x < W_c^\alpha(x) \quad (4.139)$$

$$\frac{(a - \underline{c}^\alpha)^\nu D}{\left(1 + \left(\frac{F}{\beta \Gamma_w}\right)^{\frac{1}{\nu-1}}\right)^{\nu-1}} x < W_w^\alpha(x). \quad (4.140)$$

The inequalities are illustrated graphically in Fig. 4.10 for $\beta = 0.9825$.

[31] Where all $\underline{c}^\alpha, \underline{w}^\alpha, \Gamma_c, \Gamma_w$ depend on β and α; see pages 203 and 206.

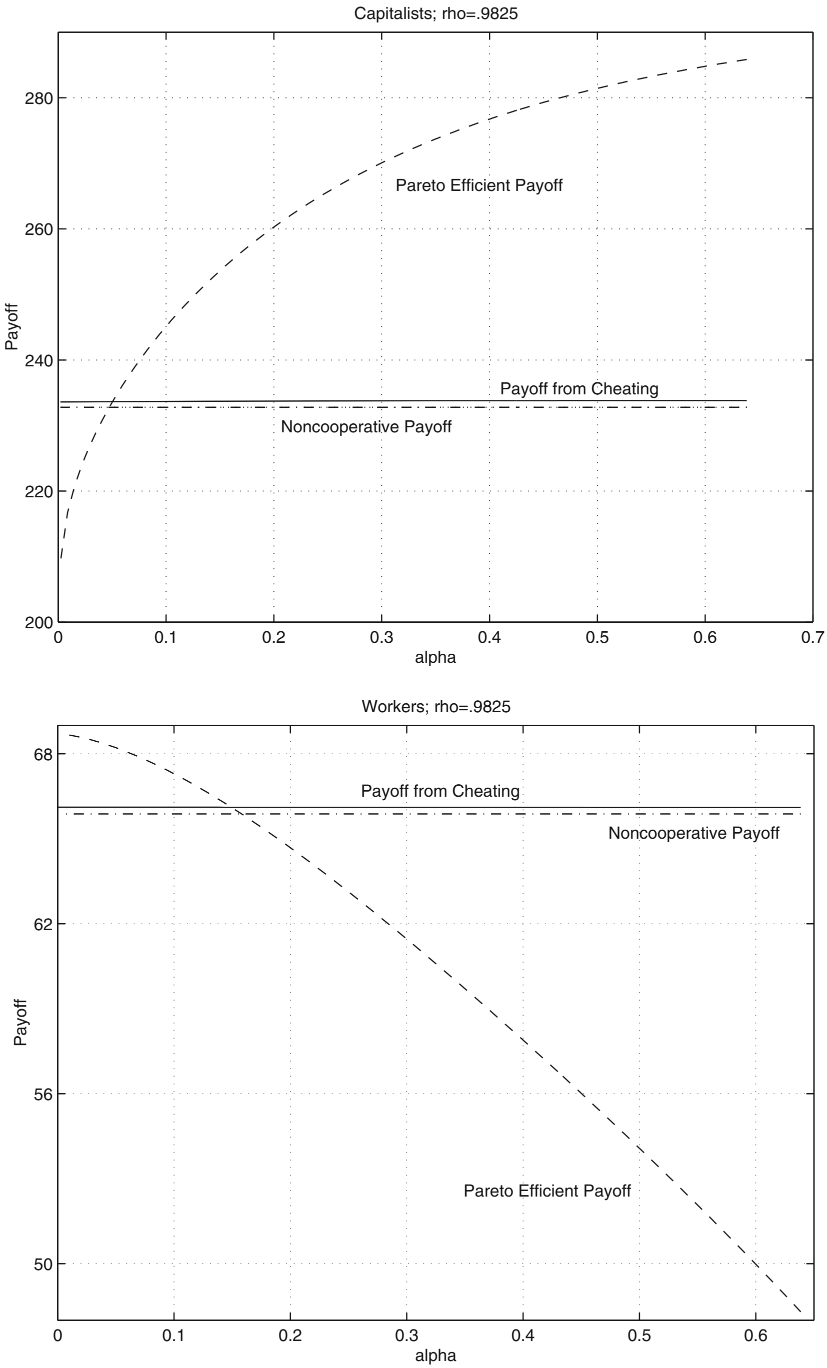


Fig. 4.10 Where is cheating non-profitable for $\beta = 0.9825$?

We observe that there is a nonempty interval of values of α, approximately from 0.05 (upper panel) to 0.15, where cheating is unprofitable for *both* players. Consequently, there exists a range of bargaining powers and a discount factor such that (4.134) is a subgame perfect equilibrium.

8.4.4 Final Remarks

We have obtained a *continuum of equilibrium growth rates for a particular set of fundamentals.* Moreover, we can infer the properties of this continuum. For example, Fig. 4.8 suggests that, for a given value of β, countries with higher α experience higher growth rates.

In other words, we have *engineered* a plausible dynamic game between trade unions and capitalists which yields a continuum of growth paths. Each path is Pareto efficient and arises in a feedback equilibrium. We believe that our model captures some of the reality underlying economic growth by acknowledging that capitalists have a passion for accumulation $(D > 0)$.

9 Conclusion

This chapter has provided a brief description of existing research on multistage games with state dynamics and illustrates the underlying theory with several examples. It highlighted the importance of information structure for the outcome of such games. We have considered various equilibrium solutions depending on what information is available to the players, in particular open-loop, closed-loop, and Markovian (or feedback) equilibria. We have also described the methods for computing these equilibria.

The chapter has distinguished between several types of settings. In some of them, all players choose their actions simultaneously, while in others only a subset of the players are able to affect the state at any given time. We have explored solutions to games in which there is a hierarchy of players (i.e., Stackelberg games). Finally, we have shown how the history of play can be used to enforce the cooperative solution in multistage games.

References

Barro RJ (1999) Ramsey meets Laibson in the neoclassical growth model. Q J Econ 114: 1125–1152

Başar T (1989) Time consistency and robustness of equilibria in noncooperative dynamic games. In: van der Ploeg F, de Zeeuw AJ (eds) Dynamic policy games in economics. North Holland, Amsterdam

Başar T, Olsder GJ (1982) Dynamic noncooperative game theory. Academic Press, New York

Başar T, Olsder GJ (1999) Dynamic noncooperative game theory. SIAM series in Classics in Applied Mathematics. SIAM, Philadelphia

Bellman R (1957) Dynamic programming. Princeton University Press, Princeton

Clark CW (1976) Mathematical bioeconomics. Wiley-Interscience, New York

DeFreitas G, Marshall A (1998) Labour surplus, worker rights and productivity growth: a comparative analysis of Asia and Latin America. Labour 12(3):515–539

Diamond P, Köszegi B (2003) Quasi-hyperbolic discounting and retirement. J Public Econ 87(9):1839–1872

Fahrig L (2002) Effect of habitat fragmentation on the extinction threshold: a synthesis*. Ecol Appl 12(2):346–353

Fan LT, Wang CS (1964) The discrete maximum principle. John Wiley and Sons, New York

Fudenberg D, Tirole J (1991) Game theory. The MIT Press, Cambridge/London

Gruber J, Koszegi B (2001) Is addiction rational? Theory and evidence. Q J Econ 116(4):1261

Halkin H (1974) Necessary conditions for optimal control problems with infinite horizons. Econom J Econom Soc 42:267–272

Haurie A, Krawczyk JB, Zaccour G (2012) Games and dynamic games. Business series, vol 1. World Scientific/Now Publishers, Singapore/Hackensack

Kaldor N (1961) Capital accumulation and economic growth. In: Lutz F, Hague DC (eds) Proceedings of a conference held by the international economics association. McMillan, London, pp 177–222

Kocherlakota NR (2001) Looking for evidence of time-inconsistent preferences in asset market data. Fed Reserve Bank Minneap Q Rev-Fed Reserve Bank Minneap 25(3):13

Krawczyk JB, Shimomura K (2003) Why countries with the same technology and preferences can have different growth rates. J Econ Dyn Control 27(10):1899–1916

Krawczyk JB, Tidball M (2006) A discrete-time dynamic game of seasonal water allocation. J Optim Theory Appl 128(2):411–429

Laibson DI (1996) Hyperbolic discount functions, undersaving, and savings policy. Technical report, National Bureau of Economic Research

Laibson D (1997) Golden eggs and hyperbolic discounting. Q J Econ 112(2):443–477

Levhari D, Mirman LJ (1980) The great fish war: an example using a dynamic Cournot-Nash solution. Bell J Econ 11(1):322–334

Long NV (1977) Optimal exploitation and replenishment of a natural resource. In: Pitchford J, Turnovsky S (eds) Applications of control theory in economic analysis. North-Holland, Amsterdam, pp 81–106

Luenberger DG (1969) Optimization by vector space methods. John Wiley & Sons, New York

Luenberger DG (1979) Introduction to dynamic systems: theory, models & applications. John Wiley & Sons, New York

Maskin E, Tirole J (1987) A theory of dynamic oligopoly, iii: Cournot competition. Eur Econ Rev 31(4):947–968

Michel P (1982) On the transversality condition in infinite horizon optimal problems. Econometrica 50:975–985

O'Donoghue T, Rabin M (1999) Doing it now or later. Am Econ Rev 89(1):103–124

Paserman MD (2008) Job search and hyperbolic discounting: structural estimation and policy evaluation. Econ J 118(531):1418–1452

Phelps ES, Pollak RA (1968) On second-best national saving and game-equilibrium growth. Rev Econ Stud 35(2):185–199

Pulliam HR (1988) Sources, sinks, and population regulation. Am Nat 652–661

Romer PM (1986) Increasing returns and long-run growth. J Polit Econ 94(5):1002–1037

Samuelson PA (1937) A note on measurement of utility. Rev Econ Stud 4(2):155–161

Selten R (1975) Rexaminition of the perfectness concept for equilibrium points in extensive games. Int J Game Theory 4(1):25–55

Simaan M, Cruz JB (1973) Additional aspects of the Stackelberg strategy in nonzero-sum games. J Optim Theory Appl 11:613–626

Spencer BJ, Brander JA (1983a) International R&D rivalry and industrial strategy. Working Paper 1192, National Bureau of Economic Research, http://www.nber.org/papers/w1192

Spencer BJ, Brander JA (1983b) Strategic commitment with R&D: the symmetric case. Bell J Econ 14(1):225–235

Strotz RH (1955) Myopia and inconsistency in dynamic utility maximization. Rev Econ Stud 23(3):165–180

Zero-Sum Stochastic Games

5

Anna Jaśkiewicz and Andrzej S. Nowak

Contents

Abstract

In this chapter, we describe a major part of the theory of zero-sum discrete-time stochastic games. We review all basic streams of research in this area, in the context of the existence of value and uniform value, algorithms, vector payoffs, incomplete information, and imperfect state observation. Also some models related to continuous-time games, e.g., games with short-stage duration, semi-Markov games, are mentioned. Moreover, a number of applications of stochastic

A. Jaśkiewicz (✉)
Faculty of Pure and Applied Mathematics, Wrocław University of Science and Technology, Wrocław, Poland
e-mail: anna.jaskiewicz@pwr.edu.pl

A. S. Nowak
Faculty of Mathematics, Computer Science and Econometrics, University of Zielona Góra, Zielona Góra, Poland
e-mail: a.nowak@wmie.uz.zgora.pl

T. Başar, G. Zaccour (eds.), *Handbook of Dynamic Game Theory*,
https://doi.org/10.1007/978-3-319-44374-4_8

games are pointed out. The provided reference list reveals a tremendous progress made in the field of zero-sum stochastic games since the seminal work of Shapley (Proc Nat Acad Sci USA 39:1095–1100, 1953).

Keywords
Zero-sum game · Stochastic game · Borel space · Unbounded payoffs · Incomplete information · Measurable strategy · Maxmin optimization · Limsup payoff · Approachability · Algorithms

1 Introduction

Stochastic games extend the model of strategic form games to situations in which the environment changes in time in response to the players' actions. They also extend the Markov decision model to competitive situations with more than one decision maker. The choices made by the players have two effects. First, together with the current state, the players' actions determine the immediate payoff that each player receives. Second, the current state and the players' actions have an influence on the chance of moving to a new state, where future payoffs will be received. Therefore, each player has to observe the current payoffs and take into account possible evolution of the state. This issue is also present in one-player sequential decision problems, but the presence of additional players who have their own goals adds complexity to the analysis of the situation. Stochastic games were introduced in a seminal paper of Shapley (1953). He considered zero-sum dynamic games with finite state and action spaces and a positive probability of termination. His model is often considered as a stochastic game with discounted payoffs. Gillette (1957) studied a similar model but with zero stop probability. These two papers inspired an enormous stream of research in dynamic game theory and Markov decision processes. There is a large variety of mathematical tools used in studying stochastic games. For example, the asymptotic theory of stochastic games is based on some algebraic methods such as semi-algebraic functions. On the other hand, the theory of stochastic games with general state spaces has a direct connection to the descriptive set theory. Furthermore, the algorithmic aspects of stochastic games yield interesting combinatorial problems. The other basic mathematical tools make use of martingale limit theory. There is also a known link between nonzero-sum stochastic games and the theory of fixed points in infinite-dimensional spaces. The principal goal of this chapter is to provide a comprehensive overview of the aforementioned aspects of zero-sum stochastic games.

To begin a literature review, let us mention that a basic and clear introduction to dynamic games is given in Başar and Olsder (1995) and Haurie et al. (2012). Mathematical programming problems occurring in algorithms for stochastic games with finite state and action spaces are broadly discussed in Filar and Vrieze (1997). Some studies of stochastic games by the methods developed in gambling theory with many informative examples are described in Maitra and Sudderth (1996). An advanced material on repeated and stochastic games is presented in Sorin (2002) and Mertens et al. (2015). The two edited volumes by Raghavan et al. (1991) and Neyman and Sorin (2003) contain a survey of a large part of the area of stochastic games developed for almost fifty years since Shapley's seminal work. This chapter and the chapter of Jaśkiewicz and Nowak (2018) include a very broad overview of state-of-the-art results on stochastic games. Moreover, the surveys given by Mertens (2002), Vieille (2002), Solan (2009), Krishnamurthy and Parthasarathy (2011), Solan and Vieille (2015), and Laraki and Sorin (2015) constitute relevant complementary material.

There is a great deal of applications of stochastic games in science and engineering. Here, we only mention the ones concerning zero-sum games. For instance, Altman and Hordijk (1995) applied stochastic games to queueing models. On the other hand, wireless communication networks were examined in terms of stochastic games by Altman et al. (2005). For use of stochastic games in models that arise in operations research, the reader is referred to Charnes and Schroeder (1967), Winston (1978), Filar (1985), or Patek and Bertsekas (1999). There is also a growing literature on applications of zero-sum stochastic games in theoretical computer science (see, for instance, Condon (1992), de Alfaro et al. (2007) and Kehagias et al. (2013) and references cited therein). Applications of zero-sum stochastic games to economic growth models and robust Markov decision processes are described in Sect. 3, which is mainly based on the paper of Jaśkiewicz and Nowak (2011). The class of possible applications of nonzero-sum stochastic games is larger than in the zero-sum case. They are discussed in our second survey in this handbook.

The chapter is organized as follows: In Sect. 2 we describe some basic material needed for a study of stochastic games with general state spaces. It incorporates auxiliary results on set-valued mappings (correspondences), their measurable selections, and the measurability of the value of a parameterized zero-sum game. This part naturally is redundant in a study of stochastic games with discrete state and action spaces. Sect. 3 is devoted to a general maxmin decision problem in discrete-time and Borel state space. The main motivation is to show its applications to stochastic economic growth models and some robust decision problems in macroeconomics. Therefore, the utility (payoff) function in illustrative examples is unbounded and the transition probability function is weakly continuous. In Sect. 4 we consider standard discounted and positive Markov games with Borel state spaces and simultaneous moves of the players. Sect. 5 is devoted to semi-Markov games with Borel state space and weakly continuous transition probabilities satisfying some stochastic stability assumptions. In the limit-average payoff case, two criteria are compared, the time average and ratio average payoff criterion, and a question of path optimality is discussed. Furthermore, stochastic games with a general Borel payoff function on the spaces of infinite plays are examined in Sect. 6. This part includes results on games with limsup payoffs and limit-average payoffs as special cases. In Sect. 7 we present some basic results from the asymptotic theory of stochastic games, mainly with finite state space, the notion of uniform value. This part of the theory exhibits nontrivial algebraic aspects. Some algorithms for solving zero-sum stochastic games of different types are described in Sect. 8. In Sect. 9 we provide an overview of zero-sum stochastic games with incomplete information and imperfect monitoring. This is a vast subarea of stochastic games, and therefore, we deal only with selected cases of recent contributions. Stochastic games with vector payoffs and Blackwell's approachability concept, on the other hand, are discussed briefly in Sect. 10. Finally, Sect. 11 gives a short overview of stochastic Markov games in continuous time. We mainly focus on Markov games with short-stage duration. This theory is based on an asymptotic analysis of discrete-time games when the stage duration tends to zero.

2 Preliminaries

Let $\mathbb{R}$ be the set of all real numbers, $\underline{\mathbb{R}} = \mathbb{R} \cup \{-\infty\}$ and $\mathbb{N} = \{1, 2, \ldots\}$. By a *Borel space* X we mean a nonempty Borel subset of a complete separable metric space endowed with the relative topology and the Borel σ-algebra $\mathcal{B}(X)$. We denote by $\Pr(X)$ the set of all Borel probability measures on X. Let $\mathcal{B}_\mu(X)$ be the completion of $\mathcal{B}(X)$ with respect to some $\mu \in \Pr(X)$. Then $\mathcal{U}(X) = \cap_{\mu \in \Pr(X)} \mathcal{B}_\mu(X)$ is the σ-algebra of all universally measurable subsets of X. There are a couple of ways to define analytic sets in X (see Chap. 12 in Aliprantis and Border 2006 or Chap. 7 in Bertsekas and Shreve 1996). One can say that $C \subset X$ is an *analytic set* if and only if there is a Borel set $D \subset X \times X$ whose projection on X is C. If X is uncountable, then there exist analytic sets in X which are not Borel (see Example 12.33 in Aliprantis and Border 2006). Every analytic set $C \subset X$ belongs to $\mathcal{U}(X)$. A function $\psi : X \to \underline{\mathbb{R}}$ is called *upper semianalytic* (*lower semianalytic*) if for any $c \in \mathbb{R}$ the set $\{x \in X : \psi(x) \geq c\}$ ($\{x \in X : \psi(x) \leq c\}$) is analytic. It is known that ψ is both upper and lower semianalytic if and only if ψ is Borel measurable. Let Y be also a Borel space. A mapping $\phi : X \to Y$ is *universally measurable* if $\phi^{-1}(C) \in \mathcal{U}(X)$ for each $C \in \mathcal{B}(Y)$.

A set-valued mapping $x \to \Phi(x) \subset Y$ (also called a correspondence from X to Y) is *upper semicontinuous* (*lower semicontinuous*) if the set $\Phi^{-1}(C) := \{x \in X : \Phi(x) \cap C \neq \emptyset\}$ is closed (open) for each closed (open) set $C \subset Y$. Φ is *continuous* if it is both lower and upper semicontinuous. Φ is *weakly* or *lower measurable* if $\Phi^{-1}(C) \in \mathcal{B}(X)$ for each open set $C \subset Y$. Assume that $\Phi(x) \neq \emptyset$ for every $x \in X$. If Φ is compact valued and upper semicontinuous, then by Theorem 1 in Brown and Purves (1973), Φ admits a measurable selector, that is, there exists a Borel measurable mapping $g : X \to Y$ such that $g(x) \in \Phi(x)$ for each $x \in X$. Moreover, the same holds if Φ is weakly measurable and has complete values $\Phi(x)$ for all $x \in X$ (see Kuratowski and Ryll-Nardzewski 1965). Assume that $D \subset X \times Y$ is a Borel set such that $D(x) := \{y \in Y : (x, y) \in D\}$ is nonempty and compact for each $x \in X$. If C is an open set in Y, then $D^{-1}(C) := \{x \in X : D(x) \cap C \neq \emptyset\}$ is the projection on X of the Borel set $D_0 = (X \times C) \cap D$ and $D_0(x) = \{y \in Y : (x, y) \in D_0\}$ is σ-compact for any $x \in X$. By Theorem 1 in Brown and Purves (1973), $D^{-1}(C) \in \mathcal{B}(X)$. For a broad discussion of semicontinuous or measurable correspondences, the reader is referred to Himmelberg (1975), Klein and Thompson (1984) or Aliprantis and Border (2006). For any Borel space Y, let $C(Y)$ be the space of all bounded continuous real-valued functions on Y. Assume that $\Pr(Y)$ is endowed with the weak topology and the Borel σ-algebra $\mathcal{B}(\Pr(Y))$ (see Bertsekas and Shreve 1996; Billingsley 1968 or Parthasarathy 1967). The σ-algebra $\mathcal{B}(\Pr(Y))$ of all Borel subsets of $\Pr(Y)$ coincides with the smallest σ-algebra on $\Pr(Y)$ for which all the mappings $p \to p(D)$ from $\Pr(Y)$ to $[0, 1]$ are measurable for each $D \in \mathcal{B}(Y)$ (see Proposition 7.25 in Bertsekas and Shreve 1996). Recall that a sequence $(p_n)_{n\in\mathbb{N}}$ *converges weakly* to some $p \in \Pr(Y)$ if and only if for any $\phi \in C(Y)$,

$$\int_Y \phi(y)p_n(dy) \to \int_Y \phi(y)p(dy) \quad \text{as} \quad n \to \infty.$$

If Y is a Borel space, then $\Pr(Y)$ is a Borel space too, and if Y is compact, so is $\Pr(Y)$ (see Corollary 7.25.1 and Proposition 7.22 in Bertsekas and Shreve 1996).

Consider the correspondence $x \to \Psi(x) := \Pr(\Phi(x)) \subset \Pr(Y)$. The following result from Himmelberg and Van Vleck (1975) is useful in studying stochastic games.

Proposition 1. *If Φ is upper (lower) semicontinuous and compact valued, then so is Ψ.*

A *transition probability* or a *stochastic kernel* from X to Y is a function $\varphi : \mathcal{B}(Y) \times X \to [0, 1]$ such that $\varphi(D|\cdot)$ is a Borel measurable function on X for every $D \in \mathcal{B}(Y)$ and $\varphi(\cdot|x) \in \Pr(Y)$ for each $x \in X$. It is well known that every Borel measurable mapping $f : X \to \Pr(Y)$ may be regarded as a transition probability φ from X to Y. Namely, $\varphi(D|x) = f(x)(D)$, $D \in \mathcal{B}(Y)$, $x \in X$ (see Proposition 7.26 in Bertsekas and Shreve 1996). We shall write $f(dy|x)$ instead of $f(x)(dy)$. Clearly, any Borel measurable mapping $f : X \to Y$ is a special transition probability φ from X to Y such that for each $x \in X$, $\varphi(\cdot|x)$ is the Dirac measure concentrated at the point $f(x)$. Similarly, universally measurable transition probabilities are defined, when $\mathcal{B}(X)$ is replaced by $\mathcal{U}(X)$.

In studying zero-sum stochastic games with Borel state spaces, we must use in the proofs some results on minmax measurable selections in *parameterized games*. Let X, A, and B be Borel spaces. Assume that $K_A \in \mathcal{B}(X \times A)$ and $K_B \in \mathcal{B}(X \times B)$ and suppose that the sets $A(x) := \{a \in A : (x, a) \in A\}$ and $B(x) := \{b \in B : (x, b) \in B\}$ are nonempty for all $x \in X$. Let $K := \{(x, a, b) : x \in X, a \in A(x), b \in B(x)\}$. Then K is a Borel subset of $X \times A \times B$. Let $r : K \to \mathbb{R}$ be a Borel measurable payoff function in a zero-sum game parameterized by $x \in X$. If players 1 and 2 choose mixed strategies $\mu \in \Pr(A(x))$ and $\nu \in \Pr(B(x))$, respectively, then the expected payoff to player 1 (cost to player 2) depends on $x \in X$ and is of the form

$$R(x, \mu, \nu) := \int_{A(x)} \int_{B(x)} r(x, a, b)\nu(db)\mu(da)$$

provided that the double integral is well defined. Assuming this and that $B(x)$ is compact for each $x \in X$ and $r(x,a,\cdot)$ is lower semicontinuous on $B(x)$ for each $(x,a) \in K_A$, we conclude from the minmax theorem of Fan (1953) that the game has a value, that is, the following equality holds

$$v^*(x) := \min_{\nu\in\Pr(B(x))} \sup_{\mu\in\Pr(A(x))} R(x,\mu,\nu) = \sup_{\mu\in\Pr(A(x))} \min_{\nu\in Pr(B(x))} R(x,\mu,\nu), \quad x \in X.$$

A universally (Borel) measurable strategy for player 1 is a universally (Borel) measurable transition probability f from X to A such that $f(A(x)|x) = 1$ for all $x \in X$. By the Jankov-von Neumann theorem (see Theorem 18.22 in Aliprantis and Border 2006), there exists a universally measurable function $\varphi : X \to A$ such that $\varphi(x) \in A(x)$ for all $x \in X$. Thus, the set of universally measurable strategies for player 1 is nonempty. Universally (Borel) measurable strategies for player 2 are defined similarly. A strategy g^* is *optimal* for player 2 if

$$v^*(x) = \sup_{\mu\in\Pr(A(x))} \int_{A(x)} \int_{B(x)} r(x,a,b) g^*(db|x)\mu(da) \quad \text{for all} \quad x \in X.$$

Let $\varepsilon \geq 0$. A strategy f^* is ε-*optimal* for player 1 if

$$v^*(x) \leq \inf_{\nu\in\Pr(B(x))} \int_{A(x)} \int_{B(x)} r(x,a,b)\nu(db) f^*(da|x) + \varepsilon \quad \text{for all} \quad x \in X.$$

A 0-optimal strategy is called optimal.

The following result follows from Nowak (1985b). For a much simpler proof, see Nowak (2010).

Proposition 2. *Under the above assumptions the value function v^* is upper semianalytic. Player 2 has a universally measurable optimal strategy and, for any $\varepsilon > 0$, player 1 has a universally measurable ε-optimal strategy. If, in addition, we assume that $A(x)$ is compact for each $x \in X$ and $r(x,\cdot,b)$ is upper semicontinuous for each $(x,b) \in K_B$, then v^* is Borel measurable and both players have Borel measurable optimal strategies.*

As a corollary to Theorem 5.1 in Nowak (1986), we can state the following result.

Proposition 3. *Assume that $x \to A(x)$ is lower semicontinuous and has complete values in A and $x \to B(x)$ is upper semicontinuous and compact valued. If $r : K \to \mathbb{R}$ is lower semicontinuous on K, then v^* is lower semicontinuous, player 2 has a Borel measurable optimal strategy, and for any $\varepsilon > 0$, player 1 has a Borel measurable ε-optimal strategy.*

The lower semicontinuity of v^* in Proposition 3 is a corollary to the maximum theorem of Berge (1963). In some games or minmax control models, one can consider the minmax value

$$\underline{v}^*(x) := \inf_{\nu \in \Pr(B(x))} \sup_{\mu \in \Pr(A(x))} R(x, \mu, \nu), \quad x \in X,$$

if the mixed strategies are used, or

$$\underline{w}^*(x) := \inf_{b \in B(x)} \sup_{a \in A(x)} r(x, a, b), \quad x \in X,$$

if the attention is restricted to pure strategies. If the assumption on semicontinuity of the function r is dropped, then the measurability of $\underline{v}^*$ or $\underline{w}^*$ is connected with the measurability of projections of coanalytic sets. This issue leads to some considerations in the classical descriptive set theory. A comprehensive study of the measurability of upper or lower value of a game with Borel payoff function r is given in Prikry and Sudderth (2016).

3 Robust Markov Decision Processes

A discounted *maxmin Markov decision process* is defined by the objects X, A, B, K_A, K, u, q, and β, where:

- X is a Borel *state space*;
- A is the *action space* of the *controller* (player 1) and B is the action space of the *opponent* (player 2). It is assumed that A and B are Borel spaces;
- $K_A \in \mathcal{B}(X \times A)$ is the *constraint set* for the *controller*. It is assumed that

  $$A(x) := \{a \in A : (x, a) \in A\} \neq \emptyset$$

  for each $x \in X$. This is the *set of admissible actions* of the *controller* in the state $x \in X$;
- $K \in \mathcal{B}(X \times A \times B)$ is the *constraint set* for the *opponent*. It is assumed that

  $$B(x, a) := \{b \in B : (x, a, b) \in B\} \neq \emptyset$$

  for each $(x, a) \in K_A$. This is the *set of admissible actions* of the *opponent* for $(x, a) \in K_A$;
- $u : K \to \mathbb{R}$ is a Borel measurable *stage payoff function*;
- q is a transition probability from K to X, called the *law of motion* among states. If x_n is a state at the beginning of period n of the process and actions $a_n \in A(x_n)$ and $b_n \in B(x_n, a_n)$ are selected by the players, then $q(\cdot|x_n, a_n, b_n)$ is the probability distribution of the next state x_{n+1};

- $\beta \in (0,1)$ is the *discount factor*.

We make the following assumptions on the admissible action sets.

(C1) For any $x \in X$, $A(x)$ is compact and the set-valued mapping $x \to A(x)$ is upper semicontinuous.
(C2) The set-valued mapping $(x,a) \to B(x,a)$ is lower semicontinuous.
(C3) There exists a Borel measurable mapping $g : K_A \to B$ such that $g(x,a) \in B(x,a)$ for all $(x,a) \in K_A$.

Remark 1. From Sect. 2, it follows that condition (C3) holds if $B(x,a)$ is σ-compact for each $(x,a) \in K_A$ (see Brown and Purves 1973) or if B is a complete separable metric space and each set $B(x,a)$ is closed (see Kuratowski and Ryll-Nardzewski 1965).

Let $H_1 := X$, $H_n := K^n \times X$ for $n \geq 2$. Put $H_1^* := K_A$ and $H_n^* := K^n \times K_A$ if $n \geq 2$. Generic elements of H_n and H_n^* are *histories* of the process, and they are of the form $h_1 = x_1$, $h_1^* = (x_1, a_1)$ and for each $n \geq 2$, $h_n = (x_1, a_1, b_1, \ldots .x_{n-1}, a_{n-1}, b_{n-1}, x_n)$, $h_n^* = (h_n, a_n)$.

A *strategy* for the controller is a sequence $\pi = (\pi_n)_{n \in \mathbb{N}}$ of stochastic kernels π_n from H_n to A such that $\pi_n(A(x_n)|h_n) = 1$ for each $h_n \in H_n$. The class of all strategies for the controller will be denoted by Π. A *strategy* for the opponent is a sequence $\gamma = (\gamma_n)_{n \in \mathbb{N}}$ of stochastic kernels γ_n from H_n^* to B such that $\gamma_n(B(x_n, a_n)|h_n^*) = 1$ for all $h_n^* \in H_n^*$. The class of all strategies for the opponent will be denoted by Γ^*. Let F be the set of Borel measurable mappings f from X to A such that $f(x) \in A(x)$ for each $x \in X$. A *deterministic stationary strategy* for the controller is a sequence $\pi = (f_n)_{n \in \mathbb{N}}$ where $f_n = f$ for all $n \in \mathbb{N}$ and some $f \in F$. Such a strategy can obviously be identified with the mapping $f \in F$. Let

$$u^+(x,a,b) := \max\{u(x,a,b), 0\} \quad \text{and}$$
$$u^-(x,a,b) := \min\{u(x,a,b), 0\}, \quad (x,a,b) \in K.$$

For each initial state $x_1 = x$ and any strategies $\pi \in \Pi$ and $\gamma \in \Gamma^*$, define

$$J_\beta^+(x,\pi,\gamma) = E_x^{\pi\gamma}\left(\sum_{n=1}^{\infty} \beta^{n-1} u^+(x_n, a_n, b_n)\right), \tag{5.1}$$

$$J_\beta^-(x,\pi,\gamma) = E_x^{\pi\gamma}\left(\sum_{n=1}^{\infty} \beta^{n-1} u^-(x_n, a_n, b_n)\right). \tag{5.2}$$

Here, $E_x^{\pi\gamma}$ denotes the expectation operator corresponding to the unique conditional probability measure $P_x^{\pi\gamma}$ defined on the space of histories, starting at state x, and endowed with the product σ-algebra, which is induced by strategies π, γ

and the transition probability q according to the Ionescu-Tulcea Theorem (see Proposition 7.45 in Bertsekas and Shreve 1996 or Proposition V.1.1 in Neveu 1965). In the sequel, we give conditions under which $J^+_\beta(x,\pi,\gamma) < \infty$ for any $x \in X$, $\pi \in \Pi$, $\gamma \in \Gamma^*$. They enable us to define the *expected discounted payoff* over an infinite time horizon as follows:

$$J_\beta(x,\pi,\gamma) = E_x^{\pi\gamma}\left(\sum_{n=1}^{\infty}\beta^{n-1}u(x_n,a_n,b_n)\right). \tag{5.3}$$

Then, for every $x \in X$, $\pi \in \Pi$, $\gamma \in \Gamma^*$ we have that $J_\beta(x,\pi,\gamma) \in \underline{\mathbb{R}}$ and

$$J_\beta(x,\pi,\gamma) = J^+_\beta(x,\pi,\gamma) + J^-_\beta(x,\pi,\gamma) = \sum_{n=1}^{\infty}\beta^{n-1}E_x^{\pi\gamma}u(x_n,a_n,b_n).$$

Let

$$v_\beta(x) := \sup_{\pi\in\Pi}\inf_{\gamma\in\Gamma^*} J_\beta(x,\pi,\gamma), \quad x \in X.$$

This is the *maxmin or lower value* of the game starting at the state $x \in X$. A strategy $\pi^* \in \Pi$ is called *optimal* for the controller if $\inf_{\gamma\in\Gamma^*} J_\beta(x,\pi^*,\gamma) = v_\beta(x)$ for every $x \in X$.

It is worth mentioning that if u is unbounded, then an optimal strategy π^* need not exist even if $0 \le v_\beta(x) < \infty$ for every $x \in X$ and the available action sets $A(x)$ and $B(x)$ are finite (see Example 1 in Jaśkiewicz and Nowak 2011).

The maxmin control problems with Borel state spaces have been already considered by González-Trejo et al. (2003), Hansen and Sargent (2008), Iyengar (2005), and Küenle (1986) and are referred to as *games against nature* or *robust dynamic programming (Markov decision) models*. The idea of using maxmin decision rules was introduced in statistics (see Blackwell and Girshick 1954). It is also used in economics (see, e.g., the variational preferences in Maccheroni et al. 2006).

3.1 One-Sided Weighted Norm Approach

We now describe our regularity assumptions imposed on the payoff and transition probability functions.

(W1) The payoff function $u : K \to \underline{\mathbb{R}}$ is upper semicontinuous.
(W2) For any $\phi \in C(X)$ the function

$$(x,a,b) \to \int_X \phi(y)q(dy|x,a,b)$$

is continuous.

(M1) There exist a continuous function $\omega : X \to [1, \infty)$ and a constant $\alpha > 0$ such that

$$\sup_{(x,a,b)\in K} \frac{\int_X \omega(y)q(dy|x,a,b)}{\omega(x)} \le \alpha \quad \text{and} \quad \beta\alpha < 1. \tag{5.4}$$

Moreover, the function $(x, a, b) \to \int_X \omega(y)q(dy|x, a, b)$ is continuous.

(M2) There exists a constant $d > 0$ such that

$$\sup_{a\in A(x)} \sup_{b\in B(x,a)} u^+(x, a, b) \le d\omega(x)$$

for all $x \in X$.

Note that under conditions (M1) and (M2), the discounted payoff function is well defined, since

$$0 \le E_x^{\pi\gamma} \left(\sum_{n=1}^{\infty} \beta^{n-1} u^+(x_n, a_n, b_n) \right) \le d \sum_{n=1}^{\infty} \beta^{n-1}\alpha^{n-1}\omega(x) < \infty.$$

Remark 2. Assumption (W2) states that transition probabilities are weakly continuous. It is worth emphasizing that this property, in contrast to the setwise continuous transitions, is satisfied in a number of models arising in operations research, economics, etc. Indeed, Feinberg and Lewis (2005) studied the typical inventory model:

$$x_{n+1} = x_n + a_n - \xi_{n+1}, \quad n \in \mathbb{N},$$

where x_n is the inventory at the end of period n, a_n is the decision on how much should be ordered, and ξ_n is the demand during period n and each ξ_n has the same distribution as the random variable ξ. Assume that $X = \mathbb{R}$, $A = \mathbb{R}_+$. Let $q(\cdot|x, a)$ be the transition law for this problem. In view of Lebesgue's dominated convergence theorem, it is clear that q is weakly continuous. On the other hand, recall that the setwise continuity means that $q(D|x, a^k) \to q(D|x, a^0)$ as $a^k \to a^0$ for any $D \in \mathcal{B}(X)$. Suppose that the demand is deterministic $d = 1$, $a^k = a + 1/k$ and $D = (-\infty, x + a - 1]$. Then, $q(D|x, a) = 1$, but $q(D|x, a^k) = 0$.

For any function $\phi : X \to \mathbb{R}$, define the ω-norm as follows:

$$\|\phi\|_\omega = \sup_{x\in X} \frac{|\phi(x)|}{\omega(x)}, \tag{5.5}$$

provided that it is finite. Let $U_\omega(X)$ be the space of all upper semicontinuous functions endowed with the metric induced by the ω-norm. By $\underline{U}_\omega(X)$ we denote the set of all upper semicontinuous functions $\phi : X \to \underline{\mathbb{R}}$ such that $\phi^+ \in U_\omega(X)$.

Define $u_k := \max\{u, -k\}$, $k \in \mathbb{N}$. For any $\phi \in \underline{U}_\omega(X)$, $(x, a, b) \in K$, and $k \in \mathbb{N}$, let

$$L_{\beta,k}\phi(x,a,b) = u_k(x,a,b) + \beta \int_X \phi(y)q(dy|x,a,b)$$

and

$$L_\beta\phi(x,a,b) = u(x,a,b) + \beta \int_X \phi(y)q(dy|x,a,b).$$

The maximum theorem of Berge (1963) (see also Proposition 10.2 in Schäl 1975) implies the following auxiliary result.

Lemma 1. *Assume* (C1)–(C3), (W1)–(W2), *and* (M1)–(M2). *Then for any* $\phi \in \underline{U}_\omega(X)$, *the functions*

$$\inf_{b\in B(x,a)} L_{\beta,k}\phi(x,a,b) \quad and \quad \max_{a\in A(x)} \inf_{b\in B(x,a)} L_{\beta,k}\phi(x,a,b)$$

are upper semicontinuous on K_A *and* X, *respectively. Similar properties hold if* $L_{\beta,k}\phi(x,a,b)$ *is replaced by* $L_\beta\phi(x,a,b)$.

For any $x \in X$, define

$$T_{\beta,k}\phi(x) = \max_{a\in A(x)} \inf_{b\in B(x,a)} L_{\beta,k}\phi(x,a,b) \quad \text{and}$$
$$T_\beta\phi(x) = \max_{a\in A(x)} \inf_{b\in B(x,a)} L_\beta\phi(x,a,b). \tag{5.6}$$

By Lemma 1, the operators $T_{\beta,k}$ and T_β are well defined. Additionally, note that

$$T_\beta\phi(x) = \max_{a\in A(x)} \inf_{\rho\in \Pr(B(x,a))} \int_{B(x,a)} L_\beta\phi(x,a,b)\rho(db).$$

We can now state the main result in Jaśkiewicz and Nowak (2011).

Theorem 1. *Assume* (C1)–(C3), (W1)–(W2), *and* (M1)–(M2). *Then* $v_\beta \in \underline{U}_\omega(X)$, $T_\beta v_\beta = v_\beta$ *and there exists a stationary strategy* $f^* \in F$ *such that*

$$v_\beta(x) = \inf_{b\in B(x,a)} L_\beta v_\beta(x, f^*(x), b)$$

for $x \in X$. *Moreover,*

$$v_\beta(x) = \inf_{\gamma\in\Gamma^*} J_\beta(x, f^*, \gamma) = \sup_{\pi\in\Pi} \inf_{\gamma\in\Gamma^*} J_\beta(x,\pi,\gamma)$$

for all $x \in X$, *so* f^* *is an optimal stationary strategy for the controller.*

The proof of Theorem 1 consists of two steps. First, we deal with truncated models, in which the payoff function u is replaced by u_k. Then, making use of the fixed point argument, we obtain an upper semicontinuous solution to the Bellman equation, say $v_{\beta,k}$. Next, we observe that the sequence $(v_{\beta,k})_{k\in\mathbb{N}}$ is nonincreasing. Letting $k \to \infty$ and making use of Lemma 1, we arrive at the conclusion.

Remark 3. The weighted supremum norm approach in Markov decision processes was proposed by Wessels (1977) and further developed, e.g., by Hernández-Lerma and Lasserre (1999). This method has been also adopted to zero-sum stochastic games (see Couwenbergh 1980; González-Trejo et al. 2003; Jaśkiewicz 2009, 2010; Jaśkiewicz and Nowak 2006, 2011; Küenle 2007 and references cited therein). The common feature of the aforementioned works is the fact that the authors use the weighted norm condition instead of assumption (M2). More precisely, in our notation it means that the following holds

$$\sup_{a\in A(x)} \sup_{b\in B(x,a)} |u(x,a,b)| \le d\omega(x), \quad x \in X \tag{5.7}$$

for some constant $d > 0$. This assumption, however, excludes many examples studied in economics where the utility function u equals $-\infty$ in some states. Moreover, inequality in (M1) and (5.7) often enforces additional constraints on the discount coefficient β in comparison with (M1) and (M2) (see Example 6 in Jaśkiewicz and Nowak 2011).

Observe that if the payoff function u accepts only negative values, then assumption (M2) is redundant. Thus, the problem comes down to the negative programming, which was solved by Strauch (1966) in the case of one-player game (Markov decision process).

3.1.1 Models with Unknown Disturbance Distributions

Consider the control system in which

$$x_{n+1} = \Psi(x_n, a_n, \xi_n), \quad n \in \mathbb{N}.$$

It is assumed that $(\xi_n)_{n\in\mathbb{N}}$ is a sequence of independent random variables with values in a Borel space S having unknown probability distributions that can change from period to period. The set B of all possible distributions is assumed to be a nonempty Borel subset of the space $\Pr(S)$ endowed with the weak topology. The mapping $\Psi : K_A \times S \to X$ is assumed to be *continuous*. Let u_0 be an *upper semicontinuous utility function* defined on $K_A \times S$ such that $u_0^+(x,a,s) \le d\omega(x)$ for some constant $d > 0$ and all $(x,a) \in K_A$, $s \in S$.

We can formulate a maxmin control model in the following way:

(a) $B(x,a) = B \subset \Pr(S)$ for each $(x,a) \in K_A$, $K = K_A \times B$;
(b) $u(x,a,b) = \int_S u_0(x,a,s)b(ds)$, $(x,a,b) \in K$;
(c) for any Borel set $D \subset X$, $q(D|x,a,b) = \int_X 1_D(\Psi(x,a,s))b(ds)$, $(x,a,b) \in K$.

Then for any bounded continuous function $\phi : X \to \mathbb{R}$, we have that

$$\int_X \phi(y)q(dy|x,a,b) = \int_X \phi(\Psi(x,a,s))b(ds). \tag{5.8}$$

From Proposition 7.30 in Bertsekas and Shreve (1996) or Lemma 5.3 in Nowak (1986) and (5.8), it follows that q is weakly continuous. Moreover, by virtue of Proposition 7.31 in Bertsekas and Shreve (1996), it is easily seen that u is upper semicontinuous on K.

The following result can be viewed as a corollary to Theorem 1.

Proposition 4. *Let Ψ and u_0 satisfy the above assumptions. If* (M1) *holds, then the controller has an optimal strategy.*

Proposition 4 is a counterpart of the results obtained in Sect. 6 of González-Trejo et al. (2003) for discounted models (see Propositions 6.1, 6.2, 6.3 and their consequences in González-Trejo et al. (2003)). However, our assumptions imposed on the primitive data are weaker than the ones used by González-Trejo et al. (2003). They are satisfied for a pretty large number of systems, in which the disturbances comprise "random noises" that are difficult to observe and often caused by external factors influencing the dynamics. Below we give certain examples which stem from economic growth theory and related topics. Mainly, they are inspired by models studied in Stokey et al. (1989), Bhattacharya and Majumdar (2007), and Hansen and Sargent (2008).

Example 1 (A growth model with multiplicative shocks). Let $X = [0, \infty)$ be the set of all possible capital stocks. If x_n is a capital stock at the beginning of period n, then the level of satisfaction of consumption of $a_n \in A(x_n) = [0, x_n]$ in this period is a_n^σ. Here $\sigma \in (0, 1]$ is a fixed parameter. The evolution of the state process is described by the following equation:

$$x_{n+1} = (x_n - a_n)^\theta \xi_n, \quad n \in \mathbb{N},$$

where $\theta \in (0, 1)$ is some constant and ξ_n is a random shock in period n. Assume that each ξ_n follows a probability distribution $b \in B$ for some Borel set $B \subset \Pr([0, \infty))$. We assume that b is unknown.

Consider the maxmin control model, where $X = [0, \infty)$, $A(x) = [0, x]$, $B(x, a) = B$, and $u(x, a, b) = a^\sigma$ for $(x, a, b) \in K$. Then, the transition probability q is of the form

$$q(D|x,a,b) = \int_0^\infty 1_D((x-a)^\theta s)b(ds),$$

where $D \in \mathcal{B}(X)$. If $\phi \in C(X)$, then the integral

$$\int_X \phi(y)q(dy|x,a,b) = \int_0^\infty \phi((x-a)^\theta s)b(ds)$$

is continuous at $(x, a, b) \in K$. We further assume that

$$\bar{s} = \sup_{b \in B} \int_0^\infty s b(ds) < \infty.$$

Define now

$$\omega(x) = (r + x)^\sigma, \quad x \in X, \tag{5.9}$$

where $r \geq 1$ is a constant. Clearly, $u^+(x, a, b) = a^\sigma \leq \omega(x)$ for any $(x, a, b) \in K$. Hence, condition (M2) is satisfied. Moreover, by Jensen's inequality we obtain

$$\int_X \omega(y) q(dy|x, a, b) = \int_0^\infty (r + (x - a)^\theta s)^\sigma b(ds) \leq (r + x^\theta \bar{s})^\sigma.$$

Thus,

$$\frac{\int_X \omega(y) q(dy|x, a, b)}{\omega(x)} \leq \eta^\sigma(x), \quad \text{where} \quad \eta(x) := \frac{r + \bar{s} x^\theta}{r + x}, \quad x \in X. \tag{5.10}$$

If $x \geq \bar{x} := \bar{s}^{1/(1-\theta)}$, then $\eta(x) \leq 1$, and consequently, $\eta^\sigma(x) \leq 1$. If $x < \bar{x}$, then

$$\eta(x) < \frac{r + \bar{s} x^\theta}{r + x} \leq \frac{r + \bar{s} \bar{x}^\theta}{r} = 1 + \frac{\bar{x}}{r},$$

and

$$\eta^\sigma(x) \leq \alpha := \left(1 + \frac{\bar{x}}{r}\right)^\sigma. \tag{5.11}$$

Let $\beta \in (0, 1)$ be any discount factor. Then, there exists $r \geq 1$ such that $\alpha\beta < 1$, and from (5.10) and (5.11) it follows that assumption (M1) is satisfied.

Example 2. Let us consider again the model from Example 1 but with $u(x, a, b) = \ln a$, $a \in A(x) = [0, x]$. This utility function has a number of applications in economics (see Stokey et al. 1989). Nonetheless, the two-sided weighted norm approach cannot be employed, because $\ln(0) = -\infty$. Assume now that the state evolution equation is of the form

$$x_{n+1} = (1 + \rho_0)(x_n - a_n)\xi_n, \quad n \in \mathbb{N},$$

where $\rho_0 > 0$ is a constant rate of growth and ξ_n is an additional random income (shock) received in period n. Let $\omega(x) = r + \ln(1 + x)$ for all $x \in X$ and some $r \geq 1$. Clearly, $u^+(x, a, b) = \max\{0, \ln a\} \leq \max\{0, \ln x\} \leq \omega(x)$ for all $(x, a, b) \in K$. By Jensen's inequality it follows that

$$\int_X \omega(y) q(dy|x, a, b) = \int_0^\infty \omega((x-a)(1+\rho_0)+s) b(ds) \leq r + \ln(1 + x(1+\rho_0)\bar{s})$$

for all $(x, a, b) \in K$. Thus

$$\frac{\int_X \omega(y)q(dy|x,a)}{\omega(x)} \le \psi(x) := \frac{r + \ln(1 + x(1+\rho_0)\bar{s})}{r + \ln(1+x)}. \tag{5.12}$$

If we assume that $\bar{s}(1+\rho_0) > 1$, then

$$\psi(x) - 1 = \frac{\ln\left(\frac{1+(1+\rho_0)\bar{s}x}{1+x}\right)}{r + \ln(1+x)} \le \frac{1}{r}\ln\left(\frac{1+(1+\rho_0)\bar{s}x}{1+x}\right) \le \frac{1}{r}\ln(\bar{s}(1+\rho_0)).$$

Hence

$$\psi(x) \le \alpha := 1 + \frac{1}{r}\ln(\bar{s}(1+\rho_0)).$$

Choose now any $\beta \in (0,1)$. If r is sufficiently large, then $\alpha\beta < 1$ and by (5.12) condition (M1) holds.

Example 3 (A growth model with additive shocks). Consider the model from Example 1 with the following state evolution equation:

$$x_{n+1} = (1+\rho_0)(x_n - a_n) + \xi_n, \quad n \in \mathbb{N},$$

where ρ_0 is constant introduced in Example 2. The transition probability q is now of the form

$$q(D|x,a,b) = \int_0^\infty 1_D((1+\rho_0)(x-a) + s)b(ds),$$

where $D \in \mathcal{B}(X)$. If $\phi \in C(X)$, then the integral

$$\int_X \phi(y)q(dy|x,a) = \int_0^\infty \phi((1+\rho_0)(x-a) + s)b(ds)$$

is continuous in $(x,a,b) \in K$. Let the function ω be as in (5.9). Applying Jensen's inequality we obtain

$$\int_X \omega(y)q(dy|x,a,b) = \int_0^\infty \omega((x-a)(1+\rho_0) + s)b(ds)$$
$$\le \omega(x(1+\rho_0) + \bar{s}) = (r + x(1+\rho_0) + \bar{s})^\sigma.$$

Thus,

$$\frac{\int_X \omega(y)q(dy|x,a,b)}{\omega(x)} \le \eta_0^\sigma(x), \quad \text{where} \quad \eta_0(x) := \frac{r + x(1+\rho_0) + \bar{s}}{r + x}, \quad x \in X.$$

Take $r > \bar{s}/\rho_0$ and note that

$$\lim_{x\to 0+} \eta_0(x) = 1 + \frac{\bar{s}}{r} < \lim_{x\to\infty} \eta_0(x) = 1 + \rho_0.$$

Hence,

$$\sup_{(x,a,b)\in K} \frac{\int_X \omega(y) q(dy|x,a,b)}{\omega(x)} \le \sup_{x\in X} \eta_0^\sigma(x) = (1+\rho_0)^\sigma.$$

Therefore, condition (M1) holds for all $\beta \in (0,1)$ such that $\beta(1+\rho_0)^\sigma < 1$.

For other examples involving quadratic cost/payoff functions and linear evolution of the system, the reader is referred to Jaśkiewicz and Nowak (2011).

3.1.2 An Application to the Hansen-Sargent Model in Macroeconomics

In this subsection, we study maxmin control model, in which minimizing player (nature) helps the controller to design a decision rule that is robust to misspecification of a dynamic approximating model linking controls today to state variables tomorrow. The constraint on nature is represented by a cost based on a reference transition probability q. Nature can deviate away from q, but the larger the deviation, the higher the cost. In particular, this cost is proportional to the relative entropy $I(\hat{q}||q)$ between the chosen probability $\hat{q}$ and the reference probability q, i.e., the cost equals to $\theta_0 I(\hat{q}||q)$, where $\theta_0 > 0$. Such preferences in macroeconomics are called multiplier preferences (see Hansen and Sargent 2008).

Let us consider the following scalar system:

$$x_{n+1} = x_n + a_n + \varepsilon_n + b_n, \quad n \in \mathbb{N}, \tag{5.13}$$

where $x_n \in X = \mathbb{R}$, $a_n \in A(x_n) \equiv A = [0, \hat{a}]$ is an action selected by the controller and $b_n \in B(x_n, a_n) \equiv B = (-\infty, 0]$ is a parameter chosen by the *malevolent nature*. The sequence of random variables $(\varepsilon_n)_{n\in\mathbb{N}}$ is i.i.d., where ε_n follows the standard Gaussian distribution with the density denoted by ϕ. At each period the controller selects a control $a \in A$, which incurs the payoff $u_0(x,a)$. It is assumed that the function u_0 is upper semicontinuous on $X \times A$. The controller has a unique explicitly specified approximating model (when $b_n \equiv 0$ for all n) but concedes that data might actually be generated by a number of set of models that surround the approximating model.

Let $n \in \mathbb{N}$ be fixed. By p we denote the conditional density of variable $Y = x_{n+1}$ implied by equation (5.13). Setting $a = a_n$, $x = x_n$, and $b_n = b$ we obtain that

$$p(y|x,a,b) = \frac{1}{\sqrt{2\pi}} e^{-\frac{(y-x-a-b)^2}{2}} \quad \text{for} \quad y \in \mathbb{R}.$$

Clearly, $p(\cdot|x,a,b)$ defines the probability measure q, where

$$q(D|x,a,b) = \int_D p(y|x,a,b)dy \quad \text{for } D \subset \mathcal{B}(\mathbb{R}).$$

If $b = 0$, then we deal with the baseline model. Hence, the relative entropy

$$I(q(\cdot|x,a,b)||q(\cdot|x,a,0)) = \frac{1}{2}b^2,$$

and consequently, the payoff function in the model is

$$u(x,a,b) = u_0(x,a) + \frac{1}{2}\theta_0 b^2.$$

The term $\frac{1}{2}\theta_0 b^2$ is a penalized cost paid by nature. The parameter θ_0 can be viewed as the degree of robustness. For example, if θ_0 is large, then the penalization becomes so great that only the nominal model remains and the strategy is less robust. Conversely, the lower values of θ_0 allow to design a strategy which is appropriate for a wider set of model misspecifications. Therefore, a lower θ_0 is equivalent to a higher degree of robustness.

Within such a framework, we shall consider pure strategies for nature. A strategy $\gamma = (\gamma_n)_{n\in\mathbb{N}}$ is an *admissible* strategy to nature, if $\gamma_n : H_n^* \to B$ is a Borel measurable function, i.e., $b_n = \gamma_n(h_n^*)$, $n \in \mathbb{N}$, and for every $x \in X$ and $\pi \in \Pi$

$$E_x^{\pi\gamma}\left(\sum_{n=1}^{\infty}\beta^{n-1}b_n^2\right) < \infty.$$

The set of all admissible strategies to nature is denoted by Γ_0^*.

The objective of the controller is to find a policy $\pi^* \in \Pi$ such that

$$\inf_{\gamma\in\Gamma_0^*} E_x^{\pi^*\gamma}\left(\sum_{n=1}^{\infty}\beta^{n-1}\left\{u_0(x_n,a_n) + \frac{1}{2}\theta_0 b_n^2\right\}\right) =$$
$$\max_{\pi\in\Pi}\inf_{\gamma\in\Gamma_0^*} E_x^{\pi\gamma}\left(\sum_{n=1}^{\infty}\beta^{n-1}\left\{u_0(x_n,a_n) + \frac{1}{2}\theta_0 b_n^2\right\}\right).$$

We solve the problem by proving that there exists a solution to the optimality equation. First, we note that assumption (M1) is satisfied for $\omega(x) = \max\{x,0\} + r$, where $r \geq 1$ is some constant. Indeed, on page 268 in Jaśkiewicz and Nowak (2011), it is shown that for every discount factor $\beta \in (0,1)$, we may choose sufficiently large $r \geq 1$ such that $\alpha\beta < 1$, where $\alpha = 1 + (\hat{a}+1)/r$. Further, we shall assume that $\sup_{a\in A} u_0^+(x,a) \leq d\omega(x)$ for all $x \in X$.

For any function $\phi \in \underline{U}_\omega(X)$, we define the operator $\mathcal{T}_\beta$ as follows:

$$\mathcal{T}_\beta\phi(x) = \max_{a\in A}\inf_{b\in B}\left[u_0(x,a) + \frac{1}{2}\theta_0 b^2 + \beta\int_X \phi(y)q(dy|x,a,b)\right]$$

for all $x \in X$. Clearly, $\mathcal{T}_\beta$ maps the space $\underline{U}_\omega(X)$ into itself. Indeed, we have

$$\mathcal{T}_\beta\phi(x) \le \max_{a\in A}\left[u_0(x,a) + \beta\int_X \phi(y)q(dy|x,a,b)\right] \le d\omega(x) + \beta\alpha\|\phi^+\|_\omega\omega(x)$$

for all $x \in X$. Hence, $(\mathcal{T}_\beta\phi)^+ \in U_\omega(X)$ and by Lemma 1, $\mathcal{T}_\beta\phi$ is upper semicontinuous. Proceeding analogously as in the proof of Theorem 1, we infer that $v_\beta \in \underline{U}_\omega(X)$, where $v_\beta = \mathcal{T}_\beta v_\beta$ and there exits $f^* \in F$ such that

$$\begin{aligned} v_\beta(x) = \mathcal{T}_\beta v_\beta(x) &= \max_{a\in A}\inf_{b\in B}\left[u_0(x,a) + \frac{1}{2}\theta_0 b^2 + \beta\int_X v_\beta(y)q(dy|x,a,b)\right] \\ &= \inf_{b\in B}\left[u_0(x,f^*(x)) + \frac{1}{2}\theta_0 b^2 + \beta\int_X v_\beta(y)q(dy|x,f^*(x),b)\right] \end{aligned} \tag{5.14}$$

for $x \in X$. Finally, we may formulate the following result.

Proposition 5. *Consider the system given in (5.13). Then, $v_\beta \in \underline{U}_\omega(X)$ and there exists a stationary strategy f^* such that (5.14) is satisfied for all $x \in X$. The strategy f^* is optimal for the controller.*

3.2 Average Reward Robust Markov Decision Process

In this subsection, we assume that u takes values in $\mathbb{R}$ rather than in $\underline{\mathbb{R}}$. Moreover, the action set of nature is independent of $(x,a) \in K_A$, i.e., $B(x,a) \equiv B$, where B is a compact metric space. Obviously, (C3) is then immediately satisfied. Since we consider the average payoff in the maxmin control problem, we impose a bit stronger assumptions than in the previous subsection. Below are their counterparts.

($\tilde{\text{C}}$1) For any $x \in X$, $A(x)$ is compact and the set-valued mapping $x \to A(x)$ is continuous.

($\tilde{\text{W}}$1) The payoff function u is continuous on K.

A strategy for the opponent is a sequence $\gamma = (\gamma_n)_{n\in\mathbb{N}}$ of Borel measurable mappings $\gamma_n : H_n^* \to B$ rather than a sequence of stochastic kernels. The set of all strategies for the opponent is denoted by Γ_0^*.

For any initial state $x \in X$ and strategies $\pi \in \Pi$, $\gamma \in \Gamma_0^*$, we set $u_n^-(x,\pi,\gamma) = E_x^{\pi\gamma}[u^-(x_n,a_n,b_n)]$, $u_n^+(x,\pi,\gamma) = E_x^{\pi\gamma}[u^+(x_n,a_n,b_n)]$, and

$u_n(x, \pi, \gamma) = E_x^{\pi\gamma}[u(x_n, a_n, b_n)]$, provided that the integral is well defined, i.e., either $u_n^+(x, \pi, \gamma) < +\infty$ or $u_n^-(x, \pi, \gamma) > -\infty$. Note that $u_n(x, \pi, \gamma)$ is the n-stage expected payoff. For $x \in X$, strategies $\pi \in \Pi$, $\gamma \in \Gamma_0^*$, and $\beta \in (0, 1)$, we define $J_\beta^-(x, \pi, \gamma)$ and $J_\beta^+(x, \pi, \gamma)$ as in (5.1) and in (5.2). Assuming that these expressions are finite, we define the expected discounted payoff to the controller as in (5.3). Clearly, the maxmin value v_β is defined as in the previous subsection, i.e.,

$$v_\beta(x) = \sup_{\pi \in \Pi} \inf_{\gamma \in \Gamma_0^*} J_\beta(x, \pi, \gamma).$$

For any initial state $x \in X$, strategies $\pi \in \Pi$, $\gamma \in \Gamma_0^*$, and $n \in \mathbb{N}$, we let

$$J_n^-(x, \pi, \gamma) := E_x^{\pi\gamma}\left[\sum_{m=1}^{n} u^-(x_m, a_m, b_m)\right] \quad \text{and}$$

$$J_n^+(x, \pi, \gamma) := E_x^{\pi\gamma}\left[\sum_{m=1}^{n} u^+(x_m, a_m, b_m)\right].$$

If these expressions are finite, we can define the total expected n-stage payoff to the controller as follows:

$$J_n(x, \pi, \gamma) := J_n^+(x, \pi, \gamma) + J_n^-(x, \pi, \gamma).$$

Clearly, we have that

$$J_n(x, \pi, \gamma) = \sum_{m=1}^{n} u_m(x, \pi, \gamma).$$

Furthermore, we set

$$\overline{J}_n^-(x, \pi, \gamma) = \frac{J_n^-(x, \pi, \gamma)}{n}, \qquad \overline{J}_n^+(x, \pi, \gamma) = \frac{J_n^+(x, \pi, \gamma)}{n},$$

and

$$\overline{J}_n(x, \pi, \gamma) = \frac{J_n(x, \pi, \gamma)}{n}.$$

The robust expected average payoff per unit time (average payoff, for short) is defined as follows:

$$\hat{R}(x, \pi) = \liminf_{n \to \infty} \inf_{\gamma \in \Gamma_0^*} \overline{J}_n(x, \pi, \gamma). \tag{5.15}$$

A strategy $\bar{\pi} \in \Pi$ is called an *optimal robust strategy* for the controller in the average payoff case, if $\sup_{\pi \in \Pi} \hat{R}(x, \pi) = \hat{R}(x, \bar{\pi})$ for each $x \in X$.

We can now formulate our assumption.

(D) There exist functions $D^+ : X \to [1, \infty)$ and $D^- : X \to [1, \infty)$ such that

$$\overline{J}_n^+(x, \pi, \gamma) \le D^+(x) \quad \text{and} \quad |\overline{J}_n^-(x, \pi, \gamma)| \le D^-(x)$$

for every $x \in X, \pi \in \Pi, \gamma \in \Gamma_0^*$ and $n \in \mathbb{N}$. Moreover, D^+ is continuous and the function $(x, a, b) \to \int_X D^+(y)q(dy|x, a, b)$ is continuous on K.

Condition (D) trivially holds if the payoff function u is bounded. The models with unbounded payoffs satisfying (D) are given in Jaśkiewicz and Nowak (2014) (see Examples 1 and 2). Our aim is to consider the robust expected average payoff per unit time. The analysis is based upon studying the so-called optimality inequality, which is obtained via vanishing discount factor approach. However, we note that we cannot use the results from previous subsection, since in our approach we must take a sequence of discount factors converging to one. Theorem 1 was obtained under assumption (M1). Unfortunately, in our case this assumption is useless. Clearly, if $\alpha > 1$, as it happens in Examples 1, 2, and 3, the requirement $\alpha\beta < 1$ is a limitation and makes impossible to define a desirable sequence $(\beta_n)_{n\in\mathbb{N}}$ converging to one. Therefore, we first reconsider the robust discounted payoff model under different assumption.

Put $w(x) = D^+(x)/(1 - \beta)$, $x \in X$. Let $\tilde{U}_w(X)$ be the space of all real-valued upper semicontinuous functions $v : X \to \mathbb{R}$ such that $v(x) \le w(x)$ for all $x \in X$. Assume now that $\phi \in \tilde{U}_w(X)$ and $f \in F$. For every $x \in X$ we set (recall (5.6))

$$T_\beta\phi(x) = \sup_{a\in A(x)} \inf_{b\in B} \left[u(x, a, b) + \beta \int_X \phi(y)q(dy|x, a, b) \right]. \tag{5.16}$$

The following result is Theorem 1 in Jaśkiewicz and Nowak (2014).

Theorem 2. *Assume* ($\tilde{\text{C}}$1),($\tilde{\text{W}}$1), (W2)*, and* (D)*. Then, for each* $\beta \in (0, 1)$*,* $v_\beta \in \tilde{U}_w(X)$*,* $v_\beta = T_\beta v_\beta$*, and there exists* $f^* \in F$ *such that*

$$v_\beta(x) = \inf_{b\in B} \left[u(x, f^*(x), b) + \beta \int_X v_\beta(y)q(dy|x, f^*(x), b) \right], \quad x \in X.$$

Moreover, $v_\beta(x) = \sup_{\pi\in\Pi} \inf_{\gamma\in\Gamma_0^*} J_\beta(x, \pi, \gamma) = \inf_{\gamma\in\Gamma_0^*} J_\beta(x, f^*, \gamma)$ *for each* $x \in X$*, i.e.,* f^* *is optimal.*

Remark 4. The proof of Theorem 2 is to some extent standard, but as mentioned we cannot apply the Banach contraction principle (see for instance Blackwell 1965 or Bertsekas and Shreve 1996). The majority of papers that deal with maximization of the expected discounted payoff assume that the one-stage payoff function is bounded from above (see Hernández-Lerma and Lasserre 1996; Schäl 1975) or it satisfies inequality (5.7). Neither requirement is met in this framework. Therefore, we have to consider truncated models and finite horizon maxmin problems.

In order to establish the optimality inequality, we shall need a generalized Tauberian relation, which plays a crucial role in proving Theorem 3 stated below.

For any sequence $(u_k)_{k\in\mathbb{N}}$ of real numbers, let $\overline{u}_n := \frac{1}{n}\sum_{k=1}^{n} u_k$ for any $n \in \mathbb{N}$. Fix a constant $D \geq 1$ and consider the set S_D of all sequences $(u_k)_{k\in\mathbb{N}}$ such that $|\overline{u}_n| \leq D$ for each $n \in \mathbb{N}$. Assume now that the elements of the sequence $(u_k(\xi))_{k\in\mathbb{N}} \in S_D$ may depend on ξ belonging to some set $\varXi$. Define

$$\overline{u}_n(\xi) = \frac{1}{n}\sum_{k=1}^{n} u_k(\xi)$$

and

$$v_\beta = \inf_{\xi\in\varXi}(1-\beta)\sum_{k=1}^{\infty}\beta^{k-1}u_k(\xi) \quad \text{for} \quad \beta \in (0,1), \quad v_n := \inf_{\xi\in\varXi}\overline{u}_n(\xi).$$

Proposition 6. *Assume that $(u_n(\xi))_{n\in\mathbb{N}} \in S_D$ for each $\xi \in \varXi$. Then, we have the following*

$$\liminf_{\beta\to 1^-} v_\beta \geq \liminf_{n\to\infty} v_n.$$

Proposition 6 extends Proposition 4 and Corollary 5 in Lehrer and Sorin (1992) that are established under the assumption that $0 \leq u_n(\xi) \leq 1$ for every $n \in \mathbb{N}$ and $\xi \in \varXi$. This result is related to the so-called Tauberian relations. Recent advances on this issue can be found in Renault (2014) (see also the discussion in Sect. 7). It is worth mentioning that Proposition 6 is also useful in the study of risk-sensitive control models (see Jaśkiewicz 2007 or Appendix in Jaśkiewicz and Nowak 2014).

Let us fix a state $z \in X$ and define

$$h_\beta(x) := V_\beta(x) - V_\beta(z), \quad \text{for } x \in X \text{ and } \beta \in (0,1).$$

Furthermore, we make the following assumptions.

(B1) There exists a function $M : X \to (-\infty, 0]$ such that $\inf_{\beta\in(0,1)} h_\beta(x) \geq M(x)$, and there exists a continuous function $Q : X \to [0, +\infty)$ such that $\sup_{\beta\in(0,1)} h_\beta(x) \leq Q(x)$ for every $x \in X$. Moreover, the function $(x,a,b) \to \int_X Q(y)q(dy|x,a,b)$ is continuous on K.

(B2) For any $x \in X$, $\pi \in \Pi$, and $\gamma \in \Gamma_0^*$, it holds that

$$\lim_{n\to\infty}\frac{E_x^{\pi\gamma}[Q(x_n)]}{n} = 0.$$

The main result in Jaśkiewicz and Nowak (2014) is as follows.

Theorem 3. *Assume* ($\tilde{\text{C}}$1), ($\tilde{\text{W}}$1), (W2), (D), *and* (B1)–(B2). *Then, there exist a constant g, a real-valued upper semicontinuous function h, and a stationary strategy $\bar{f} \in F$ such that*

$$h(x) + g \le \sup_{a \in A(x)} \inf_{b \in B} \left[u(x,a,b) + \int_X h(y) q(dy|x,a,b) \right]$$
$$= \inf_{b \in B} \left[u(x,\bar{f}(x),b) + \int_X h(y) q(dy|x,\bar{f}(x),b) \right]$$

for $x \in X$. Moreover, $g = \sup_{\pi \in \Pi} \hat{R}(x,\pi) = \hat{R}(x,\bar{f})$ for all $x \in X$, i.e., $\bar{f}$ is the optimal robust strategy.

4 Discounted and Positive Stochastic Markov Games with Simultaneous Moves

From now on we assume that $B(x,a) = B(x)$ is independent of $a \in A(x)$ for each $x \in X$. Therefore, we now have $K_A \in \mathcal{B}(X \times A)$,

$$K_B \in \mathcal{B}(X \times B), \quad \text{and} \quad K := \{(x,a,b) : x \in X, a \in A(x), b \in B(x)\}. \tag{5.17}$$

Thus, at every stage $n \in \mathbb{N}$, player 2 does not observe player 1's action $a_n \in A(x_n)$ in state $x_n \in X$. One can say that the players act simultaneously and play the standard discounted stochastic game as in the seminal work of Shapley (1953). It is assumed that both players know at every stage $n \in \mathbb{N}$ the entire history of the game up to state $x_n \in X$. Now a *strategy* for player 2 is a sequence $\gamma = (\gamma_n)_{n \in \mathbb{N}}$ of Borel (or universally measurable) transition probabilities γ_n from H_n to B such that $\gamma_n(B(x_n)|h_n) = 1$ for each $h_n \in H_n$. The set of all Borel (universally) measurable strategies for player 2 is denoted by Γ (Γ_u). Let G (G_u) be the set of all Borel (universally) measurable mappings $g : X \to \Pr(B)$ such that $g(x) \in \Pr(B(x))$ for all $x \in X$. Every $g \in G_u$ induces a transition probability $g(db|x)$ from X to B and is recognized as a randomized *stationary strategy* for player 2. A *semistationary strategy* for player 2 is determined by a Borel or universally measurable function $g : X \times X \to \Pr(B)$ such that $g(x,x') \in \Pr(B(x'))$ for all $(x,x') \in X \times X$. Using a semistationary strategy, player 2 chooses an action $b_n \in B(x_n)$ on any stage $n \ge 2$ according to the probability measure $g(x_1,x_n)$ depending on x_n and the initial state x_1. Let F (F_u) be the set of all Borel (universally) measurable mappings $f : X \to \Pr(A)$ such that $f(x) \in \Pr(A(x))$ for all $x \in X$. Then, F (F_u) can be considered as the set of all randomized *stationary strategies* for player 1. The set of all Borel (universally) measurable strategies for player 1 is denoted by Π (Π_u). For any initial state $x \in X$, $\pi \in \Pi_u$, $\gamma \in \Gamma_u$, the expected discounted payoff function

$J_\beta(x,\pi,\gamma)$ is well defined under conditions (M1) and (M2). Since $\Pi \subset \Pi_u$ and $\Gamma \subset \Gamma_u$, $J_\beta(x,\pi,\gamma)$ is well defined for all $\pi \in \Pi$, $\gamma \in \Gamma$. If we restrict attention to Borel measurable strategies, then the *lower value* of the game is

$$\underline{v}_\beta(x) = \sup_{\pi\in\Pi} \inf_{\gamma\in\Gamma} J_\beta(x,\pi,\gamma)$$

and the *upper value* of the game is

$$\overline{v}_\beta(x) = \inf_{\gamma\in\Gamma} \sup_{\pi\in\Pi} J_\beta(x,\pi,\gamma),\ x \in X.$$

Suppose that the stochastic game has a *value*, i.e., $v_\beta(x) := \underline{v}_\beta(x) = \overline{v}_\beta(x)$, for each $x \in X$. Then, under our assumptions (M1) and (M2), $v_\beta(x) \in \underline{\mathbb{R}}$. Let $\underline{X} := \{x \in X : v_\beta(x) = -\infty\}$. A strategy $\pi^* \in \Pi$ is *optimal* for player 1 if

$$\inf_{\gamma\in\Gamma} J_\beta(x,\pi^*,\gamma) = v_\beta(x) \ \text{ for all } x \in X.$$

Let $\varepsilon > 0$ be fixed. A strategy $\gamma^* \in \Gamma$ is ε-*optimal* for player 2 if

$$\sup_{\pi\in\Pi} J_\beta(x,\pi,\gamma^*) = v_\beta(x) \ \text{ for all } \ x \in X \setminus \underline{X} \quad \text{and}$$

$$\sup_{\pi\in\Pi} J_\beta(x,\pi,\gamma^*) < -\frac{1}{\varepsilon} \ \text{ for all } x \in \underline{X}.$$

Similarly, the value v_β and ε-optimal or optimal strategies can be defined in the class of universally measurable strategies. Let

$$\bar{K}_A := \{(x,\nu) : x \in X, \nu \in \Pr(A(x))\}, \quad \bar{K}_B := \{(x,\rho) : x \in X, \rho \in \Pr(B(x))\},$$

and

$$\bar{K} := \{(x,\nu,\rho) : x \in X, \nu \in \Pr(A(x)), \rho \in \Pr(B(x))\}.$$

For any $(x,\nu,\rho) \in \bar{K}$ and $D \in \mathcal{B}(X)$, define

$$u(x,\nu,\rho) := \int_{A(x)}\int_{B(x)} u(x,a,b)\rho(db)\nu(da)$$

and

$$q(D|x,\nu,\rho) := \int_{A(x)}\int_{B(x)} q(D|x,a,b)\rho(db)\nu(da).$$

If $f \in F_u$ and $g \in G_u$, then

$$u(x,f,g) := u(x,f(x),g(x)) \quad \text{and} \quad q(D|x,f,g) := q(D|x,f(x),g(x)). \tag{5.18}$$

For any $(x, \nu, \rho) \in \bar{K}$ and $\phi \in \underline{U}_\omega(X)$, define

$$L_\beta\phi(x,\nu,\rho) = u(x,\nu,\rho) + \beta \int_X \phi(y)q(dy|x,\nu,\rho) \tag{5.19}$$

and

$$T_\beta\phi(x) = \max_{\nu\in\Pr(A(x))} \inf_{\rho\in\Pr(B(x))} L_\beta\phi(x,\nu,\rho). \tag{5.20}$$

By Lemma 7 in Jaśkiewicz and Nowak (2011), the operator T_β is well defined, and using the maximum theorem of Berge (1963), it can be proved that $T_\beta\phi \in \underline{U}_\omega(X)$ for any $\phi \in \underline{U}_\omega(X)$.

Theorem 4. *Assume* (C1), (W1)–(W2), *and* (M1)–(M2). *In addition, let the correspondence $x \to B(x)$ be lower semicontinuous and let every set $B(x)$ be a complete subset of B. Then, the game has a value $v_\beta \in \underline{U}_\omega(X)$, player 1 has an optimal stationary strategy $f^* \in F$ and*

$$T_\beta v_\beta(x) = v_\beta(x) = \max_{\nu\in\Pr(A(x))} \inf_{\rho\in\Pr(B(x))} L_\beta v_\beta(x,\nu,\rho) = \inf_{\rho\in\Pr(B(x))} L_\beta v_\beta(x, f^*(x), \rho)$$

for each $x \in X$. Moreover, for any $\varepsilon > 0$, player 2 has an ε-optimal Borel measurable semistationary strategy.

The assumption that every $B(x)$ is complete in B is made to assure that $G \neq \emptyset$ (see Kuratowski and Ryll-Nardzewski 1965). The construction of an ε-optimal semistationary strategy for player 2 is based on using "truncated games" $\mathcal{G}_k$ with the payoff functions $u_k := \max\{u, -k\}$, $k \in \mathbb{N}$. In every game $\mathcal{G}_k$ player 2 has an $\frac{\varepsilon}{2}$-optimal stationary strategy, say $g_k^* \in G$. If $v_{\beta,k}$ is the value function of the game $\mathcal{G}_k$, then it is shown that $v_\beta(x) = \inf_{k\in\mathbb{N}} v_{\beta,k}(x)$ for all $x \in X$. This fact can be easily used to construct a measurable partition $\{X_n\}_{n\in Z}$ of the state space ($Z \subset \mathbb{N}$) such that $v_\beta(x) > v_{\beta,k}(x) - \frac{\varepsilon}{2}$ for all $x \in X_k$, $k \in Z$. If $g^*(x,x') := g_n^*(x')$ for every $x \in X_n$, $n \in Z$ and for each $x' \in X$, then g^* is an ε-optimal semistationary strategy for player 2. The above definition is valid, if $v_\beta(x) > -\infty$ for all $x \in X$. If $v_\beta(x) = -\infty$ for some state $x \in X$, then the reader is referred to the proof of Theorem 2 in Jaśkiewicz and Nowak (2011), where a modified construction of the ε-optimal semistationary strategy is provided.

Remark 5. Zero-sum discounted stochastic games with a compact metric state space and weakly continuous transitions were first studied by Maitra and Parthasarathy (1970). Kumar and Shiau (1981) extended their result to Borel state space games with bounded continuous payoff functions and weakly continuous transitions. Couwenbergh (1980) studied continuous games with unbounded payoffs and a metric state space using the weighted supremum norm approach introduced by

Wessels (1977). He proved that both players possess optimal stationary strategies. In order to obtain such a result, additional conditions should be imposed. Namely, the function u is continuous and such that $|u(x,a,b)| \le d\omega(x)$ for some constant $d > 0$ and all $(x,a,b) \in K$. Moreover, the mappings $x \to A(x)$ and $x \to B(x)$ are compact valued and continuous. It should be noted that our condition (M2) allows for much larger class of models and is less restrictive for discount factors compared with the weighted supremum norm approach. We also point out that a class of zero-sum lower semicontinuous stochastic games with weakly continuous transition probabilities and bounded from below nonadditive payoff functions was studied by Nowak (1986).

A similar result can also be proved under the following conditions:

(C4) $A(x)$ is compact for each $x \in X$.
(C5) The payoff function u is Borel measurable and $u(x,\cdot,b)$ is upper semicontinuous and $q(D|x,\cdot,b)$ is continuous on $A(x)$ for any $D \in \mathcal{B}(X)$, $x \in X$, $b \in B(x)$.

A simple modification of the proof of Theorem 2 in Jaśkiewicz and Nowak (2011) using appropriately adapted theorems on measurable minmax selections proved in Nowak (1985b) yields the following result:

Theorem 5. *Assume* (C4)–(C5) *and* (M1)–(M2). *Then, the game has a value v_β, which is a lower semianalytic function on X. Player 1 has an optimal stationary strategy $f^* \in F_u$ and*

$$T_\beta v_\beta(x) = v_\beta(x) = \max_{\nu \in \Pr(A(x))} \inf_{\rho \in \Pr(B(x))} L_\beta v_\beta(x,\nu,\rho) = \inf_{\rho \in \Pr(B(x))} L_\beta v_\beta(x, f^*(x), \rho)$$

for each $x \in X$. Moreover, for any $\varepsilon > 0$, player 2 has an ε-optimal universally measurable semistationary strategy.

Maitra and Parthasarathy (1971) first studied *positive stochastic games,* where the stage payoff function $u \ge 0$ and $\beta = 1$. The extended payoff in a positive stochastic game is

$$J_p(x,\pi,\gamma) := E_x^{\pi\gamma}\left(\sum_{n=1}^{\infty} u(x_n,a_n,b_n)\right), \quad x = x_1 \in X,\ \pi \in \Pi,\ \gamma \in \Gamma.$$

Using standard iteration arguments as in Strauch (1966) or Bertsekas and Shreve (1996), one can show that $J_p(x,\pi,\gamma) < \infty$ if and only if there exists a nonnegative universally measurable function w on X such that the following condition holds:

$$u(x,a,b) + \int_X w(y)q(dy|x,a,b) \le w(x) \quad \text{for all} \quad (x,a,b) \in K. \qquad (5.21)$$

Value functions and ε-optimal strategies are defined in positive stochastic games in an obvious manner. Studying positive stochastic games, it is convenient to use approximation of $J_p(x,\pi,\gamma)$ from below by $J_\beta(x,\pi,\gamma)$ as β goes to 1. To make this method effective we must change our assumptions on the primitives in the way described below.

(C6) $B(x)$ is compact for each $x \in X$.
(C7) The payoff function u is Borel measurable and $u(x,a,\cdot)$ is lower semicontinuous and $q(D|x,a,\cdot)$ is continuous on $B(x)$ for any $D \in \mathcal{B}(X)$, $x \in X$, $a \in A(x)$.

As noted in the preliminaries, assumption (C6) implies that $\emptyset \neq G \subset G_u$ and $F_u \neq \emptyset$. Let L_1 and T_1 be the operators defined as in (5.19) and (5.20), respectively, but with $\beta = 1$.

Theorem 6. *Assume that* (5.21) *and* (C6)–(C7) *hold. Then the positive stochastic game has a value function v_p which is upper semianalytic and $v_p(x) = \sup_{\beta\in(0.1)} v_\beta(x)$ for all $x \in X$. Moreover, v_p is the smallest nonnegative upper semianalytic solution to the equation*

$$T_1 v(x) = v(x), \quad x \in X.$$

Player 2 has an optimal stationary strategy $g^* \in G_u$ such that

$$T_1 v_p(x) = \sup_{\nu\in\Pr(A(x))} \min_{\rho\in\Pr(B(x))} L_1 v_p(x,\nu,\rho) = \sup_{\nu\in\Pr(A(x))} L_1 v_p(x,\nu,g^*(x)), \quad x \in X$$

and for any $\varepsilon > 0$, player 1 has an ε-optimal universally measurable semistationary strategy.

Theorem 6 is a version of Theorem 5.4 in Nowak (1985a). Some special cases under much stronger continuity assumptions were considered by Maitra and Parthasarathy (1971) for games with compact state spaces and by Kumar and Shiau (1981) for games with a Borel state space and finite action sets in each state. An essential part of the proof of Theorem 6 is Proposition 2.

A similar result holds for positive semicontinuous games satisfying the following conditions:

(C8) For any $x \in X$, $A(x)$ is a complete set in A and the correspondence $x \to A(x)$ is lower semicontinuous.
(C9) For any $x \in X$, $B(x)$ is compact and the correspondence $x \to B(x)$ is upper semicontinuous.
(W3) $u \geq 0$ and u is lower semicontinuous on K.

Theorem 7. *Assume* (5.21), (C8)–(C9), *and* (W3). *Then, the positive stochastic game has a value function* v_p *which is lower semicontinuous and* $v_p(x) = \sup_{\beta\in(0,1)} v_\beta(x)$ *for all* $x \in X$. *Moreover,* v_p *is the smallest nonnegative lower semicontinuous solution to the equation*

$$T_1 v(x) = v(x), \quad x \in X. \tag{5.22}$$

Player 2 has an optimal stationary strategy $g^* \in G$ *such that*

$$T_1 v_p(x) = \sup_{\nu\in\Pr(A(x))} \min_{\rho\in\Pr(B(x))} L_1 v_p(x,\nu,\rho) = \sup_{\nu\in\Pr(A(x))} L_1 v_p(x,\nu,g^*(x)), \quad x \in X$$

and for any $\varepsilon > 0$, *player 1 has an* ε*-optimal Borel measurable semistationary strategy.*

The proof of Theorem 7 is similar to that of Theorem 6 and makes use of Proposition 3.

Player 1 need not have an optimal strategy even if X is finite. This is shown in Kumar and Shiau (1981) in Example 1 (see also pages 192–193 in Maitra and Sudderth 1996), which was inspired by Everett (1957). We present this example below.

Example 4. Let $X = \{-1,0,1\}$, $A = \{0,1\}$, $B = \{0,1\}$. States $x = -1$ and $x = 1$ are absorbing with zero payoffs. If $x = 0$ and both players choose the same actions ($a = 1 = b$ or $a = 0 = b$), then $u(x,a,b) = 1$ and $q(-1|0,a,b) = 1$. Moreover, $q(0|0,0,1) = q(1|0,1,0) = 1$ and $u(0,0,1) = u(0,1,0) = 0$. It is obvious that $v_p(-1) = 0 = v_p(1)$. In state $x = 0$ we obtain the equation $v_p(0) = 1/(2-v_p(0))$, which yields the solution $v_p(0) = 1$. In this game player 1 has no optimal strategy.

If player 2 is dummy, i.e., every set $B(x)$ is a singleton, X is a countable set and v_p is bounded on X, then by Ornstein (1969) player 1 has a stationary ε-optimal strategy. A counterpart of this result does not hold for positive stochastic games.

Example 5. Let $X = \mathbb{N} \cup \{0\}$, $A = \{1,2\}$, $B = \{1,2\}$. State $x = 0$ is absorbing with zero payoffs. Let $x \geq 2$ and $a = 1$. Then $u(x,1,b) = 0$ for $b \in B$ and $q(x-1|x,1,1) = q(x+1|x,1,2) = 1$. If $x \geq 2$ and $a = 2$, then $u(x,2,1) = 0$ and $u(x,2,2) = 1$. In both cases ($b = 1$ or $b = 2$) the game moves to the absorbing state $x = 0$ with probability one. If $x = 1$, then $u(1,a,b) = 1$ and $q(0|1,a,b) = 1$ for all $a \in A$ and $b \in B$. It is obvious that $v_p(0) = 0$ and $v_p(1) = 1$. It is shown that $v_p(x) = (x+1)/2x$ for $x \geq 2$ and player 1 has no stationary ε-optimal strategy. It is easy to check that the function v_p given here is a solution to equation (5.22). It may be interesting to note that also $v(0) = 0$, $v(x) = 1$ for $x \geq 1$ is also a solution to equation (5.22) and $v(x) > v_p(x)$ for $x > 1$. For details see counterexample in

Nowak and Raghavan (1991), whose interesting modification called the "Big Match on the integers" was studied by Fristedt et al. (1995).

The assumption that $q(D|x,a,\cdot)$ is continuous on $B(x)$ for each $(x,a) \in K_A$ and $D \in \mathcal{B}(X)$ is weaker than the norm continuity of $q(\cdot|x,a,b)$ in $b \in B(x)$. However, from the point of view of applications, e.g., in dynamic economic models or engineering problems, the weak continuity assumption of $q(\cdot|x,a,b)$ in $(x,a,b) \in K$ is more useful (see Remark 2).

We close this section with a remark on the weighted evaluation proposed for Markov decision models in Krass et al. (1992) and for zero-sum stochastic games in Filar and Vrieze (1992). The criterion is either a convex combination of discounted evaluation and an average evaluation or a convex combination of two discounted evaluations. In the first case, it is proved that the value of the game exists and that both players have ϵ-optimal strategies. In the second case, it is shown that the value is the unique solution of some system of functional equations and that both players have optimal Markov policies. The idea of using the weighted evaluations was applied to the study of nonzero-sum stochastic games (with finite state and action sets) by Flesch et al. (1999). Zero-sum perfect information games under the weighted discounted payoff criterion were studied by Altman et al. (2000). We would like to point out that discounted utility (payoff) functions belong to the class of "recursive utilities" extensively examined in economics (see Miao 2014). It seems, however, that the weighted discounted utilities are not in this class.

5 Zero-Sum Semi-Markov Games

In this section, we study zero-sum semi-Markov games on a general state space with possibly unbounded payoffs. Different limit-average expected payoff criteria can be used for such games, but under some conditions they turn out to be equivalent. Such games are characterized by the fact that the time between jumps is a random variable with distribution dependent on the state and actions chosen by the players. Most primitive data for a game model considered here are as in Sect. 4. More precisely, let $K_A \in \mathcal{B}(X \times A)$ and $K_B \in \mathcal{B}(X \times B)$. Then, the set K in (5.17) is Borel. As in Sect. 4 we assume that $A(x)$ and $B(x)$ are the admissible action sets for the player 1 and 2, respectively, in state $x \in X$. Let Q be a transition probability from K to $[0,\infty) \times X$. Hence, if $a \in A(x)$ and $b \in B(x)$ are actions chosen by the players in state x, then for $D \in \mathcal{B}(X)$ and $t \geq 0$, $Q([0,t] \times D|x,a,b)$ is the probability that the sojourn time of the process in x will be smaller than t, and the next state x' will be in D. Let $k = (x,a,b) \in K$. Clearly, $q(D|k) = Q([0,\infty] \times D|k)$ is the transition law of the next state. The mean holding time given k is defined as

$$\tau(k) = \int_0^{+\infty} tH(dt|k),$$

where $H(t|k) = Q([0,t] \times X|k)$ is a distribution function of the sojourn time. The payoff function to player 1 is a Borel measurable function $u : K \to \mathbb{R}$ and is usually of the form

$$u(x,a,b) = u^1(x,a,b) + u^2(x,a,b)\tau(x,a,b), \quad (x,a,b) \in K, \tag{5.23}$$

where $u^1(x,a,b)$ is an immediate reward obtained at the transition time and $u^2(x,a,b)$ is the reward rate in the time interval between successive transitions.

The game starts at $T_1 := 0$ and is played as follows. If the initial state is $x_1 \in X$ and the actions $(a_1,b_1) \in A(x_1) \times B(x_1)$ are selected by the players, then the immediate payoff $u^1(x_1,a_1,b_1)$ is incurred for player 1 and the game remains in state x_1 for a random time T_2 that enjoys the probability distribution $H(\cdot|x_1,a_1,b_1)$. The payoff $u^2(x_1,a_1,b_1)$ to player 1 is incurred until the next transition occurs. Afterwards the system jumps to the state x_2 according to the transition law $q(\cdot|x_1,a_1,b_1)$. The situation repeats itself yielding a trajectory $(x_1,a_1,b_1,t_2,x_2,a_2,b_2,t_3,\ldots)$ of some stochastic process, where x_n,a_n,b_n and t_{n+1} describe the state, the actions chosen by the players, and the decision epoch, respectively, on the nth stage of the game. Clearly, t_{n+1} is a realization of the random variable T_{n+1}, and $H(\cdot|x_n,a_n,b_n)$ is a distribution function of the random variable $T_{n+1} - T_n$ for any $n \in \mathbb{N}$.

Strategies and their sets for both players are defined in a similar way as in Sect. 4. The only difference now is that the history of the process also includes the jump epochs, i.e., $h_n = (x_1,a_1,b_1,t_2,\ldots,x_n)$ is the history of the process up to the nth state.

Let $N(t)$ be the number of jumps that have occurred prior to time t, i.e., $N(t) = \max\{n \in \mathbb{N} : T_n \le t\}$. Under our assumptions for each initial state $x \in X$ and any strategies $(\pi,\gamma) \in \Pi \times \Gamma$, we have $P_x^{\pi\gamma}(N(t) < 1) = 1$ for any $t \ge 0$.

For any pair of strategies $(\pi,\gamma) \in \Pi \times \Gamma$ and an initial state $x \in X$, we define

- the *ratio average* payoff

$$\hat{J}(x,\pi,\gamma) = \liminf_{n\to\infty} \frac{E_x^{\pi\gamma}(\sum_{k=1}^n u(x_k,a_k,b_k))}{E_x^{\pi\gamma}(\sum_{k=1}^n \tau(x_k,a_k,b_k))}; \tag{5.24}$$

- the *time average* payoff

$$\hat{j}(x,\pi,\gamma) = \liminf_{t\to\infty} \frac{E_x^{\pi\gamma}(\sum_{n=1}^{N(t)} u(x_n,a_n,b_n))}{t}, \tag{5.25}$$

where $E_x^{\pi\gamma}$ is the expectation operator corresponding to the unique measure $P_x^{\pi\gamma}$ defined on the space of all histories of the process starting at x and induced by q, H, and strategies $\pi \in \Pi$ and $\gamma \in \Gamma$.

Remark 6. (a) The definition of average reward in (5.25) is more natural for semi-Markov games, since it takes into account continuous nature of such processes. Formally, the time average payoff should be defined as follows:

$$\hat{j}(x,\pi,\gamma)=\liminf_{t\to\infty}\frac{E_x^{\pi\gamma}(\sum_{n=1}^{N(t)}u(x_n,a_n,b_n)+(T_{N(t)+1}-t)u_2(x_{N(t)},a_{N(t)},b_{N(t)})}{t}.$$

However, from Remark 3.1 in Jaśkiewicz (2009), it follows that the assumptions imposed on the game model with the time average payoff imply that

$$\lim_{t\to\infty}\frac{E_x^{\pi\gamma}(T_{N(t)+1}-t)u_2(x_{N(t)},a_{N(t)},b_{N(t)})}{t}=0.$$

Finally, it is worth emphasizing that the payoff defined in (5.25) requires additional tools and methods for the study (such as renewal theory, martingale theory, and analysis of the underlying process to the so-called small set) than the model with average payoff (5.24).

(b) It is worth mentioning that payoff criteria (5.24) and (5.25) need not coincide even for stationary policies and may lead to different optimal policies. Such situations happen if the Markov chain induced by stationary strategies is not ergodic (see Feinberg 1994).

We shall need the following continuity-compactness, ergodicity, and regularity assumptions.

(C10) The set-valued mappings $x \to A(x)$ and $x \to B(x)$ are continuous; moreover, $A(x)$ and $B(x)$ are compact for each $x \in X$.

(C11) The functions u and τ are continuous on K, and there exist a positive constant d and continuous function $\omega : X \to [1,\infty)$ such that

$$\tau(x,a,b) \le d\omega(x), \qquad |u(x,a,b)| \le d\omega(x),$$

for all $(x,a,b) \in K$.

(C12) The function $(x,a,b) \to \int_X \omega(y)q(dy|x,a,b)$ is continuous.

(GE1) There exists a Borel set $C \subset X$ such that for some $\hat{\lambda} \in (0,1)$ and $\eta > 0$, we have

$$\int_X \omega(y)q(dy|x,a,b) \le \hat{\lambda}\omega(x) + \eta 1_C(x),$$

for each $(x,a,b) \in K$, with ω introduced in (C11).

(GE2) The function ω is bounded on C, that is,

$$\omega_C := \sup_{x\in C}\omega(x) < \infty.$$

(GE3) There exist some $\delta \in (0,1)$ and a probability measure on C with the property that

$$q(D|x,a,b) \ge \delta\mu(D),$$

for each Borel set $D \subset C$, $x \in C$, $a \in A(x)$, and $b \in B(x)$.

(R1) There exist $\kappa > 0$ and $\xi < 1$ such that

$$H(\kappa|x, a, b) \leq \xi,$$

for all $x \in C$, $a \in A(x)$ and $b \in B(x)$. Moreover, $\tau(x, a, b) \leq d$ for all $(x, a, b) \in K$.

(R2) There exists a decreasing function α such that $\alpha(0) \leq d$, $\alpha(\infty) = 0$ and

$$\int_t^{\infty} sH(ds|x, a, b) \leq \alpha(t)$$

for all $(x, a, b) \in K$. Moreover, $\lim_{t\to\infty} \sup_{x\in C} \sup_{a\in A(x), b\in B(x)}[1 - H(t|x, a, b)] = 0$.

(C13) There exists an open set $\widetilde{C} \subset C$ such that $\mu(\widetilde{C}) > 0$.

For any Borel function $v : X \to \mathbb{R}$, we define the ω-norm as in (5.5). By $\mathcal{B}_\omega(X)$ we denote the set of all Borel measurable functions with finite ω-norm.

Remark 7. (a) Assumption (GE3) in the theory of Markov chains implies that the process generated by the stationary strategies of the players and the transition law q is φ-irreducible and aperiodic. The irreducible measure can be defined as follows:

$$\varphi(D) := \delta\mu(D \cap C) \quad \text{for} \quad D \in \mathcal{B}(X).$$

In other words, if $\varphi(D) > 0$, then the probability of reaching the set D is positive, independent of the initial state. The set C is called "small set."

The function ω in (GE1, GE2) up to the multiplicative constant is a bound for the average time of first entry of the process to the set C (Theorem 14.2.2 in Meyn and Tweedie 2009).

Assumptions (GE) imply that the underlying Markov chain $(x_n)_{n\in\mathbb{N}}$ induced by a pair of stationary strategies $(f, g) \in F \times G$ of the players possesses a unique invariant probability measure π_{fg}. Moreover, $(x_n)_{n\in\mathbb{N}}$ is ω-uniformly ergodic (see Meyn and Tweedie 1994), i.e., there exist constants $\theta > 0$ and $\hat{\alpha} < 1$ such that

$$\left| \int_X \phi(y)q(dy|x, f, g) - \int_X \phi(y)\pi_{fg}(dy) \right| \leq \|\phi\|_\omega \theta\omega(x)\hat{\alpha}^n \tag{5.26}$$

for every $\phi \in \mathcal{B}_\omega(X)$ and $x \in X$, $n \geq 1$. Here $q^{(n)}(\cdot|x, f, g)$ denotes the n-step transition probability induced by q, $f \in F$, and $g \in G$. Clearly, for integers $n \geq 2$ and $D \in \mathcal{B}(X)$, we have

$$q^{(n)}(D|x, f, g) := \int_X q^{(n-1)}(D|y, f, g)q(dy|x, f, g)$$

and $q^{(1)}(D|x, f, g) := q(D|x, f, g)$. From (5.26) we conclude that

$$\hat{J}(f, g) := \hat{J}(x, f, g) = \frac{\int_X u(y, f, g)\pi_{fg}(dy)}{\int_X \tau(y, f, g)\pi_{fg}(dy)}, \quad x \in X, \tag{5.27}$$

for every $f \in F$ and $g \in G$, that is, the average payoff is independent of the initial state. Obviously, $\tau(x, f, g) = \tau(x, f(x), g(x))$ (see (5.18)). Consult also Proposition 10.2.5 in Hernández-Lerma and Lasserre (1999) and Theorem 3.6 in Kartashov (1996) for similar type of assumptions that lead to ω-ergodicity of the underlying Markov chains induced by stationary strategies of the players. The reader is also referred to Arapostathis et al. (1993) for an overview of ergodicity assumptions.

(b) Condition (R1) ensures that infinite number of transitions does not occur in a finite time interval when the process is in the set C. Indeed, when the process is outside the set C, then assumption (GE) implies that the process governed by any strategies of the players returns to the set C within a finite number of transitions with probability one. Then, (R1) prevents the process in the set C from the explosion. As an immediate consequence of (R1), we get that $\tau(x,a,b) > \kappa(1-\xi)$ for all $x \in C$ and $(x,a,b) \in K$. Assumption (R2) is a technical assumption used in the proof of the equivalence of the aforementioned two average payoff criteria.

In order to formulate the first result, we replace the function ω by a new one $W(x) := \omega(x) + \frac{\eta}{\delta}$ that satisfies the following inequality:

$$\int_X W(y)q(dy|x,a,b) \le \lambda^* W(x) + \delta 1_C(x) \int_C W(y)\mu(dy),$$

for $(x,a,b) \in K$ and a suitably chosen $\lambda^* \in (0,1)$ (see Lemma 3.2 in Jaśkiewicz 2009). Observe that if we define the subprobability measure $p(\cdot|x,a,b) := q(\cdot|x,a,b) - \delta 1_C(x)\mu(\cdot)$, then

$$\int_X W(y)p(dy|x,a,b) \le \lambda^* W(x).$$

The above inequality plays a crucial role in the application of the fixed point argument in the proof of Theorem 8 given below.

Similarly as in (5.5) we define $\|\cdot\|_W$ and the set $\mathcal{B}_W(X)$. For each average payoff, we define the lower value, upper value, and the value of the game in an obvious way.

The first result summarizes Theorem 4.1 in Jaśkiewicz (2009) and Theorem 1 in Jaśkiewicz (2010).

Theorem 8. *Assume* (C10)–(C13), (GE1)–(GE3), and (W2). *Then, the following hold:*

(a) There exist a constant v *and* $h^* \in \mathcal{B}_W(X)$, *which is continuous and such that*

$$h^*(x) = \mathrm{val}\left[u(x,\cdot,\cdot) - v\tau(x,\cdot,\cdot) + \int_X h^*(y)q(dy|x,\cdot,\cdot)\right] \tag{5.28}$$

$$= \sup_{\nu\in\Pr(A(x))}\inf_{\rho\in\Pr(B(x))}\left[u(x,\nu,\rho) - v\tau(x,\nu,\rho) + \int_X h^*(y)q(dy|x,\nu,\rho)\right]$$

$$= \inf_{\rho\in\Pr(B(x))}\sup_{\nu\in\Pr(A(x))}\left[u(x,\nu,\rho) - v\tau(x,\nu,\rho) + \int_X h^*(y)q(dy|x,\nu,\rho)\right]$$

for all $x \in X$.

(b) The constant v *is the value of the game with the average payoff defined in* (5.24).

(c) There exists a pair $(\hat{f},\hat{g}) \in F \times G$ *such that*

$$h^*(x) = \inf_{\rho\in\Pr(B(x))}\left[u(x,\hat{f}(x),\rho) - v\tau(x,\hat{f}(x),\rho) + \int_X h^*(y)q(dy|x,\hat{f}(x),\rho)\right]$$

$$= \sup_{\nu\in\Pr(A(x))}\left[u(x,\nu,\hat{g}(x)) - v\tau(x,\nu,\hat{g}(x)) + \int_X h^*(y)q(dy|x,\nu,\hat{g}(x))\right]$$

for all $x \in X$. *The stationary strategy* $\hat{f} \in F$ *(*$\hat{g} \in G$*) is optimal for player* 1 *(player* 2*).*

The proof of Theorem 8 owes much to the approach introduced by Vega-Amaya (2003), who used a fixed point argument in the game model with setwise continuous transition probabilities. However, we cannot directly apply a fixed point argument. First, we have to regularize (to smooth in some sense) certain functions. Using this smoothing method, we are able to apply the Banach fixed point theorem in the space of lower semicontinuous functions that are bounded in the W-norm. It is worth mentioning that the contraction operator for any lower semicontinuous function $h : X \to \mathbb{R}$ is of the form

$$(\hat{T}h)(x) := \inf_{\rho\in\Pr(B(x))}\sup_{\nu\in\Pr(A(x))}\Phi^h(x,\nu,\rho),$$

where

$$\Phi^h(\bar{k}) := \liminf_{d(k',\bar{k})\to 0}\left(u(k') - \mathcal{V}\tau(k') + \int_X h(y)p(dy|k')\right),$$

d is a metric on $X \times \Pr(A) \times \Pr(B)$, and

$$\mathcal{V} := \sup_{f\in F}\inf_{g\in G}\hat{J}(f,g)$$

is the lower value (in the class of stationary strategies) of the game with the payoff function defined in (5.24). Next, it is proved that $k \to \Phi^h(k)$ is indeed lower semicontinuous. The definition of the operator $\hat{T}$ is more involved when compared to the one studied by Vega-Amaya (2003), who assumed that the transition law is setwise continuous in actions, i.e., for which the function $(x, a, b) \to q(D|x, a, b)$ is continuous in (a, b) for every set $D \in \mathcal{B}(X)$. Within such a framework he obtained a solution to the optimality equation $h^* \in \mathcal{B}_W(X)$. The operator $\hat{T}$, on the other hand, enables us to get a lower semicontinuous solution to the optimality equation. In order to obtain a continuous solution, we have to repeat this procedure for a game with the payoff $-u$. Then, it is sufficient to show that the obtained lower semicontinuous solution for the game with the payoff $-u$ coincides with the solution to the optimality equation obtained for the original game. Hence, it must be continuous. The optimal strategies and the conclusion that $\mathcal{V} = v$ are deduced immediately from the optimality equation.

The problem of finding optimal strategies for the players in ergodic zero-sum Markov games on a general state space was considered by, among others, Ghosh and Bagchi (1998), who assumed that the transition law q has a majorant, i.e., there exists a probability measure $\hat{\nu}$ such that $q(\cdot|x, a, b) \geq \hat{\nu}(\cdot)$ for all $(x, a, b) \in K$. Then, the solution to the optimality equation is obtained via the Banach fixed point theorem, since due to the aforementioned assumption, one can introduce a contractive operator in the so-called span semi-norm: $\|h\|_{sp} := \sup_{x \in X} h(x) - \inf_{x \in X} h(x)$, where $h : X \to \mathbb{R}$ is a bounded Borel function. Nowak (1994) studied Markov games with state-independent transitions and obtained some optimality inequalities using standard vanishing discount factor approach. Finally, the results of Meyn and Tweedie (1994, 2009) and Kartashov (1996) allowed to study other classes of stochastic (Markov or semi-Markov) games satisfying general ergodicity conditions. These assumptions were used to prove the existence of the game value with the average payoff criteria and the existence of optimal strategies for the players in games with unbounded payoff functions (see Jaśkiewicz 2002; Vega-Amaya 2003 or Jaśkiewicz and Nowak 2006; Küenle 2007, and references cited therein). For instance, the first two papers mentioned above deal with semi-Markov zero-sum games with setwise continuous transition probabilities. The payoffs and transitions in Jaśkiewicz (2002) and Vega-Amaya (2003) need not be continuous with respect to the state variable. Within such a framework, the authors proved that the optimality equation has a solution, there exists a value of the game, and both players possess optimal stationary strategies. However, the proofs in these papers are based on different methods. For instance, Jaśkiewicz (2002) analyzes auxiliary perturbed models, whereas Vega-Amaya (2003) makes use of a fixed point theorem, which directly leads to a solution of the optimality equation. Moreover, neither of these works deals with the time average payoff criterion.

Jaśkiewicz and Nowak (2006) and Küenle (2007), on the other hand, examine Markov games with weakly continuous transition probabilities. Jaśkiewicz and Nowak (2006) proved that such a Markov game has a value and both players have optimal stationary strategies. Their approach relies on applying Fatou's lemma for weakly convergent measures, which in turn leads to the optimality inequalities instead of the optimality equation. Moreover, the proof employs Michael's theorem on a continuous selection. A completely different approach was presented by Küenle (2007). Under slightly weaker assumptions, he introduced certain contraction operators that lead to a parameterized family of functional equations. Making use of some continuity and monotonicity properties of the solutions to these equations (with respect to the parameter), he obtained a lower semicontinuous solution to the optimality equation.

Remark 8. Jaśkiewicz (2009) and Küenle (2007) imposed a weaker version of basic assumption (C10). In particular, they assumed that the payoff function u is lower semicontinuous, $A(x)$ is a complete metric space, and the mapping $x \to A(x)$ is lower semicontinuous, while the correspondence $x \to B(x)$ is upper semicontinuous and $B(x)$ is a compact metric space. Then, it was shown that the game has a value and the second player has an optimal stationary strategy, whereas the first player has an ϵ-optimal stationary strategy for any $\epsilon > 0$.

The next result is concerned with the second payoff criterion.

Theorem 9. *Assume* (C10)–(C13), (GE1)–(GE3), (W2), and (R1)–(R2). *Then, v is the value of the game and the pair of stationary strategies $(\hat{f}, \hat{g})$ is also optimal for the players in the game with the time average payoff defined in* (5.25).

The proof of Theorem 9 requires different methods than the proof of Theorem 8 and was formulated as Theorem 5.1 in Jaśkiewicz (2009). The point of departure of its proof is the optimality equation (5.28). It allows to define a certain martingale or a super- (sub-) martingale, to which the optional sampling theorem is applied. Use of this result requires an analysis of returns of the process to the small set C and certain consequences of ω-geometric ergodicity as well as some facts from the renewal theory. Theorem 5.1 in Jaśkiewicz (2009) refers to the result in Jaśkiewicz (2004) on the equivalence of the expected time and ratio average payoff criteria for semi-Markov control processes with setwise continuous transition probabilities. Some adaptation to the weakly continuous transition probability case is needed. Moreover, the conclusion of Lemma 7 in Jaśkiewicz (2004) that is also used in the proof of Theorem 9 requires an additional assumption as (R2) given above.

The third result deals with the sample path optimality. For any pair of strategies $(\pi, \gamma) \in \Pi \times \Gamma$ and an initial state $x \in X$, we define three payoffs:

- the sample path ratio average payoff (I)

$$\hat{J}^1(x, \pi, \gamma) = \liminf_{n \to \infty} \frac{\sum_{k=1}^{n} u(x_k, a_k, b_k)}{T_n}; \tag{5.29}$$

- the sample path ratio average payoff (II)

$$\hat{J}^2(x,\pi,\gamma) = \liminf_{n\to\infty} \frac{\sum_{k=1}^{n} u(x_k,a_k,b_k)}{\sum_{k=1}^{n} \tau(x_k,a_k,b_k)}; \tag{5.30}$$

- the sample path time average payoff

$$\hat{j}(x,\pi,\gamma) = \liminf_{t\to\infty} \frac{\sum_{n=1}^{N(t)} u(x_n,a_n,b_n)}{t}. \tag{5.31}$$

A pair of strategies $(\pi^*,\gamma^*) \in \Pi \times \Gamma$ is said to be *sample path optimal* with respect to (5.29), if there exists a function $v_1 \in \mathcal{B}_\omega(X)$ such that for all $x \in X$ it holds

$$\hat{J}^1(x,\pi^*,\gamma^*) = v_1(x) \quad P_x^{\pi^*\gamma^*}\ a.s.$$

$$\text{for every} \quad \gamma \in \Gamma \quad \hat{J}^1(x,\pi^*,\gamma) \geq v_1(x) \quad P_x^{\pi^*\gamma}\ a.s.$$

$$\text{for every} \quad \pi \in \Pi \quad \hat{J}^1(x,\pi,\gamma^*) \leq v_1(x) \quad P_x^{\pi\gamma^*}\ a.s.$$

Analogously, we define sample path optimality with respect to (5.30) and (5.31). In order to prove sample path optimality, we need additional assumptions.

(C14) There exist positive constants d_1, d_2, and $p \in [1,2)$ such that

$$d_2 \leq \tau(x,a,b)^p \leq d_1\omega(x), \quad \text{and} \quad |u(x,a,b)|^p \leq d_1\omega(x),$$

for all $(x,a,b) \in K$.

(C15) If we introduce

$$\hat{\eta}(x,a,b) = \int_0^\infty t^p H(dt|x,a,b),$$

where the constant p is introduced in (C14) and $(x,a,b) \in K$, then there exists a constant $d_3 > 0$ such that

$$\hat{\eta}(x,a,b) \leq d_3\omega(x), \quad (x,a,b) \in K.$$

The following result states that the sample path average payoff criteria coincide. The result was proved by Vega-Amaya and Luque-Vásquez (2000) (see Theorems 3.7 and 3.8). for semi-Markov control processes (one-player games).

Theorem 10. *Assume* (C10)–(C15), (W2), and (GE1)–(GE2). *Then, the pair of optimal strategies* $(\bar{f},\bar{g}) \in F \times G$ *from Theorem* 8 *is sample path optimal with respect to each of the payoffs in* (5.29)*,* (5.30)*, and* (5.31)*. Moreover,* $\hat{J}^1(x,\bar{f},\bar{g}) = \hat{J}^2(x,\bar{f},\bar{g}) = \hat{j}(x,\bar{f},\bar{g}) = v$.

The point of departure in the proof of Theorem 10 is the optimality equation from Theorem 8. Namely, from (5.28) we get two inequalities. The first one is obtained with the optimal stationary strategy $\bar{f}$ for player 1, whereas the second one is connected with the optimal stationary strategy $\bar{g}$ for player 2. Then, the proofs proceed as in Vega-Amaya and Luque-Vásquez (2000) and make use of strong law of large numbers for Markov chains and for martingales (see Hall and Heyde 1980).

6 Stochastic Games with Borel Payoffs

Consider a game $\mathcal{G}$ with countable state space X, finite action spaces, and the transition law q. Let $r : H_\infty \to \mathbb{R}$ be a bounded Borel measurable *payoff function* defined on the set H_∞ of all plays $(x_t, a_t, b_t)_{t\in\mathbb{N}}$ endowed with the product topology and the Borel σ-algebra. (X, A, and B are given the discrete topology.) For any initial state $x = x_1$ and each pair of strategies (π, γ), the *expected payoff* is

$$R(x, \pi, \gamma) := E_x^{\pi\gamma} r(x_1, a_1, b_1, x_2, a_2, b_2, \ldots).$$

If X is a singleton, then $\mathcal{G}$ is called the Blackwell game (see Martin 1998). Blackwell (1969, 1989) proved the following result:

Theorem 11. *The game $\mathcal{G}$ has a value if $r = 1_Z$ is the indicator function of a G_δ-set $Z \subset H_\infty$.*

Martin (1998) proved the following remarkable result:

Theorem 12. *The Blackwell game $\mathcal{G}$ has a value for any bounded Borel measurable payoff function $r : H_\infty \to \mathbb{R}$.*

Maitra and Sudderth (2003b) noted that Theorem 12 can be extended easily to stochastic games with countable set of states X. It is interesting that the proof of the above result is in some part based on the theorem of Martin (1975, 1985) on the determinacy of infinite Borel games with perfect information extending the classical work of Gale and Steward (1953) on clopen games. A further discussion of games with perfect information can be found in Mycielski (1992). An extension to games with delayed information was studied by Shmaya (2011). Theorem 12 was extended by Maitra and Sudderth (1998) in a finitely additive measure setting to a pretty large class of stochastic games with arbitrary state and action spaces endowed with the discrete topology and the history space H_∞ equipped with the product topology. The payoff function r in their approach is Borel measurable. Since Fubini's theorem is not true for finite additive measures, the integration order is fixed in the model. The proof of Maitra and Sudderth (1998) is based on some considerations described in Maitra and Sudderth (1993b) and basic ideas of Martin (1998).

As shown in Maitra and Sudderth (1992), Blackwell G_δ-games (as in Theorem 11) belong to a class of games where the payoff function $r = \limsup_{n\to\infty} r_n$ and r_n depends on finite histories of play. Clearly, the limsup payoffs include the discounted ones. A "partial history trick" on page 181 in Maitra and Sudderth (1996) or page 358 in Maitra and Sudderth (2003a) can be used to show that the limsup payoffs also generalize the usual limiting average ones. Using the operator approach of Blackwell (1989) and some ideas from gambling theory developed in Dubins and Savage (2014) and Dubins et al. (1989), Maitra and Sudderth (1992) showed that every stochastic game with the limsup payoff, countable state, and action spaces has a value. The approach is algorithmic in some sense and was extended to a Borel space framework by Maitra and Sudderth (1993a), where some measurability issues were resolved by using the minmax measurable selection theorem from Nowak (1985a) and some methods from the theory of inductive definability. The authors first studied "leavable games," where player 1 can use a stop rule. Then, they considered approximation of a non-leavable game by leavable ones. The limsup payoffs are Borel measurable, but the methods used in Martin (1998) and Maitra and Sudderth (1998) are not suitable for the countably additive games considered in Maitra and Sudderth (1993a). On the other hand, the proof given in Maitra and Sudderth (1998) has no algorithmic aspect compared with Maitra and Sudderth (1993a). As mentioned above the class of games with the limsup payoffs includes the games with the average payoffs defined as follows: Let X, A, and B be Borel spaces and let $u : X \times A \times B \to \mathbb{R}$ be a *bounded* Borel measurable stage payoff function defined on the Borel set K. Assume that the players are allowed to use *universally measurable strategies*. For any initial state $x = x_1$ and each strategy pair (π, γ), the expected limsup payoff is

$$R(x, \pi, \gamma) := E_x^{\pi\gamma} \left(\limsup_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} u(x_k, a_k, b_k) \right). \tag{5.32}$$

By a minor modification of the proof of Theorem 1.1 in Maitra and Sudderth (1993a) together with the "partial history trick" mentioned above, one can conclude the following result:

Theorem 13. *Assume that X, A, and B are Borel spaces, $K_A \in \mathcal{B}(X \times A)$, $K_B \in \mathcal{B}(X \times B)$, and the set $B(x)$ is compact for each $x \in X$. If $u : K \to \mathbb{R}$ is bounded Borel measurable, $u(x, a, \cdot)$ is lower semicontinuous and $q(D|x, a, \cdot)$ is continuous on $B(x)$ for all $(x, a) \in K_A$ and $D \in \mathcal{B}(X)$, then the game with the expected limiting average payoff defined in (5.32) has a value and for any $\varepsilon > 0$ both players have ε-optimal universally measurable strategies.*

The methods of gambling theory were also used to study "games of survival" of Milnor and Shapley (1957) (see Theorem 16.4 in Maitra and Sudderth 1996). As defined by Everett (1957) a *recursive game* is a stochastic game, where the payoff is zero in every state from which the game can move after some choice of actions to a different state. Secchi (1997, 1998) gave conditions for recursive games with countably many states and finite action sets under which the value exists and the players have stationary ε-optimal strategies. He used techniques from gambling theory.

The lower semicontinuous payoffs $r : H_\infty \to \mathbb{R}$ used in Nowak (1986) are of the limsup type. However, Theorem 4.2 on the existence of value in a semicontinuous game established in Nowak (1986) is not a special case of the aforementioned works of Maitra and Sudderth. The reason is that the transition law in Nowak (1986) is *weakly continuous*. If r is bounded and continuous and the action correspondences are compact valued and continuous, then Theorem 4.2 in Nowak (1986) implies that both players have "persistently optimal strategies." This notion comes from gambling theory (see Kertz and Nachman 1979). A pair of persistently optimal strategies forms a sub-game perfect equilibrium in the sense of Selten (1975).

We close this section with a famous example of Gillette (1957) called the Big Match.

Example 6. Let $X = \{0, 1, 2\}$, $A(x) = A = \{0, 1\}$, and $B(x) = B = \{0, 1\}$. The state $x = 0$ is absorbing with zero payoffs and $x = 2$ is absorbing with payoffs 1. The game starts in state $x = 1$. As long as player 1 picks 0, she gets one unit on each stage that player 2 picks 0 and gets nothing on stages when player 2 chooses 1. If player 1 plays 0 forever, then she gets

$$\limsup_{n\to\infty} \frac{r_1 + \cdots + r_n}{n},$$

where r_k is the number of units obtained on stage $k \in \mathbb{N}$. However, if player 1 picks 1 on some stage (goes to "Big Match") and the choice of player 2 is also 1, then the game moves to the absorbing state 2 and she will get 1 from this stage on. If player 1 picks 1 on some stage and the choice of player 2 is 0, then the game moves to the absorbing state 0 and all future payoffs will be zero. The definition of the transition probability is obvious. Blackwell and Ferguson (1968) proved the following: The Big Match has no value in the class of stationary strategies. However, if the players know the entire history at every stage of the game, then the game has a value in general classes of strategies. Player 2 has a stationary optimal strategy (toss a coin in state $x = 1$), and for any $\varepsilon > 0$ player 1 has an ε-optimal strategy. The value of the game in state 1 is $1/2$. An important feature of this example (that belongs to the class of games studied by Maitra and Sudderth 1992) is that player 1 must remember the entire history of the game at every moment of play. Blackwell and Ferguson (1968) gave two different constructions of an ε-optimal strategy for player 1. One of them relies on using a sequence of optimal stationary strategies in the discounted games with the discount factor tending to one. The idea was to switch from one discounted optimal strategy to another on the basis of some statistics defined on the past plays. This concept was used by Mertens and Neyman (1981) in their fundamental work on stochastic games with average payoffs. The Big Match was generalized by Kohlberg (1974), who considered finite state and finite action games in which all states but one are absorbing. Useful comments on the Big Match can be found in Mertens (2002) or Solan (2009).

7 Asymptotic Analysis and the Uniform Value

In this section, we briefly review some results found in the literature in terms of "normalized discounted payoffs." Let $x = x_1 \in X$, $\pi \in \Pi$, and $\gamma \in \Gamma$. The normalized discounted payoff is of the form

$$J_\lambda(x,\pi,\gamma) := E_x^{\pi\gamma}\left(\lambda\sum_{n=1}^{\infty}(1-\lambda)^{n-1}u(x_n,a_n,b_n)\right).$$

The discount factor is $\beta = 1-\lambda$ where $\lambda \in (0,1)$. Clearly $J_\lambda(x,\pi,\gamma) = (1-\beta) J_\beta(x,\pi,\gamma)$. If the value $w_\lambda(x)$ exists for the normalized game for an initial state $x \in X$, then $w_\lambda(x) = (1-\beta)v_\beta(x)$. By $v_n(x)$ we denote the value function of the n-stage game with the payoff function:

$$\overline{J}_n(x,\pi,\gamma) := E_x^{\pi\gamma}\left(\frac{\sum_{k=1}^{n} u(x_k,a_k,b_k)}{n}\right).$$

A function $v_\infty : X \to \mathbb{R}$ is called a *uniform value* for the stochastic game if for any $\epsilon > 0$, there exist a pair of strategies $(\pi^\epsilon,\gamma^\epsilon) \in \Pi\times\Gamma$, some $n_0 \in \mathbb{N}$ and $\lambda_0 \in (0,1)$ such that for all $n \geq n_0$ and $x \in X$,

$$\sup_{\pi\in\Pi}\overline{J}_n(x,\pi,\gamma^\epsilon) - \epsilon \leq v_\infty(x) \leq \inf_{\gamma\in\Gamma}\overline{J}_n(x,\pi^\epsilon,\gamma)+\epsilon \tag{5.33}$$

and for all $\lambda \in (0,\lambda_0)$ and $x \in X$,

$$\sup_{\pi\in\Pi}J_\lambda(x,\pi,\gamma^\epsilon) - \epsilon \leq v_\infty(x) \leq \inf_{\gamma\in\Gamma}J_\lambda(x,\pi^\epsilon,\gamma)+\epsilon. \tag{5.34}$$

If v_∞ exists, then from (5.33) and (5.34), it follows that $v_\infty(x) = \lim_{n\to\infty} v_n(x) = \lim_{\lambda\to 0+} w_\lambda(x)$. Moreover, $(\pi^\epsilon,\gamma^\epsilon)$ is a pair of nearly optimal strategies in all sufficiently long finite games as well as in all discounted games with the discount factor β (or λ) sufficiently close to one (zero).

Mertens and Neyman (1981) gave sufficient conditions for the existence of v_∞ for arbitrary state space games. For a proof of the following result, see Mertens and Neyman (1981) or Chap. VII in Mertens et al. (2015).

Theorem 14. *Assume that*

– *the payoff function u is bounded,*
– *for any $\lambda \in (0,1)$, w_λ exists, and both players have ε-optimal stationary strategies,*
– *for any $\alpha < 1$, there exists a sequence $(\lambda_i)_{i\in\mathbb{N}}$ such that $0 < \lambda_i < 1$, $\lambda_{i+1} \geq \alpha\lambda_i$ for all $i \in \mathbb{N}$, $\lim_{i\to\infty}\lambda_i = 0$ and*

$$\sum_{i=1}^{\infty} \sup_{x \in X} |w_{\lambda_i}(x) - w_{\lambda_{i+1}}(x)| < \infty.$$

Then, the uniform value v_∞ exists. Moreover, if $x = x_1$ is an initial state and

$$\overline{U}_n(h_n, a_n, b_n) = \frac{u(x_1, a_1, b_1) + \cdots + u(x_n, a_n, b_n)}{n},$$

then we have

$$\sup_{\pi \in \Pi} E_x^{\pi \gamma^\epsilon} \left(\limsup_{n \to \infty} \overline{U}_n(h_n, a_n, b_n) \right) - \epsilon \le v_\infty(x) \tag{5.35}$$
$$\le \inf_{\gamma \in \Gamma} E_x^{\pi^\epsilon \gamma} \left(\liminf_{n \to \infty} \overline{U}_n(h_n, a_n, b_n) \right) + \epsilon.$$

Mertens and Neyman (1981) proved additionally that w_λ and v_n converge to v_∞ uniformly on X. It is worth emphasizing that their ϵ-optimal strategy has a simple intuition behind it. Namely, at every step, the strategy updates a fictitious discount factor and plays an optimal strategy for that fictitious parameter. This parameter summarizes past play and its updating is based on payoffs received in the previous steps. If payoffs received so far are high, the player places higher weight on the future and increases his patience by letting the fictitious discount factor get closer to one. If, on the other hand, payoffs received so far are low, he focuses more about short-term payoffs and therefore decreases this fictitious discount factor. The construction idea of such a strategy lies in the fine-tuning and hinges on algebraic properties of the value of the discounted game as a function of the discount factor (see Bewley and Kohlberg 1976a). For a detailed discussion of the assumptions made in Theorem 14, consult Mertens (2002) and Mertens et al. (2015). It should be noted that neither the existence of uniform value nor (5.35) follows from the general minmax theorems of Maitra and Sudderth (1992, 1993a).

Assume that X, A, and B are finite. Bewley and Kohlberg (1976a,b) proved that the limits $\lim_{\lambda \to 0^+} w_\lambda(x)$ and $\lim_{n \to \infty} v_n(x)$ exist and have a common value $v(x)$, called the asymptotic value. Using their results, Mertens and Neyman (1982) proved that $v(x)$ is actually the uniform value $v_\infty(x)$. Independent of this result, it is possible to show using Bewley and Kohlberg (1976a) that the assumptions of Theorem 14 hold for games with a finite state space and finite action sets (see Remark VII.3.2 in Mertens et al. 2015). Bewley and Kohlberg (1976a) actually proved more, i.e., $w_\lambda(x)$ has in the neighborhood of zero the Puiseux series expansion. More precisely, there exist $\lambda' \in (0, 1)$, $M \in \mathbb{N}$, and numbers $a_i(x)$ $(i = 0, 1, \ldots)$ (depending on $x \in X$) such that for all $\lambda \in (0, \lambda')$, we have

$$w_\lambda(x) = \sum_{i=0}^{\infty} a_i(x) \lambda^{i/M}. \tag{5.36}$$

Recently, Oliu-Barton (2014) gave a direct proof of the existence of $\lim_{\lambda\to 0^+} w_\lambda$. His proof does not utilize the Tarski-Seidenberg elimination from real algebraic geometry as in Bewley and Kohlberg (1976a). (An excellent introduction to semi-algebraic functions and their usage in finite state and action stochastic games can be found in Neyman 2003a.) Moreover, based upon the explicit description of asymptotically optimal strategies, Oliu-Barton (2014) showed that his approach can also be used to obtain the uniform value as in Mertens and Neyman (1981). Further generalization of the abovementioned results to other stochastic games was provided by Ziliotto (2016).

A similar Puiseux expansion can be obtained for stationary optimal strategies in discounted games. Mertens (1982, 2002) showed how to get (5.36) for normalized discounted payoffs in finite nonzero-sum games. Different proofs of (5.36) are given in Milman (2002), Szczechla et al. (1997), and Neyman (2003a). It is also worth mentioning that the values v_n of finite stage games can be approximated by also some series of expansions. Bewley and Kohlberg (1976b) proved that there exist $M \in \mathbb{N}$ and real numbers $b_i(x)$ $(i = 0, 1, 2 \ldots)$ such that for n sufficiently large we have

$$\left| v_n(x) - \sum_{i=0}^{\infty} b_i(x) n^{-i/M} \right| = O(\ln n/n) \tag{5.37}$$

and the bound in (5.37) is tight. A result on a uniform polynomial convergence rate of the values v_n to v_∞ is given in Milman (2002). The results on the values w_λ described above generalize the paper of Blackwell (1962) on dynamic programming (one-person games), where it was shown that the normalized value is a bounded and rational function of the discount factor.

The Puiseux series expansions can also be used to characterize average payoff games, in which the players have optimal stationary strategies (see Bewley and Kohlberg 1978, Chap. 8 in Vrieze 1987 or Filar and Vrieze 1997). For example, one can prove that the average payoff game has a constant value v_0 and both players have optimal stationary strategies if and only if $a_0(x) = v_0$ and $a_1(x) = \cdots = a_{M-1}(x) = 0$ in (5.36) for all $x \in X$(see, e.g.,Theorem 5.3.3 in Filar and Vrieze 1997).

We recall that a stochastic game is *absorbing* if all states but one are absorbing. A recursive or an absorbing game is called continuous if the action sets are compact metric, the state space is countable, and the payoffs and transition probabilities depend continuously on actions. Mertens and Neyman (1981) gave sufficient conditions for $\lim_{\lambda\to 0^+} w_\lambda = \lim_{n\to\infty} v_n$ to hold that include the finite case as well as a more general situation, e.g., when the function $\lambda \to w_\lambda$ is of bounded variation or satisfies some integrability condition (see also Remark 2 in Mertens 2002 and Laraki and Sorin 2015). However, their conditions are not known to hold in continuous absorbing or recursive games. Rosenberg and Sorin (2001) studied the asymptotic properties of w_λ and v_n using some non-expansive operators called Shapley operators, naturally connected with stochastic games (see also Kohlberg

1974; Neyman 2003b; Sorin 2004). They obtained results implying that equality $\lim_{\lambda\to 0^+} w_\lambda = \lim_{n\to\infty} v_n$ holds for continuous absorbing games with finite state spaces. Their result was used by Mertens et al. (2009) to show that every game in this class has a uniform value (consult also Sect. 3 in Ziliotto 2016).

Recursive games were introduced by Everett (1957), who proved the existence of value and of stationary ε-optimal strategies, when the state space and action sets are finite. Recently, Li and Venel (2016) proved that recursive games on a countable state space with finite action spaces have the uniform value, if the family $\{v_n\}$ is totally bounded. Their proofs follow the same idea as in Solan and Vieille (2002). Moreover, the result in Li and Venel (2016) together with the ones in Rosenberg and Vieille (2000) provides the uniform Tauberian theorem for recursive games: (v_n) converges uniformly if and only if (v_λ) converges uniformly and both limits are the same. For finite state continuous recursive games, the existence of $\lim_{\lambda\to 0^+} w_\lambda$ was recently proved by Sorin and Vigeral (2015a).

We also mention one more class of stochastic games, the so-called definable games, studied by Bolte et al. (2015). Such games involve a finite number of states, and it is additionally assumed that all their data (action sets, payoffs, and transition probabilities) are definable in an o-minimal structure. Bolte et al. (2015) proved that these games have the uniform value. The reason for that lies in the fact that definability allows to avoid highly oscillatory phenomena in various settings (partial differential equations, control theory, continuous optimization) (see Bolte et al. 2015 and the references cited therein).

Generally, the asymptotic value $\lim_{\lambda\to 0^+} w_\lambda$ or $\lim_{n\to\infty} v_n$ may not exist for stochastic games with *finitely many states*. An example with four states (two of them being absorbing) and compact action sets was recently given by Vigeral (2013). Moreover, there are problems with asymptotic theory in stochastic games with finite state space and countable action sets (see Ziliotto 2016). In particular, the example given in Ziliotto (2016) contradicts the famous hypothesis formulated by Mertens (1987) on the existence of asymptotic value. A generalization of examples due to Vigeral (2013) and Ziliotto (2016) is presented in Sorin and Vigeral (2015b).

A new approach to the asymptotic value in games with finite state and action sets was recently given by Oliu-Barton (2014). His proof when compared to Bewley and Kohlberg (1976a) is direct, relatively short, and more elementary. It is based on the theory of finite-dimensional systems and the theory of finite Markov chains. The existence of uniform value is obtained without using algebraic tools. A simpler proof for the existence of the asymptotic value $\lim_{\lambda\to 0} w_\lambda$ of finite λ-discounted absorbing games was provided by Laraki (2010), who obtained explicit formulas for this value. According to the author's comments, certain extensions to absorbing games with finite state and compact action spaces are also possible, but under some continuity assumptions on the payoff function. The convergence of the values of n-stage games (as $n \to \infty$) and the existence of the uniform value in stochastic games with a general state space and finite action spaces were studied by Venel (2015) who assumed that the transition law is in certain sense commutative with respect to the actions played at two consecutive periods. Absorbing games can be reformulated as commutative stochastic games.

8 Algorithms for Zero-Sum Stochastic Games

Let $P = [p_{ij}]$ be a payoff matrix in a zero-sum game where $1 \le i \le m_1$, $1 \le j \le m_2$. By $valP$ we denote the value for this game in mixed strategies. We assume in this section that X, A, and B are finite sets. For any function $\phi : X \to \mathbb{R}$, we can consider the zero-sum game $\Gamma_\phi(x)$ where the payoff matrix is

$$P_\phi(x) := \left[\lambda u(x,i,j) + (1-\lambda) \sum_{y \in X} \phi(y) q(y|x,i,j) \right], \quad x \in X.$$

Recall that $\beta = 1 - \lambda$. Similar to (5.20) we define $T_\lambda \phi(x)$ as the value of the game $\Gamma_\phi(x)$, i.e., $T_\lambda \phi(x) = valP_\phi(x)$. If $\phi(x) = \phi_0(x) = 0$ for all $x \in X$, then $T_\lambda^n \phi_0(x)$ is the value of the n-stage discounted stochastic game starting at the state $x \in X$. As we know from Shapley (1953), the value function w_λ of the normalized discounted game is a unique solution to the equation $w_\lambda(x) = T_\lambda w_\lambda(x)$, $x \in X$. Moreover, $w_\lambda(x) = \lim_{n\to\infty} T_\lambda^n \phi_0(x)$. The procedure of computing $T_\lambda^n \phi_0(x)$ is known as the *value iteration* and can be used as an algorithm to approximate the value function w_λ. However, this algorithm is rather slow. If $f^*(x)$ ($g^*(x)$) is an optimal mixed strategy for player 1 (player 2) in game $\Gamma_{w_\lambda}(x)$, then the functions f^* and g^* are stationary optimal strategies for the players in the infinite horizon discounted game.

Example 7. Let $X = \{1,2\}$, $A(x) = B(x) = \{1,2\}$ for $x \in X$. Assume that state $x = 2$ is absorbing with zero payoffs. In state $x = 1$, we have $u(1,1,1) = 2$, $u(1,2,2) = 6$, and $u(1,i,j) = 0$ for $i \ne j$. Further, we have $q(1|1,1,1) = q(1|1,2,2) = 1$ and $q(2|1,i,j) = 1$ for $i \ne j$. If $\lambda = 1/2$, then the Shapley equation is for $x = 1$ of the form

$$w_\lambda(1) = val \begin{bmatrix} 1 + \frac{1}{2} w_\lambda(1) & 0 + \frac{1}{2} w_\lambda(2) \\ 0 + \frac{1}{2} w_\lambda(2) & 3 + \frac{1}{2} w_\lambda(1) \end{bmatrix}.$$

Clearly, $w_\lambda(2) = 0$ and $w_\lambda(1) \ge 0$. Hence, the above matrix game has no pure saddle point and it is easy to calculate that $w_\lambda(1) = (-4 + 2\sqrt{13})/3$. This example is taken from Parthasarathy and Raghavan (1981) and shows that in general there is no finite step algorithm for solving zero-sum discounted stochastic games.

The value iteration algorithm of Shapley does not utilize any information on optimal strategies in the n-stage games. Hoffman and Karp (1966) proposed a new algorithm involving both payoffs and strategies. Let $g_1(x)$ be an optimal strategy for player 2 in the matrix game $P_{\phi_0}(x)$, $x \in X$. Define $w^1(x) = \sup_{\pi \in \Pi} J_\lambda(x, \pi, g_1)$. Then, choose an optimal strategy $g_2(x)$ for player 2 in the matrix game $P_{w^1}(x)$. Define $w^2(x) = \sup_{\pi \in \Pi} J_\lambda(x, \pi, g_2)$ and continue the procedure. It is shown that $\lim_{n\to\infty} w^n(x) = w_\lambda(x)$.

Let $X = \{1, \ldots, k\}$. Any function $w : X \to \mathbb{R}$ can be viewed as a vector $\bar{w} = (w(1), \ldots, w(k)) \in \mathbb{R}^k$. The fact that w_λ is a unique solution to the Shapley equation is equivalent to saying that the unconstrained optimization problem

$$\min_{\bar{w}\in\mathbb{R}^k} \sum_{x\in X} (T_\lambda w(x) - w(x))^2$$

has a unique global minimum. Pollatschek and Avi-Itzhak (1969) proposed a successive iterations algorithm, which corresponds to the "policy iteration" in dynamic programming. The proposed algorithm is connected with a Newton-Raphson type procedure associated with the global minimum problem mentioned above. Van der Wal (1978) showed that their algorithm does not converge in general. Filar and Tolwinski (1991) presented an improved version of the Pollatschek and Avi-Itzhak algorithm for solving discounted zero-sum stochastic games based on a "modified Newton's method." They demonstrated that it always converges to the value of the stochastic game and solved the example of Van der Wal (1978). For further comments on the abovementioned iterative algorithms, the reader is referred to Vrieze (1987), Breton (1991), Raghavan and Filar (1991), Filar and Vrieze (1997), and Raghavan (2003).

Observe now that every $f \in F$ (also $g \in G$) can be viewed as a vector in Euclidean space. If $f \in F$, then

$$u(x,f,b) = \sum_{a\in A(x)} u(x,a,b)f(a|x) \quad \text{and} \quad q(y|x,f,b) = \sum_{a\in A(x)} q(y|x,a,b)f(a|x).$$

Similarly $u(x,a,g)$ and $q(y|x,a,g)$ are defined for any $g \in G$.

In the remaining part of this section we assume that $u \geq 0$. This condition is made only for simplicity of presentation. A zero-sum discounted stochastic game can also be solved by a constrained nonlinear programming technique studied by Filar et al. (1991) (see also Chap. 3 in Filar and Vrieze 1997). Consider the problem (NP1) defined as follows:

$$\min \sum_{x\in X} (w_1(x) + w_2(x))$$

subject to $(f,g) \in F \times G$, $w_1 \geq 0$, $w_2 \leq 0$ and

$$\lambda u(x,a,g) + (1-\lambda)\sum_{y\in X} w_1(y)q(y|x,a,g) \leq w_1(x), \ \text{for all } x \in X,\ a \in A(x),$$

$$-\lambda u(x,f,b) + (1-\lambda)\sum_{y\in X} w_2(y)q(y|x,f,b) \leq w_2(x), \ \text{for all } x \in X,\ b \in B(x).$$

Note that the objective function is linear, but the constraint set is not convex. It is shown (see Chap. 3 in Filar and Vrieze 1997) that every local minimum of (NP1) is a global minimum. Hence, we have the following result.

Theorem 15. *Let* (w_1^*, w_2^*, f^*, g^*) *be a global minimum of (NP1). Then,* $\sum_{x \in X}(w_1^*(x) + w_2^*(x)) = 0$ *and* $w_1^*(x) = w_\lambda(x)$ *for all* $x \in X$. *Moreover,* (f^*, g^*) *is a pair of stationary optimal strategies for the players in the discounted stochastic game.*

In the case of single-controller stochastic game, in which $q(y|x,a,b)$ is independent of $a \in A(x)$ for each $x \in X$ and denoted by $q(y|x,b)$, the problem of finding optimal strategies for the players is much simpler. We now present a result of Parthasarathy and Raghavan (1981). Consider the following linear programming problem (LP1):

$$\max \sum_{x \in X} w(x)$$

subject to $f \in F,\ w \geq 0$ and

$$\lambda u(x,f,b) + (1-\lambda)\sum_{y \in X} w(y)q(y|x,b) \geq w(x), \text{ for all } x \in X,\ b \in B(x).$$

Note that the constraint set in (LP1) is convex.

Theorem 16. *The problem (LP1) has an optimal solution* (w^*, f^*). *Moreover,* $w^*(x) = w_\lambda(x)$ *for all* $x \in X$, *and* f^* *is an optimal stationary strategy for player 1 in the single-controller discounted stochastic game.*

Remark 9. Knowing w_λ one can find an optimal stationary strategy g^* for player 2 using the Shapley equation $w_\lambda = T_\lambda w_\lambda$, i.e., $g^*(x)$ can be any optimal strategy in the matrix game with the payoff function:

$$\lambda u(x,a,b) + (1-\lambda)\sum_{y \in X} w_\lambda(y)q(y|x,b), \quad a \in A(x),\ b \in B(x).$$

Let $X = X_1 \cup X_2$ and $X_1 \cap X_2 = \emptyset$. Assume that $q(y|x,a,b) = q_1(y|x,a)$ for $x \in X_1$ and $q(y|x,a,b) = q_2(y|x,b)$ for $x \in X_2$, $a \in A(x)$, $b \in B(x)$, $y \in X$. Then the game is called a *switching control stochastic game* (SCSG for short). Filar (1981) studied this class of games with discounting and showed the order field property saying that a solution to the game can be found in the same algebraic field as the data of the game. Other classes of stochastic games having the order field property are described in Raghavan (2003). It is interesting that the value function w_λ for the SCSG can be represented in a neighborhood of zero by the power series of λ (see Theorem 6.3.5 in Filar and Vrieze 1997) . It should be mentioned that every discounted SCSG can be solved by a finite sequence of linear programming problems (see Algorithm 3.2.1 in Filar and Vrieze 1997). This was first shown by Vrieze (1987).

We can now turn to the limiting average payoff stochastic games. We know from the Big Match example of Blackwell and Ferguson (1968) that ε-optimal stationary strategies may not exist. A characterization of limiting average payoff games, where the players have stationary optimal strategies, was given by Vrieze (1987) (see also Theorem 5.3.5 in Filar and Vrieze 1997). Below we state this result. For any function $\phi : X \to \mathbb{R}$ we consider the zero-sum game $\Gamma_\phi^0(x)$ with the payoff matrix

$$P_\phi^0(x) := \left[\sum_{y \in X} \phi(y) q(y|x,i,j)\right], \quad x \in X$$

and the zero-sum game $\Gamma_\phi^1(x)$ with the payoff matrix

$$\tilde{P}_\phi(x) := \left[u(x,i,j) + \sum_{y \in X} \phi(y) q(y|x,i,j)\right], \quad x \in X.$$

Theorem 17. *Consider a function $v^* : X \to \mathbb{R}$ and $f^* \in F$, $g^* \in G$. Then, v^* is the value of the limiting average payoff stochastic game and f^*, g^* are stationary optimal strategies for players 1 and 2, respectively, if and only if for each $x \in X$*

$$v^*(x) = \mathrm{val} P_{v^*}^0(x), \tag{5.38}$$

$(f^*(x), g^*(x))$ is a pair of optimal mixed strategies in the zero-sum game with the payoff matrix $P_{v^*}^0(x)$, and there exist functions $\phi_i : X \to \mathbb{R}$ $(i = 1, 2)$ such that for every $x \in X$, we have

$$v^*(x) + \phi_1(x) = \mathrm{val} \tilde{P}_{\phi_1}(x) = \min_{b \in B(x)} \left[u(x, f^*, b) + \sum_{y \in X} \phi_1(y) q(y|x, f^*, b)\right], \tag{5.39}$$

and

$$v^*(x) + \phi_2(x) = val\, \tilde{P}_{\phi_2}(x) = \max_{a \in A(x)} \left[u(x,a,g^*) + \sum_{y \in X} \phi_2(y) q(y|x,a,g^*)\right]. \tag{5.40}$$

Remark 10. If the Markov chain induced by any stationary strategy pair is irreducible, then v^* is a constant. Then, (5.38) holds trivially and $\phi_1(x)$, $\phi_2(x)$ satisfying (5.39) and (5.40) are such that $\phi_1(x) - \phi_2(x)$ is independent of $x \in X$. In such a case we may take $\phi_1 = \phi_2$. However, in other cases (without irreducibility) $\phi_1(x) - \phi_2(x)$ may depend on $x \in X$. For details the reader is referred to Chap. 8 in Vrieze (1987).

A counterpart to the optimization problem (NP1) with non-convex constraints can also be formulated for the limiting average payoff case. Consider the problem (NP2):

$$\min \sum_{x \in X} (v_1(x) + v_2(x))$$

subject to $(f, g) \in F \times G, v_1 \geq 0,\ v_2 \leq 0, \phi_1 \geq 0,\ \phi_2 \geq 0$ and

$$\sum_{y \in X} v_1(y) q(y|x, a, g) \leq v_1(x),\ \ u(x, a, g) + \sum_{y \in X} \phi_1(y) q(y|x, a, g) \leq v_1(x) + \phi_1(x)$$

for all $x \in X,\ a \in A(x)$ and

$$\sum_{y \in X} v_2(y) q(y|x, f, b) \leq v_2(x),\ \ -u(x, f, b) + \sum_{y \in X} \phi_2(y) q(y|x, f, b) \leq v_2(x) + \phi_2(x)$$

for all $x \in X,\ b \in B(x)$.

Theorem 18. *If $(\phi_1^*, \phi_2^*, v_1^*, v_2^*, f^*, g^*)$ is a feasible solution of (NP2) with the property that $\sum_{x \in X} (v_1(x) + v_2(x)) = 0$, then it is a global minimum and (f^*, g^*) is a pair of optimal stationary strategies. Moreover, $v_1^*(x) = R(x, f^*, g^*)$ (see (5.32)) for all $x \in X$.*

For a proof consult Filar et al. (1991) or pages 127–129 in Filar and Vrieze (1997). Single-controller average payoff stochastic games can also be solved by linear programming. The formulation is more involved than in the discounted case and generalizes the approach known in the theory of Markov decision processes. Two independent studies on this topic are given in Hordijk and Kallenberg (1981) and Vrieze (1981). Similarly as in the discounted case, the SCSG with the average payoff criterion can be solved by a finite sequence of nested linear programs (see Vrieze et al. 1983).

If $X = X_1 \cup X_2$, $X_1 \cap X_2 = \emptyset$, and $A(x)$ ($B(x)$) is a singleton for each $x \in X_1$ ($x \in X_2$), then the stochastic game is of *perfect information*. Raghavan and Syed (2003) gave a policy-improvement type algorithm to find optimal pure stationary strategies for the players in discounted stochastic games of perfect information. Avrachenkov et al. (2012) proposed two algorithms to find the uniformly optimal strategies in discounted games. Such strategies are also optimal in the limiting average payoff stochastic game. Fresh ideas for constructing optimal stationary strategies in zero-sum limiting average payoff games can be found in Boros et al. (2013). In particular, Boros et al. (2013) introduced a potential transformation of the original game to an equivalent canonical form and applied this method to games with additive transitions (AT games) as well as to stochastic games played on a directed graph. The existence of a canonical form was also provided for stochastic games with perfect information, switching control games, or ARAT (additive reward-additive transition) games. Such a potential transformation has an impact on solving some classes of games in sub-exponential time. Additional results can be found in Boros et al. (2016). It is worth to note that a finite step algorithm of Cottle-Dantzig's type was recently applied for solving discounted zero-sum semi-Markov ARAT games by Mondal et al. (2016).

Computation of the uniform value is a difficult task. Chatterjee et al. (2008) provided a finite algorithm for finding the approximation of the uniform value. As mentioned in the previous section, Bewley and Kohlberg (1976a) showed that the function $\lambda \to w_\lambda$ is semi-algebraic. It can be function of λ. It can be expressed as a Taylor series in fractional powers of λ (called Puiseux series) in the neighborhood of zero. By Mertens and Neyman (1981), the uniform value $v(x) = \lim_{\lambda \to 0^+} w_\lambda(x)$. Chatterjee et al. (2008) noted that, for a given $\alpha > 0$, determining whether $v > \alpha$ is equivalent to finding the truth value of a sentence in the theory of real-closed fields. A generalization of the quantifier elimination algorithm of Tarski (1951) due to Basu (1999) (see also Basu et al. 2003) can be used to compute this truth value. The uniform value v is bounded by the maximum payoffs of the game; it is therefore sufficient to repeat this algorithm for finitely many different values of α to get a good approximation of v. An ε-approximation of $v(x)$ at a given state x can be computed in time bounded by an exponential in a polynomial of the size of the game times a polynomial function of $\log(1/\varepsilon)$. This means that the approximating uniform value $v(x)$ lies in the computational complexity class EXPTIME (see Papadimitriou 1994). Solan and Vieille (2010) applied the methods of Chatterjee et al. (2008) to calculate the uniform ε-optimal strategies described by Mertens and Neyman (1981). These strategies are good for all sufficiently long finite horizon games as well as for all (normalized) discounted games with λ sufficiently small. Moreover, they use unbounded memory. As shown by Bewley and Kohlberg (1976a), any pair of stationary optimal strategies in discounted games (which are obviously functions of λ) can also be represented by a Taylor series of fractional powers of λ for $\lambda \in (0, \lambda_0)$ with λ_0 sufficiently small. This result, the theory of real-closed fields, and the methods of formal logic developed in Basu (1999) are basic ideas for Solan and Vieille (2010). A complexity bound on the algorithm of Solan and Vieille (2010) is not determined yet.

9 Zero-Sum Stochastic Games with Incomplete Information or Imperfect Monitoring

The following model of a general two-player zero-sum stochastic game, say $\mathcal{G}$, is described in Sorin (2003a).

- X is a finite state space.
- A and B are finite admissible action sets for players 1 and 2, respectively.
- Ω is a finite state of signals.
- $r : X \times A \times B \to [0, 1]$ is a payoff function to player 1.
- q is a transition probability mapping from $X \times A \times B$ to $\Pr(X \times \Omega)$.

Let p be an initial probability distribution on $X \times \Omega$. The game evolves as follows. At stage one nature chooses (x_1, ω_1) according to p and the players learn ω_1. Then, simultaneously player 1 selects $a_1 \in A$ and player 2 selects $b_1 \in B$. The stage payoff $r(x_1, a_1, b_1)$ is paid by player 2 to player 1 and a pair (x_2, ω_2) is drawn according to $q(\cdot|x_1, a_1, b_1)$. The game proceeds to stage two and the situation is repeated. The standard stochastic game with incomplete information is obtained, when $\Omega = A \times B$. Such a game with finite horizon of play was studied by Krausz and Rieder (1997), who showed the existence of the game value and presented an algorithm to compute optimal strategies for the players via linear programming. Their model assumes incomplete information on one side, i.e., player 2 is never informed about the state of the underlying Markov chain in contrast to player 1. In addition, both players have perfect recall. Renault (2006) studied a similar model. Namely, he assumed that the sequence of states follows a Markov chain, i.e., q is independent of the actions of the players. At the beginning of each stage, only player 1 is informed of the current state, the actions are selected simultaneously, and they are observed by both players. The play proceeds to the next stage. Renault (2006) showed that such a game has a uniform value and the second player has an optimal strategy.

Clearly, if Ω is a singleton, the game is a standard stochastic game. For general stochastic games with incomplete information, little is known, but some classes were studied in the literature. For the Big Match game, Sorin (1984, 1985) and Sorin and Zamir (1991) proved the existence of the maxmin value and the minmax value. These values may be different. Moreover, they showed that the values of the n-stage games (λ-discounted games with normalized payoffs) converge as $n \to \infty$ (as $\lambda \to 0^+$) to the maxmin value.

Another model was considered by Rosenberg et al. (2004). Namely, at the beginning of the game a signal ω is chosen according to $p \in \Pr(\Omega)$. Only player 1 is informed of ω. At stage $n \in \mathbb{N}$ players simultaneously choose actions $a_n \in A$ and $b_n \in B$. The stage payoff $r^\omega(x_n, a_n, b_n)$ is incurred and the next state x_{n+1} is drawn according to $q(\cdot|x_n, a_n, b_n)$. Both players are informed of (a_n, b_n, x_{n+1}). Note that in this setting $r^\omega(x_n, a_n, b_n)$ is told to player 1, but not to player 2. Rosenberg et al. (2004) proved the following result

Theorem 19. *If player 1 controls the transition probability, the game value exists. If player 2 controls the transition probability, both the minmax value and the maxmin value exist.*

Recursive games with incomplete information on one side were studied by Rosenberg and Vieille (2000), who proved that the maxmin value exists and is equal to the limit of the values of n-stage games (λ-discounted games) as $n \to \infty$ (as $\lambda \to 0^+$). Rosenberg (2000), on the other hand, considered absorbing games. She proved the existence of the limit of the values of finitely repeated absorbing games (discounted absorbing games) with incomplete information on one side as the number of repetitions goes to infinity ($\lambda \to 0^+$). Additional discussion on stochastic games with incomplete information on one side can be found in Sorin (2003b) and Laraki and Sorin (2015).

Coulomb (1992, 1999, 2001) was the first who studied stochastic games with imperfect monitoring. These games are played as follows. At every stage, the game is in one of finitely many states. Each player chooses an action, independently of his opponent. The current state, together with the pair of actions, determines a daily payoff, a probability distribution according to which a new state is chosen, and a probability distribution over pairs of signals, one for each player. Each player is then informed of his private signal and of the new state. However, no player is informed of his opponent's signal and of the daily payoff (see also the detailed model in Coulomb 2003a). Coulomb (1992, 1999, 2001) studied the class of absorbing games and proved that the uniform maxmin and minmax values exist. In addition, he provided a formula for both values. One of his main findings is that the maxmin value does not depend on the signaling structure of player 2. Similarly, the minmax value does not depend on the signaling structure of player 1. In general, the maxmin and minmax values do not coincide, hence stochastic games with imperfect monitoring need not have a uniform value. Based on these ideas, Coulomb (2003c) and Rosenberg et al. (2003) independently proved that the uniform maxmin value always exists in a stochastic game, in which each player observes the state and his/her own action. Moreover, the uniform maxmin value is independent of the information structure of player 2. Symmetric results hold for the uniform minmax value.

We now consider the general model of zero-sum dynamic game presented in Mertens et al. (2015) and Coulomb (2003b). These games are known as games of incomplete information on both sides.

- X, A, and B are as above.
- S and T are finite signal spaces for players 1 and 2, respectively.
- The payoff function is defined as above, and the transition probability function is $q : X \times A \times B \to \Pr(X \times S \times T)$.

The evolution of the game is as follows. At stage one nature chooses (x_1, s_1, t_1) according to a given distribution $p \in \Pr(X \times S \times T)$. Player 1 learns s_1 and player 2 is informed of t_1. Then, simultaneously player 1 selects $a_1 \in A$ and player 2 selects $b_1 \in B$. The stage payoff $r(x_1, a_1, b_1)$ is incurred and a new triple (x_2, s_2, t_2) is drawn according to $q(\cdot|x_1, a_1, b_1)$. The game proceeds to stage two and the process repeats. Let us denote this game by $\mathcal{G}_0$. Renault (2012) proved that such a game has a value under an additional condition.

Theorem 20. *Assume that player 1 can always deduce the state and player 2's signal from his own signal. Then, the game $\mathcal{G}_0$ has a uniform value.*

Further examples of games for which Theorem 20 holds were recently provided by Gensbittel et al. (2014). In particular, they showed that if player 1 is more informed than player 2 and controls the evolution of information on the state, then the uniform value exists. This result, from one side, extends results on Markov decision processes with partial observation given by Rosenberg et al. (2002), and, on the other hand, it extends a result on repeated games with an informed controller studied by Renault (2012).

An extension of the repeated game in Renault (2006) to a game with incomplete information on both sides was examined by Gensbittel and Renault (2015). The model is described by two finite action sets A and B and two finite sets of states S and T. The payoff function is $r : S \times T \times A \times B \to [-1, 1]$. There are given two initial probabilities $p_1 \in \Pr(S)$ and $p_2 \in \Pr(T)$ and two transition probability functions $q_1 : S \to \Pr(S)$ and $q_2 : T \to \Pr(T)$. The Markov chains $(s_n)_{n\in\mathbb{N}}$, $(t_n)_{n\in\mathbb{N}}$ are independent. At the beginning of stage $n \in \mathbb{N}$, player 1 observes s_n and player 2 observes t_n. Then, both players simultaneously select actions $a_n \in A$ and $b_n \in B$. Player 1's payoff in stage n is $r(s_n, t_n, a_n, b_n)$. Then, (a_n, b_n) is publicly announced and the play goes to stage $n + 1$. Notice that the payoff $r(s_n, t_n, a_n, b_n)$ is not directly known and cannot be deduced. The main theorem states that $\lim_{n\to\infty} v_n$ exists and is a unique continuous solution to the so-called Mertens-Zamir system of equations (see Mertens et al. 2015). Recently, Sorin and Vigeral (2015a) showed in a simpler model (repeated game model, where s_1 and t_1 are chosen once and they are kept throughout the play) that v_λ converges uniformly as $\lambda \to 0$.

In this section, we should also mention the Mertens conjecture (see Mertens 1987) and its solution. His hypothesis is twofold: the first statement says that in any general model of zero-sum repeated game, the asymptotic value exists, and the second one says that if player 1 is always more informed than player 2 (in the sense that player 2's signal can be deduced from player 1's private signal), then in the long run player 1 is able to guarantee the asymptotic value. Ziliotto (2016) showed that in general the Mertens hypothesis is false. Namely, he constructed an example of a seven-state symmetric information game, in which each player has two action sets. The set of signals is public. The game is played as the game $\mathcal{G}$ described above. More details can be found in Solan and Ziliotto (2016) where related issues are also discussed.

Although the Mertens conjecture does not generally hold, there are some classes of games for which it is true. The interested reader is referred to Sorin (1984, 1985), Rosenberg et al. (2004), Renault (2012), Gensbittel et al. (2014), Rosenberg and Vieille (2000), and Li and Venel (2016). For instance, Li and Venel (2016) dealt with a stochastic game $\mathcal{G}_0$ with incomplete information on both sides and proved the following (see Theorem 5.8 in Li and Venel 2016).

Theorem 21. *Let $\mathcal{G}_0$ be a recursive game such that player 1 is more informed than player 2. Then, for every initial distribution $p \in \Pr(X \times S \times T)$, both the asymptotic value and the uniform maxmin exist and are equal, i.e.,*

$$\underline{v}_\infty = \lim_{n\to\infty} v_n = \lim_{\lambda\to 0} v_\lambda.$$

Different notions of value in two-person zero-sum repeated games were recently examined by Gimbert et al. (2016). Assuming that the state evolves and players receive signals, they showed that the uniform value (limsup value) may not exist. However, the value exists if the payoff function is Borel measurable and the players observe a public signal including the actions played. The existence of the uniform value was proved for recursive games with nonnegative payoffs without any special assumptions on signals.

Stochastic games with partial observations, in which one player observes the sequence of states, while the other player observes the sequence of state-dependent signals, are examined in Basu and Stettner (2015) and its references. A class of dynamic games in which a player is informed of his opponent's actions and states after some time delay were studied by Dubins (1957), Scarf and Shapley (1957), and Levy (2012). For obvious reasons, this survey does not cover all models and cases of games with incomplete information. Further references and applications can be found in Laraki and Sorin (2015), Neyman and Sorin (2003), or Solan and Ziliotto (2016).

10 Approachability in Stochastic Games with Vector Payoffs

In this section, we consider games with payoffs in Euclidean space $\mathbb{R}^k$, where the inner product is denoted by $\langle \cdot, \cdot \rangle$ and the norm of any $\bar{c} \in \mathbb{R}^k$ is $\|\bar{c}\| = \sqrt{\langle \bar{c}, \bar{c} \rangle}$. Let A and B be finite sets of pure strategies for players 1 and 2, respectively. Let $u^0 : A \times B \to \mathbb{R}^k$ be a vector payoff function. For any mixed strategies $s_1 \in \Pr(A)$ and $s_2 \in \Pr(B)$, $\bar{u}^0(s_1, s_2)$ stands for the expected vector payoff. Consider a *two*-person infinitely repeated game G_∞ defined as follows. At each stage $t \in \mathbb{N}$, players 1 and 2 choose simultaneously $a_t \in A$ and $b_t \in B$. Behavioral strategies $\hat{\pi}$ and $\hat{\gamma}$ for the players are defined in the usual way. The corresponding vector outcome is $g_t = u^0(a_t, b_t) \in \mathbb{R}^k$. The couple of actions (a_t, b_t) is announced to both players. The average vector outcome up to stage n is $\bar{g}_n = (g_1 + \cdots + g_n)/n$. The aim of player 1 is to make $\bar{g}_n$ approach a *target set* $C \subset \mathbb{R}^k$. If $k = 1$, then we usually have in mind $C = [v^0, \infty)$ where v^0 is the value of the game in mixed strategies. If $C \subset \mathbb{R}^k$ and $y \in \mathbb{R}^k$, then the distance from y to the set C is $d(y, C) = \inf_{z \in C} \|y - z\|$.

A nonempty closed set $C \subset \mathbb{R}^k$ is *approachable by player* 1 in G_∞ if for every $\epsilon > 0$ there exists a strategy $\hat{\pi}$ of player 1 and $n_\epsilon \in \mathbb{N}$ such that for any strategy $\hat{\gamma}$ of player 2 and any $n \geq n_\epsilon$, we have

$$E^{\hat{\pi}\hat{\gamma}} d(\bar{g}_n, C) \leq \epsilon.$$

The dual concept is excludability.

Let $P_C(y)$ denote the set of closest points to y in C. A closed set $C \subset \mathbb{R}^k$ satisfies the *Blackwell condition* for player 1, if for any $y \notin C$, there exist $z \in P_C(y)$ and a mixed action (depending on y) $s_1 = s_1(y) \in \Pr(A)$ such that the hyperplane through z orthogonal to the line segment $[yz]$ separates y from the set $\{\bar{u}^0(s_1, s_2) : s_2 \in \Pr(B)\}$, i.e.,

$$\langle \bar{u}^0(s_1, s_2) - z, y - z \rangle \leq 0 \quad \text{for all} \quad s_2 \in \Pr(B).$$

The following two results are due to Blackwell (1956).

Theorem 22. *If* $C \subset \mathbb{R}^k$ *is a nonempty closed set satisfying the Blackwell condition, then* C *is approachable in game* G_∞*. An approachability strategy is* $\hat{\pi}(h_n) = s_1(\bar{g}_n)$*, where* h_n *is the history of a play at stage* n*.*

Theorem 23. *A closed and convex set* $C \subset \mathbb{R}^k$ *is either approachable or excludable.*

The next result was proved by Spinat (2002).

Theorem 24. *A closed set* $C \subset \mathbb{R}^k$ *is approachable if and only if* C *contains a subset having the Blackwell property.*

Related results with applications to repeated games can be found in Sorin (2002) and Mertens et al. (2015). Applications to optimization models, learning, and games with partial monitoring can be found in Cesa-Bianchi and Lugosi (2006), Cesa-Bianchi et al. (2006), Perchet (2011a,b), and Lehrer and Solan (2016). A theorem on approachability for stochastic games with vector payoffs was proved by Shimkin and Shwartz (1993). They imposed certain ergodicity conditions on the transition probability and showed the applications of these results to queueing models. A more general theorem on approachability for vector payoff stochastic games was proved by Milman (2006). Below we briefly describe his result.

Consider a stochastic game with finite state space X and action spaces $A(x) \subset A$ and $B(x) \subset B$, where A and B are finite sets. The stage payoff function is $u : X \times A \times B \to \mathbb{R}^k$. For any strategies $\pi \in \Pi$ and $\gamma \in \Gamma$ and an initial state $x = x_1$, there exists a unique probability measure $P_x^{\pi\gamma}$ on the space of all plays (the Ionescu-Tulcea theorem) generated by these strategies and the transition probability q. By $PD_x^{\pi\gamma}$ we denote the probability distribution on the stream of vector payoffs $\overline{g} = (g_1, g_2, \ldots)$. Clearly, $PD_x^{\pi\gamma}$ is uniquely induced by $P_x^{\pi\gamma}$.

A closed set $C \subset \mathbb{R}^k$ is *approachable in probability from all initial states* $x \in X$, if there exists a strategy $\pi_0 \in \Pi$ such that for any $x \in X$ and $\epsilon > 0$ we have

$$\lim_{n\to\infty} \sup_{\gamma\in\Gamma} PD_x^{\pi_0\gamma}(\{\overline{g} : d(\overline{g}_k, C) > \epsilon\}) = 0.$$

Assume that $y \notin C$ and $z \in P_C(y)$. Let $\sigma(z, y) := (z - y)/\|z - y\|$. Consider the stochastic game with scalarized payoffs $u_\sigma(x, a, b) := \langle u(x, a, b), \sigma(z, y)\rangle$. By Mertens and Neyman (1981) this game has a uniform value, denoted here by $v_\sigma(x)$, $x \in X$. An analogue to the theorem of Blackwell (1956), due to Milman (2006), sounds as follows.

Theorem 25. *A closed set $C \subset \mathbb{R}^k$ is approachable in probability from all initial states $x \in X$ if, for each $y \notin C$, there exists $z \in P_C(y)$ such that $v_\sigma(x) \geq \langle z, \sigma(z, y)\rangle$ for all $x \in X$.*

We close this section by mentioning a recent paper by Kalathil et al. (2016) devoted to the approachability problem in Stackelberg stochastic games with vector costs. They constructed a simple and computationally tractable strategy for approachability for this class of games and gave a reinforcement learning algorithm for learning the approachable strategy when the transition kernel is unknown.

11 Stochastic Games with Short-Stage Duration and Related Models

Studying continuous-time Markov games entails some conceptual and mathematical difficulties. One of the main issues concerns randomization in continuous time. Zachrisson (1964) first considered zero-sum Markov games of a finite and commonly known duration. His method of evaluating the stream of payoffs in continuous time was simply to integrate over time. In his approach, the players use Markov strategies, i.e., they choose their actions as a function of time and the current state only. Stochastic games on Markov jump processes were studied by many authors (see, e.g., Guo and Hernández-Lerma 2003, 2005). The payoff functions and transition rates are time independent, and it is assumed that using randomized Markov strategies, the players determine an infinitesimal operator of the stochastic process, whose trajectories determine the stream of payoffs. The assumptions made on the primitives imply that the players have optimal stationary strategies in the zero-sum case (stationary equilibria in the nonzero-sum case), i.e., strategies that are independent of time, but depend on the state that changes at random time epochs. Altman and Gaitsgory (1995) studied zero-sum "hybrid games," where the state evolves according to a linear continuous-time dynamics. The parameters of the state evolution equation may change at discrete times according to a countable state Markov chain that is directly controlled by both players. Each player has a finite action space. The authors proposed a procedure (similar in form to the well-known maximum principle) that determines a pair of stationary strategies for the players, which is asymptotically a saddle point, as the number of transitions during the finite time horizon grows to infinity. Levy (2013) studied some connections of continuous-time (finite state and action spaces) n-person Markov games with differential games and the theory of differential inclusions. He also gave some results on correlated equilibria with public randomization in an approximating game. He considered Markov strategies only. We mention his paper here because no section on continuous-time games is included in our chapter on nonzero-sum stochastic games. Cardaliaguet et al. (2012) considered the asymptotic value of two-person zero-sum repeated games with incomplete information games, splitting games, and absorbing games. They used a technique relying on embedding the discrete repeated game into a continuous-time game and using the viscosity solution methods. Other approaches to continuous-time Markov games including discretization of time are briefly described in Laraki and Sorin (2015). The class of games discussed here is important for many applications, e.g., in studying queueing models involving birth and death processes and more general ones (see Altman et al. 1997).

Recently, Neyman (2013) presented a framework for fairly general strategies using an asymptotic analysis of stochastic games with stage duration converging to zero. He established some new results, especially on the uniform value and approximate equilibria. There has been very little development in this direction. In order to describe briefly certain ideas from Neyman (2013), we must introduce some notation. We assume that the state space X and the action sets A and B are finite. Let $\delta > 0$ and Γ_δ be a zero-sum stochastic game played in stages $t\delta$, $t \in \mathbb{N}$. Strategies for the players are defined in the usual way, but we should note that the players act in time epochs δ, 2δ, and so on. Following Neyman (2013), we say that δ is the *stage duration*. The stage payoff function $u_\delta : X \times A \times B \to \mathbb{R}$ is assumed to depend on δ. The evaluation of streams of payoffs in a multistage game is not specified at this moment. The transition probability q_δ also depends on δ and is defined using so-called *transition rate function* $q_\delta^0 : X \times X \times A \times B \to \mathbb{R}$ satisfying standard assumptions

$$q_\delta^0(y,x,a,b) \geq 0 \quad \text{for } y \neq x, \quad q_\delta^0(y,y,a,b) \geq -1 \quad \text{and} \quad \sum_{y \in X} q_\delta^0(y,x,a,b) = 0.$$

The transition probability is

$$q_\delta(y|x,a,b) = q_\delta^0(y,x,a,b) \text{ if } y \neq x \quad \text{and} \quad q_\delta(x|x,a,b) = q_\delta^0(x,x,a,b) + 1$$

for all $x \in X$, $a \in A$ and $b \in B$. The transition rate $q_\delta^0(y,x,a,b)$ represents the difference between the probability that the next state will be y and the probability (0 or 1) that the current state is y when the current state is x and the players' actions are a and b, respectively.

Following Neyman (2013), we say that the family of games $(\Gamma_\delta)_{\delta>0}$ is *converging* if there exist functions $\mu : X \times X \times A \times B \to \mathbb{R}$ and $u : X \times A \times B \to \mathbb{R}$ such that for all $x, y \in X$, $a \in A$, and $b \in B$, we have

$$\lim_{\delta \to 0+} \frac{q_\delta^0(y,x,a,b)}{\delta} = \mu(y,x,a,b) \quad \text{and} \quad \lim_{\delta \to 0+} \frac{u_\delta(x,a,b)}{\delta} = u(x,a,b),$$

and the family of games $(\Gamma_\delta)_{\delta>0}$ is *exact* if there exist functions $\mu : X \times X \times A \times B \to \mathbb{R}$ and $u : X \times A \times B \to \mathbb{R}$ such that for all $x, y \in X$, $a \in A$, and $b \in B$, we have $q_\delta^0(y,x,a,b)/\delta = \mu(y,x,a,b)$ and $u_\delta(x,a,b)/\delta = u(x,a,b)$.

Assume that $(x_1, a_1, b_1, \ldots)$ is a play in the game with stage duration δ. According to Neyman (2013), the unnormalized payoff in the ρ-discounted game, denoted by $\Gamma_{\delta,\rho}$, is

$$\sum_{t=1}^{\infty}(1-\rho\delta)^{t-1}u_\delta(x_t,a_t,b_t).$$

The discount factor β in the sense of Sect. 3 is $1-\delta\rho$. It is called admissible, if $\lim_{\delta\to 0^+}(1-\beta(\delta))/\delta$ exists. This limit is known as an *asymptotic discount rate*. In the case of $\beta(\delta)=1-\rho\delta$, $\rho>0$ is the asymptotic discount rate. Other example of an admissible δ-dependent discount factor is $e^{-\rho\delta}$. Assuming that the family of games $(\Gamma_\delta)_{\delta>0}$ is converging, it is proved that the value of $\Gamma_{\delta,\rho}$, denoted by $v_{\delta,\rho}(x)$, converges to some $v_\rho(x)$ (called the asymptotic ρ-discounted value) for any initial state $x\in X$ as $\delta\to 0^+$ and the players have stationary optimal strategies π_ρ and γ_ρ that are independent of δ. Optimality of π_ρ means that π_ρ is $\epsilon(\delta)$-optimal in the game $\Gamma_{\delta,\rho}$, where $\epsilon(\delta)\to 0$ as $\delta\to 0^+$. Similarly, we define the optimality for γ_ρ. For details the reader is referred to Theorem 1 in Neyman (2013).

For any play $(x_1,a_1,b_1,\ldots)$ and $s>0$, define the average per unit time payoff $g_\delta(s)$ as

$$g_\delta(s):=\frac{1}{s}\sum_{1\le t<s/\delta}u_\delta(x_t,a_t,b_t).$$

A family $(\Gamma_\delta)_{\delta>0}$ of *two*-person zero-sum stochastic games has an *asymptotic uniform value* $v(x)$ $(x\in X)$ if for every $\epsilon>0$ there are strategies π_δ of player 1 and γ_δ of player 2, a duration $\delta_0>0$ and a time $s_0>0$ such that for every $\delta\in(0,\delta_0)$ and $s>s_0$, strategy π of player 1, and strategy γ of player 2, we have

$$\epsilon+E_x^{\pi_\delta\gamma}g_\delta(s)\ge v(x)\ge E_x^{\pi\gamma_\delta}g_\delta(s)-\epsilon.$$

Theorem 6 in Neyman (2013) states that any exact family of zero-sum games $(\Gamma_\delta)_{\delta>0}$ has an asymptotic uniform value.

The paper by Neyman (2013) contains also some results on the limit-average games and n-person games with short-stage duration. His asymptotic analysis is partly based on the theory of Bewley and Kohlberg (1976a) and Mertens and Neyman (1981). His work inspired other researchers. For instance, Cardaliaguet et al. (2016) studied the asymptotics of a class of *two*-person zero-sum stochastic game with incomplete information on one side. Furthermore, Gensbittel (2016) considered a zero-sum dynamic game with incomplete information, in which one player is more informed. He analyzed the limit value and gave its characterization through an auxiliary optimization problem and as the unique viscosity solution of a Hamilton-Jacobi equation. Sorin and Vigeral (2016), on the other hand, examined stochastic games with varying duration using iterations of non-expansive Shapley operators that were successfully used in the theory of discrete-time repeated and stochastic games.

Acknowledgements We thank Tamer Başar and Georges Zaccour for inviting us to write this chapter and their help. We also thank Eilon Solan, Sylvain Sorin, William Sudderth, and two reviewers for their comments on an earlier version of this survey.

References

Aliprantis C, Border K (2006) Infinite dimensional analysis: a Hitchhiker's guide. Springer, New York

Altman E, Avrachenkov K, Marquez R, Miller G (2005) Zero-sum constrained stochastic games with independent state processes. Math Meth Oper Res 62:375–386

Altman E, Feinberg EA, Shwartz A (2000) Weighted discounted stochastic games with perfect information. Annals of the International Society of Dynamic Games, vol 5. Birkhäuser, Boston, pp 303–324

Altman E, Gaitsgory VA (1995) A hybrid (differential-stochastic) zero-sum game with fast stochastic part. Annals of the International Society of Dynamic Games, vol 3. Birkhäuser, Boston, pp 47–59

Altman E, Hordijk A (1995) Zero-sum Markov games and worst-case optimal control of queueing systems. Queueing Syst Theory Appl 21:415–447

Altman E, Hordijk A, Spieksma FM (1997) Contraction conditions for average and α-discount optimality in countable state Markov games with unbounded rewards. Math Oper Res 22:588–618

Arapostathis A, Borkar VS, Fernández-Gaucherand, Gosh MK, Markus SI (1993) Discrete-time controlled Markov processes with average cost criterion: a survey. SIAM J Control Optim 31:282–344

Avrachenkov K, Cottatellucci L, Maggi L (2012) Algorithms for uniform optimal strategies in two-player zero-sum stochastic games with perfect information. Oper Res Lett 40:56–60

Basu S (1999) New results on quantifier elimination over real-closed fields and applications to constraint databases. J ACM 46:537–555

Basu S, Pollack R, Roy MF (2003) Algorithms in real algebraic geometry. Springer, New York

Basu A, Stettner Ł (2015) Finite- and infinite-horizon Shapley games with non-symmetric partial observation. SIAM J Control Optim 53:3584–3619

Başar T, Olsder GJ (1995) Dynamic noncooperative game theory. Academic, New York

Berge C (1963) Topological spaces. MacMillan, New York

Bertsekas DP, Shreve SE (1996) Stochastic Optimal Control: the Discrete-Time Case. Athena Scientic, Belmont

Bewley T, Kohlberg E (1976a) The asymptotic theory of stochastic games. Math Oper Res 1:197–208

Bewley T, Kohlberg E (1976b) The asymptotic solution of a recursion equation occurring in stochastic games. Math Oper Res 1:321–336

Bewley T, Kohlberg E (1978) On stochastic games with stationary optimal strategies. Math Oper Res 3:104–125

Bhattacharya R, Majumdar M (2007) Random dynamical systems: theory and applications. Cambridge University Press, Cambridge

Billingsley P (1968) Convergence of probability measures. Wiley, New York

Blackwell DA, Girshick MA (1954) Theory of games and statistical decisions. Wiley and Sons, New York

Blackwell D (1956) An analog of the minmax theorem for vector payoffs. Pac J Math 6:1–8

Blackwell D (1962) Discrete dynamic programming. Ann Math Statist 33:719–726

Blackwell D (1965) Discounted dynamic programming. Ann Math Statist 36: 226–235

Blackwell D (1969) Infinite G_δ-games with imperfect information. Zastosowania Matematyki (Appl Math) 10:99–101

Blackwell D (1989) Operator solution of infinite G_δ-games of imperfect information. In: Anderson TW et al (eds) Probability, Statistics, and Mathematics: Papers in Honor of Samuel Karlin. Academic, New York, pp 83–87

Blackwell D, Ferguson TS (1968) The big match. Ann Math Stat 39:159–163

Bolte J, Gaubert S, Vigeral G (2015) Definable zero-sum stochastic games. Math Oper Res 40: 80–104

Boros E, Elbassioni K, Gurvich V, Makino K (2013) On canonical forms for zero-sum stochastic mean payoff games. Dyn Games Appl 3:128–161
Boros E, Elbassioni K, Gurvich V, Makino K (2016) A potential reduction algorithm for two-person zero-sum mean payoff stochastic games. Dyn Games Appl doi:10.1007/s13235-016-0199-x
Breton M (1991) Algorithms for stochastic games. In: Stochastic games and related topics. Shapley honor volume. Kluwer, Dordrecht, pp 45–58
Brown LD, Purves R (1973) Measurable selections of extrema. Ann Stat 1:902–912
Cardaliaguet P, Laraki R, Sorin S (2012) A continuous time approach for the asymptotic value in two-person zero-sum repeated games. SIAM J Control Optim 50:1573–1596
Cardaliaguet P, Rainer C, Rosenberg D, Vieille N (2016) Markov games with frequent actions and incomplete information-The limit case. Math Oper Res 41:49–71
Cesa-Bianchi N, Lugosi G (2006) Prediction, learning, and games. Cambridge University Press, Cambridge
Cesa-Bianchi N, Lugosi G, Stoltz G (2006) Regret minimization under partial monitoring. Math Oper Res 31:562–580
Charnes A, Schroeder R (1967) On some tactical antisubmarine games. Naval Res Logistics Quarterly 14:291–311
Chatterjee K, Majumdar R, Henzinger TA (2008) Stochastic limit-average games are in EXPTIME. Int J Game Theory 37:219–234
Condon A (1992) The complexity of stochastic games. Inf Comput 96:203–224
Coulomb JM (1992) Repeated games with absorbing states and no signals. Int J Game Theory 21:161–174
Coulomb JM (1999) Generalized big-match. Math Oper Res 24:795–816
Coulomb JM (2001) Absorbing games with a signalling structure. Math Oper Res 26:286–303
Coulomb JM (2003a) Absorbing games with a signalling structure. In: Neyman A, Sorin S (eds) Stochastic games and applications. Kluwer, Dordrecht, pp 335–355
Coulomb JM (2003b) Games with a recursive structure. In: Neyman A, Sorin S (eds) Stochastic games and applications. Kluwer, Dordrecht, pp 427–442
Coulomb JM (2003c) Stochastic games with imperfect monitoring. Int J Game Theory (2003) 32:73–96
Couwenbergh HAM (1980) Stochastic games with metric state spaces. Int J Game Theory 9:25–36
de Alfaro L, Henzinger TA, Kupferman O (2007) Concurrent reachability games. Theoret Comp Sci 386:188–217
Dubins LE (1957) A discrete invasion game. In: Dresher M et al (eds) Contributions to the theory of games III. Annals of Mathematics Studies, vol 39. Princeton University Press, Princeton, pp 231–255
Dubins LE, Maitra A, Purves R, Sudderth W (1989) Measurable, nonleavable gambling problems. Israel J Math 67:257–271
Dubins LE, Savage LJ (2014) Inequalities for stochastic processes. Dover, New York
Everett H (1957) Recursive games. In: Dresher M et al (eds) Contributions to the theory of games III. Annals of Mathematics Studies, vol 39. Princeton University Press, Princeton, pp 47–78
Fan K (1953) Minmax theorems. Proc Nat Acad Sci USA 39:42–47
Feinberg EA (1994) Constrained semi-Markov decision processes with average rewards. Math Methods Oper Res 39:257–288
Feinberg EA, Lewis ME (2005) Optimality of four-threshold policies in inventory systems with customer returns and borrowing/storage options. Probab Eng Inf Sci 19:45–71
Filar JA (1981) Ordered field property for stochastic games when the player who controls transitions changes from state to state. J Optim Theory Appl 34:503–513
Filar JA (1985). Player aggregation in the travelling inspector model. IEEE Trans Autom Control 30:723–729
Filar JA, Schultz TA, Thuijsman F, Vrieze OJ (1991) Nonlinear programming and stationary equilibria of stochastic games. Math Program Ser A 50:227–237

Filar JA, Tolwinski B (1991) On the algorithm of Pollatschek and Avi-Itzhak. In: Stochastic games and related topics. Shapley honor volume. Kluwer, Dordrecht, pp 59–70

Filar JA, Vrieze, OJ (1992) Weighted reward criteria in competitive Markov decision processes. Z Oper Res 36:343–358

Filar JA, Vrieze K (1997) Competitive Markov decision processes. Springer, New York

Flesch J, Thuijsman F, Vrieze OJ (1999) Average-discounted equilibria in stochastic games. European J Oper Res 112:187–195

Fristedt B, Lapic S, Sudderth WD (1995) The big match on the integers. Annals of the international society of dynamic games, vol 3. Birkhäuser, Boston, pp 95–107

Gale D, Steward EM (1953) Infinite games with perfect information. In: Kuhn H, Tucker AW (eds) Contributions to the theory of games II. Annals of mathematics studies, vol 28. Princeton University Press, Princeton, pp 241–266

Gensbittel F (2016) Continuous-time limit of dynamic games with incomplete information and a more informed player. Int J Game Theory 45:321–352

Gensbittel F, Oliu-Barton M, Venel X (2014) Existence of the uniform value in repeated games with a more informed controller. J Dyn Games 1:411–445

Gensbittel F, Renault J (2015) The value of Markov chain games with lack of information on both sides. Math Oper Res 40:820–841

Gillette D (1957) Stochastic games with zero stop probabilities.In: Dresher M et al (eds) Contributions to the theory of games III. Annals of mathematics studies, vol 39. Princeton University Press, Princeton, pp 179–187

Ghosh MK, Bagchi A (1998) Stochastic games with average payoff criterion. Appl Math Optim 38:283–301

Gimbert H, Renault J, Sorin S, Venel X, Zielonka W (2016) On values of repeated games with signals. Ann Appl Probab 26:402–424

González-Trejo JI, Hernández-Lerma O, Hoyos-Reyes LF (2003) Minmax control of discrete-time stochastic systems. SIAM J Control Optim 41:1626–1659

Guo X, Hernández-Lerma O (2003) Zero-sum games for continuous-time Markov chains with unbounded transitions and average payoff rates. J Appl Probab 40:327–345

Guo X, Hernńdez-Lerma O (2005) Nonzero-sum games for continuous-time Markov chains with unbounded discounted payoffs. J Appl Probab 42:303–320

Hall P, Heyde C (1980) Martingale limit theory and its applications. Academic, New York

Hansen LP, Sargent TJ (2008) Robustness. Princeton University Press, Princeton

Haurie A, Krawczyk JB, Zaccour G (2012) Games and dynamic games. World Scientific, Singapore

Hernández-Lerma O, Lasserre JB (1996) Discrete-time Markov control processes: basic optimality criteria. Springer-Verlag, New York

Hernández-Lerma O, Lasserre JB (1999) Further topics on discrete-time Markov control processes. Springer, New York

Himmelberg CJ (1975) Measurable relations. Fundam Math 87:53–72

Himmelberg CJ, Van Vleck FS (1975) Multifunctions with values in a space of probability measures. J Math Anal Appl 50:108–112

Hoffman AJ, Karp RM (1966) On non-terminating stochastic games. Management Sci 12:359–370

Hordijk A, Kallenberg LCM (1981) Linear programming and Markov games I, II. In: Moeschlin O, Pallaschke D (eds) Game theory and mathematical economics, North-Holland, Amsterdam, pp 291–320

Iyengar GN (2005) Robust dynamic programming. Math Oper Res 30:257–280

Jaśkiewicz A (2002) Zero-sum semi-Markov games. SIAM J Control Optim 41:723–739

Jaśkiewicz A (2004) On the equivalence of two expected average cost criteria for semi-Markov control processes. Math Oper Res 29:326–338

Jaśkiewicz A (2007) Average optimality for risk-sensitive control with general state space. Ann Appl Probab 17: 654–675

Jaśkiewicz A (2009) Zero-sum ergodic semi-Markov games with weakly continuous transition probabilities. J Optim Theory Appl 141:321–347

Jaśkiewicz A (2010) On a continuous solution to the Bellman-Poisson equation in stochastic games. J Optim Theory Appl 145:451–458
Jaśkiewicz A, Nowak AS (2006) Zero-sum ergodic stochastic games with Feller transition probabilities. SIAM J Control Optim 45:773–789
Jaśkiewicz A, Nowak AS (2011) Stochastic games with unbounded payoffs: Applications to robust control in economics. Dyn Games Appl 1: 253–279
Jaśkiewicz A, Nowak AS (2014) Robust Markov control process. J Math Anal Appl 420:1337–1353
Jaśkiewicz A, Nowak AS (2018) Non-zero-sum stochastic games. In: Başar T, Zaccour G (eds) Handbook of dynamic game theory. Birkhäuser, Basel
Kalathil, D, Borkar VS, Jain R (2016) Approachability in Stackelberg stochastic games with vector costs. Dyn Games Appl. doi:10.1007/s13235–016-0198-y
Kartashov NV(1996) Strong stable Markov chains. VSP, Utrecht, The Netherlands
Kehagias A, Mitschke D, Praşat P (2013) Cops and invisible robbers: The cost of drunkenness. Theoret Comp Sci 481:100–120
Kertz RP, Nachman D (1979) Persistently optimal plans for nonstationary dynamic programming: The topology of weak convergence case. Ann Probab 7:811–826
Klein E, Thompson AC (1984) Theory of correspondences. Wiley, New York
Kohlberg E (1974) Repeated games with absorbing states. Ann Statist 2:724–738
Krass D, Filar JA, Sinha S (1992) A weighted Markov decision process. Oper Res 40:1180–1187
Krausz A, Rieder U (1997) Markov games with incomplete information. Math Meth Oper Res 46:263–279
Krishnamurthy N, Parthasarathy T (2011) Multistage (stochastic) games. Wiley encyclopedia of operations research and management science. Wiley online library. doi:10.1002/9780470400531.eorms0551
Kumar PR, Shiau TH (1981) Existence of value and randomized strategies in zero-sum discrete-time stochastic dynamic games. SIAM J Control Optim 19:617–634
Kuratowski K, Ryll-Nardzewski C (1965) A general theorem on selectors. Bull Polish Acad Sci (Ser Math) 13:397–403
Küenle HU (1986) Stochastische Spiele und Entscheidungsmodelle. BG Teubner, Leipzig
Küenle HU (2007) On Markov games with average reward criterion and weakly continuous transition probabilities. SIAM J Control Optim 45:2156–2168
Laraki R (2010) Explicit formulas for repeated games with absorbing states. Int J Game Theory 39:53–69
Laraki R, Sorin S (2015) Advances in zero-sum dynamic games. In: Young HP, Zamir S (eds) Handbook of game theory with economic applications, vol 4. North Holland, pp 27–93
Lehrer E, Solan E (2016) A general internal regret-free strategy. Dyn Games Appl 6:112–138
Lehrer E, Sorin S (1992) A uniform Tauberian theorem in dynamic programming. Math Oper Res 17:303–307
Levy Y (2012) Stochastic games with information lag. Games Econ Behavior 74:243–256
Levy Y (2013) Continuous-time stochastic games of fixed duration. Dyn Games Appl 3:279–312
Li X, Venel X (2016) Recursive games: uniform value, Tauberian theorem and the Mertens conjecture "$Maxmin = \lim v_n = \lim v_\lambda$". Int J Game Theory 45:155–189
Maccheroni F, Marinacci M, Rustichini A (2006) Dynamic variational preferences. J Econ Theory 128:4–44
Maitra A, Parthasarathy T (1970) On stochastic games. J Optim Theory Appl 5:289–300
Maitra A, Parthasarathy T (1971) On stochastic games II. J Optim Theory Appl 8:154–160
Maitra A, Sudderth W (1992) An operator solution for stochastic games. Israel J Math 78:33–49
Maitra A, Sudderth W (1993a) Borel stochastic games with limsup payoffs. Ann Probab 21:861–885
Maitra A, Sudderth W (1993b) Finitely additive and measurable stochastic games. Int J Game Theory 22:201–223
Maitra A, Sudderth W (1996) Discrete gambling and stochastic games. Springer, New York

Maitra A, Sudderth W (1998) Finitely additive stochastic games with Borel measurable payoffs. Int J Game Theory 27:257–267
Maitra A, Sudderth W (2003a) Stochastic games with limsup payoff. In: Neyman A, Sorin S (eds) Stochastic games and applications. Kluwer, Dordrecht, pp 357–366
Maitra A, Sudderth W (2003b) Stochastic games with Borel payoffs. In: Neyman A, Sorin S (eds) Stochastic games and applications. Kluwer, Dordrecht, pp 367–373
Martin D (1975) Borel determinacy. Ann Math 102:363–371
Martin D (1985) A purely inductive proof of Borel determinacy. In: Nerode A, Shore RA (eds) Recursion theory. Proceedings of symposia in pure mathematics, vol 42. American Mathematical Society, Providence, pp 303–308
Martin D (1998) The determinacy of Blackwell games. J Symb Logic 63:1565–1581
Mertens JF (1982) Repeated games: an overview of the zero-sum case. In: Hildenbrand W (ed) Advances in economic theory. Cambridge Univ Press, Cambridge, pp 175–182
Mertens JF (1987) Repeated games. In: Proceedings of the international congress of mathematicians, American mathematical society, Berkeley, California, pp 1528–1577
Mertens JF (2002) Stochastic games. In: Aumann RJ, Hart S (eds) Handbook of game theory with economic applications, vol 3. North Holland, pp 1809–1832
Mertens JF, Neyman A (1981) Stochastic games. Int J Game Theory 10:53–56
Mertens JF, Neyman A (1982) Stochastic games have a value. Proc Natl Acad Sci USA 79: 2145–2146
Mertens JF, Neyman A, Rosenberg D (2009) Absorbing games with compact action spaces. Math Oper Res 34:257–262
Mertens JF, Sorin S, Zamir S (2015) Repeated games. Cambridge University Press, Cambridge, MA
Meyn SP, Tweedie RL (1994) Computable bounds for geometric convergence rates of Markov chains. Ann Appl Probab 4:981–1011
Meyn SP, Tweedie RL (2009) Markov chains and stochastic stability. Cambridge University Press, Cambridge
Miao J (2014) Economic dynamics in discrete time. MIT Press, Cambridge
Milman E (2002) The semi-algebraic theory of stochastic games. Math Oper Res 27:401–418
Milman E (2006) Approachable sets of vector payoffs in stochastic games. Games Econ Behavior 56:135–147
Milnor J, Shapley LS (1957) On games of survival. In: Dresher M et al (eds) Contributions to the theory of games III. Annals of mathematics studies, vol 39. Princeton University Press, Princeton, pp 15–45
Mondal P, Sinha S, Neogy SK, Das AK (2016) On discounted ARAT semi-Markov games and its complementarity formulations. Int J Game Theory 45:567–583
Mycielski J (1992) Games with perfect information. In: Aumann RJ, Hart S (eds) Handbook of Game Theory with Economic Applications, vol 1. North Holland, pp 41–70
Neyman A (2003a) Real algebraic tools in stochastic games. In: Neyman A, Sorin S (eds) Stochastic Games and Applications. Kluwer, Dordrecht, pp 57–75
Neyman A (2003b) Stochastic games and nonexpansive maps. In: Neyman A, Sorin S (eds) Stochastic Games and Applications. Kluwer, Dordrecht, pp 397–415
Neyman A (2013) Stochastic games with short-stage duration. Dyn Games Appl 3:236–278
Neyman A, Sorin S (eds) (2003) Stochastic games and applications. Kluwer, Dordrecht
Neveu J (1965) Mathematical foundations of the calculus of probability. Holden-Day, San Francisco
Nowak AS (1985a) Universally measurable strategies in zero-sum stochastic games. Ann Probab 13: 269–287
Nowak AS (1985b) Measurable selection theorems for minmax stochastic optimization problems, SIAM J Control Optim 23:466–476
Nowak AS (1986) Semicontinuous nonstationary stochastic games. J Math Analysis Appl 117:84–99

Nowak AS (1994) Zero-sum average payoff stochastic games wit general state space. Games Econ Behavior 7:221–232
Nowak AS (2010) On measurable minmax selectors. J Math Anal Appl 366:385–388
Nowak AS, Raghavan TES (1991) Positive stochastic games and a theorem of Ornstein. In: Stochastic games and related topics. Shapley honor volume. Kluwer, Dordrecht, pp 127–134
Oliu-Barton M (2014) The asymptotic value in finite stochastic games. Math Oper Res 39:712–721
Ornstein D (1969) On the existence of stationary optimal strategies. Proc Am Math Soc 20:563–569
Papadimitriou CH (1994) Computational complexity. Addison-Wesley, Reading
Parthasarathy KR (1967) Probability measures on metric spaces. Academic, New York
Parthasarathy T, Raghavan TES (1981) An order field property for stochastic games when one player controls transition probabilities. J Optim Theory Appl 33:375–392
Patek SD, Bertsekas DP (1999) Stochastic shortest path games. SIAM J Control Optim 37:804–824
Perchet V (2011a) Approachability of convex sets in games with partial monitoring. J Optim Theory Appl 149:665–677
Perchet V (2011b) Internal regret with partial monitoring calibration-based optimal algorithms. J Mach Learn Res 12:1893–1921
Pollatschek M, Avi-Itzhak B (1969) Algorithms for stochastic games with geometrical interpretation. Manag Sci 15:399–425
Prikry K, Sudderth WD (2016) Measurability of the value of a parametrized game. Int J Game Theory 45:675–683
Raghavan TES (2003) Finite-step algorithms for single-controller and perfect information stochastic games. In: Neyman A, Sorin S (eds) Stochastic games and applications. Kluwer, Dordrecht, pp 227–251
Raghavan TES, Ferguson TS, Parthasarathy T, Vrieze OJ, eds. (1991) Stochastic games and related topics: In honor of professor LS Shapley, Kluwer, Dordrecht
Raghavan TES, Filar JA (1991) Algorithms for stochastic games: a survey. Z Oper Res (Math Meth Oper Res) 35: 437–472
Raghavan TES, Syed Z (2003) A policy improvement type algorithm for solving zero-sum two-person stochastic games of perfect information. Math Program Ser A 95:513–532
Renault J (2006) The value of Markov chain games with lack of information on one side. Math Oper Res 31:490–512
Renault J (2012) The value of repeated games with an uninformed controller. Math Oper Res 37:154–179
Renault J (2014) General limit value in dynamic programming. J Dyn Games 1:471–484
Rosenberg D (2000) Zero-sum absorbing games with incomplete information on one side: asymptotic analysis. SIAM J Control Optim 39:208–225
Rosenberg D, Sorin S (2001) An operator approach to zero-sum repeated games. Israel J Math 121:221–246
Rosenberg D, Solan E, Vieille N (2002) Blackwell optimality in Markov decision processes with partial observation. Ann Stat 30:1178–1193
Rosenberg D, Solan E, Vieille N (2003) The maxmin value of stochastic games with imperfect monitoring. Int J Game Theory 32:133–150
Rosenberg D, Solan E, Vieille N (2004) Stochastic games with a single controller and incomplete information. SIAM J Control Optim 43:86–110
Rosenberg D, Vieille N (2000) The maxmin of recursive games with incomplete information on one side. Math Oper Res 25:23–35
Scarf HE, Shapley LS (1957) A discrete invasion game. In: Dresher M et al (eds) Contributions to the Theory of Games III. Annals of Mathematics Studies, vol 39, Princeton University Press, Princeton, pp 213–229
Schäl M (1975) Conditions for optimality in dynamic programming and for the limit of n-stage optimal policies to be optimal. Z Wahrscheinlichkeitstheorie Verw Geb 32:179–196
Secchi P (1997) Stationary strategies for recursive games. Math Oper Res 22:494–512

Secchi P (1998) On the existence of good stationary strategies for nonleavable stochastic games. Int J Game Theory 27:61–81
Selten R (1975) Re-examination of the perfectness concept for equilibrium points in extensive games. Int J Game Theory 4:25–55
Shapley LS (1953) Stochastic games. Proc Nat Acad Sci USA 39:1095–1100
Shimkin N, Shwartz A (1993) Guaranteed performance regions for Markovian systems with competing decision makers. IEEE Trans Autom Control 38:84–95
Shmaya E (2011) The determinacy of infinite games with eventual perfect monitoring. Proc Am Math Soc 139:3665–3678
Solan E (2009) Stochastic games. In: Meyers RA (ed) Encyclopedia of complexity and systems science. Springer, New York, pp 8698–8708
Solan E, Vieille N (2002) Uniform value in recursive games. Ann Appl Probab 12:1185–1201
Solan E, Vieille N (2010) Camputing uniformly optimal strategies in two-player stochastic games. Econ Theory 42:237–253
Solan E, Vieille N (2015) Stochastic games. Proc Nat Acad Sci USA 112:13743–13746
Solan E, Ziliotto B (2016) Stochastic games with signals. Annals of the international society of dynamic game, vol 14. Birkhäuser, Boston, pp 77–94
Sorin S (1984) Big match with lack of information on one side (Part 1). Int J Game Theory 13:201–255
Sorin S (1985) Big match with lack of information on one side (Part 2). Int J Game Theory 14:173–204
Sorin S (2002) A first course on zero-sum repeated games. Mathematiques et applications, vol 37. Springer, New York
Sorin S (2003a) Stochastic games with incomplete information. In: Neyman A, Sorin S (eds) Stochastic Games and Applications. Kluwer, Dordrecht, pp 375–395
Sorin S (2003b) The operator approach to zero-sum stochastic games. In: Neyman A, Sorin S (eds) Stochastic Games and Applications. Kluwer, Dordrecht
Sorin S (2004) Asymptotic properties of monotonic nonexpansive mappings. Discrete Event Dyn Syst 14:109–122
Sorin S, Vigeral G (2015a) Existence of the limit value of two-person zero-sum discounted repeated games via comparison theorems. J Optim Theory Appl 157:564–576
Sorin S, Vigeral G (2015b) Reversibility and oscillations in zero-sum discounted stochastic games. J Dyn Games 2:103–115
Sorin S, Vigeral G (2016) Operator approach to values of stochastic games with varying stage duration. Int J Game Theory 45:389–410
Sorin S, Zamir S (1991) "Big match" with lack on information on one side (Part 3). In: Shapley LS, Raghavan TES (eds) Stochastic games and related topics. Shapley Honor Volume. Kluwer, Dordrecht, pp 101–112
Spinat X (2002) A necessary and sufficient condition for approachability. Math Oper Res 27:31–44
Stokey NL, Lucas RE, Prescott E (1989) Recursive methods in economic dynamics. Harvard University Press, Cambridge
Strauch R (1966) Negative dynamic programming. Ann Math Stat 37:871–890
Szczechla W, Connell SA, Filar JA, Vrieze OJ (1997) On the Puiseux series expansion of the limit discount equation of stochastic games. SIAM J Control Optim 35:860–875
Tarski A (1951) A decision method for elementary algebra and geometry. University of California Press, Berkeley
Van der Wal I (1978) Discounted Markov games: Generalized policy iteration method. J Optim Theory Appl 25:125–138
Vega-Amaya O (2003) Zero-sum average semi-Markov games: fixed-point solutions of the Shapley equation. SIAM J Control Optim 42:1876–1894
Vega-Amaya O, Luque-Vásquez (2000) Sample path average cost optimality for semi-Markov control processes on Borel spaces: unbounded costs and mean holding times. Appl Math (Warsaw) 27:343–367
Venel X (2015) Commutative stochastic games. Math Oper Res 40:403–428

Vieille N (2002) Stochastic games: recent results. In: Aumann RJ, Hart S (eds) Handbook of Game Theory with Economic Applications, vol 3. North Holland, Amsterdam/London, pp 1833–1850
Vigeral G (2013) A zero-sum stochastic game with compact action sets and no asymptotic value. Dyn Games Appl 3:172–186
Vrieze OJ (1981) Linear programming and undiscounted stochastic games. Oper Res Spektrum 3:29–35
Vrieze OJ (1987) Stochastic games with finite state and action spaces. Mathematisch Centrum Tract, vol 33. Centrum voor Wiskunde en Informatica, Amsterdam
Vrieze OJ, Tijs SH, Raghavan TES, Filar JA (1983) A finite algorithm for switching control stochastic games. Oper Res Spektrum 5:15–24
Wessels J (1977) Markov programming by successive approximations with respect to weighted supremum norms. J Math Anal Appl 58:326–335
Winston W (1978) A stochastic game model of a weapons development competition. SIAM J Control Optim 16:411–419
Zachrisson LE (1964) Markov games. In: Dresher M, Shapley LS, Tucker AW (eds) Advances in Game Theory. Princeton University Press, Princeton, pp 211–253
Ziliotto B (2016) General limit value in zero-sum stochastic games. Int J Game Theory 45:353–374
Ziliotto B (2016) Zero-sum repeated games: counterexamples to the existence of the asymptotic value and the conjecture. Ann Probab 44:1107–1133

Nonzero-Sum Stochastic Games

6

Anna Jaśkiewicz and Andrzej S. Nowak

Contents

Abstract

This chapter describes a number of results obtained in the last 60 years on the theory of nonzero-sum discrete-time stochastic games. We provide an overview of almost all basic streams of research in this area such as the existence of stationary Nash and correlated equilibria in models on countable and general state spaces, the existence of subgame-perfect equilibria, algorithms, stopping

A. Jaśkiewicz (✉)
Faculty of Pure and Applied Mathematics, Wrocław University of Science and Technology, Wrocław, Poland
e-mail: anna.jaskiewicz@pwr.edu.pl

A. S. Nowak
Faculty of Mathematics, Computer Science and Econometrics, University of Zielona Góra, Zielona Góra, Poland
e-mail: a.nowak@wmie.uz.zgora.pl

T. Başar, G. Zaccour (eds.), *Handbook of Dynamic Game Theory*,
https://doi.org/10.1007/978-3-319-44374-4_33

games, and the existence of uniform equilibria. Our survey incorporates several examples of games studied in operations research and economics. In particular, separate sections are devoted to intergenerational games, dynamic Cournot competition and game models of resource extraction. The provided reference list includes not only seminal papers that commenced research in various directions but also exposes recent advances in this field.

Keywords

Nonzero-sum game · Stochastic game · Discounted payoff · Limit-average payoff · Markov perfect equilibrium · Subgame-perfect equilibrium · Intergenerational altruism · Uniform equilibrium · Stopping game

1 Introduction

The fundamentals of modern theory of non-cooperative dynamic games were established in the 1950s at Princeton University. First Nash (1950) introduced the notion of equilibrium for finite n-person static games and proved its existence using the fixed point theorem of Kakutani (1941). Next Shapley (1953) presented the model of infinite time horizon stochastic zero-sum game with positive stop probability. Fink (1964) and Takahashi (1964) extended his model to nonzero-sum discounted stochastic games with finite state spaces and proved the existence of equilibrium in stationary Markov strategies. Later on, Rogers (1969) and Sobel (1971) obtained similar results for irreducible stochastic games with the expected limit-average payoffs. Afterwards, the theory of discrete-time nonzero-sum stochastic games was extended in various directions inspiring a lot of interesting applications. An overview of selected basic topics in stochastic dynamic games with instructive examples can be found in the books of Başar and Olsder (1995) and Haurie et al. (2012) and the survey paper by Dutta and Sundaram (1998). Theoretically more advanced material is included in the book edited by Neyman and Sorin (2003) and the monograph of Mertens et al. (2015).

In this chapter, we provide an overview of almost all basic streams of research in the area of nonzero-sum discrete-time stochastic games such as: the existence of stationary equilibria in models on both countable and general state spaces, the existence of subgame-perfect equilibria, algorithms, stopping games, correlated and uniform equilibria. Our survey incorporates several examples of games studied in operations research and economics. In particular, separate sections are devoted to intergenerational games, dynamic Cournot competition and game models of resource extraction. The provided reference list not only includes seminal papers that initiated research in various directions but also exposes recent advances in this field.

The paper is organized as follows. In Sect. 2 we describe some basic material needed for an examination of nonzero-sum stochastic games with general state spaces. To make the presentation less technical, we restrict our attention to Borel space models. A great deal of the results described in this chapter are stated in the literature in a more general framework. However, the Borel state space models are sufficiently rich for most applications. Section 2 includes auxiliary results on set-valued mappings arising in the study of the existence of stationary Nash/correlated equilibria and certain known results in the literature such as a random version of the Carathéodory theorem or the Mertens measurable selection theorem. The second part, on the other hand, is devoted to supermodular games. Sect. 3 deals with the concept of subgame-perfect equilibrium in games on a Borel state space and introduces different classes of strategies, in which subgame-perfect equilibria may be obtained. Sect. 4 includes results on correlated equilibria with public signal proved for games on Borel state spaces, whereas Sect. 5 presents the state-of-the-art results on the existence of stationary equilibria (further called "stationary Markov perfect equilibria") in discounted stochastic games. This theory found its applications to several examples examined in operations research and economics. Namely, in Sect. 6 we provide models with special but natural transition structures, for which there exist stationary equilibria. Sect. 7 recalls the papers, where the authors proved the existence of an equilibrium for stochastic games on denumerable state spaces. This section embraces the discounted models as well as models with the limit-average payoff criterion. Moreover, it is also shown that the discounted game with a Borel state space can be approximated, under some assumptions, by simpler games with countable state spaces. Sect. 8 reviews algorithms applied to nonzero-sum stochastic games. In particular, it is shown how a formulation of a linear complementarity problem can be helpful in solving games with discounted and limit-average payoff criteria with special transition structure. In addition, we also mention the homotopy methods applied to this issue. Sect. 9 presents material on games with finite state and action spaces, while Sect. 10 deals with games with product state spaces. In Sect. 11 we formulate results proved for various intergenerational games. Our models incorporate paternalistic and non-paternalistic altruistic economic growth models: games with one, finite, or infinitely many descendants as well as games on one- and multidimensional commodity spaces. Finally, Sect. 12 provides a short overview of stopping games beginning from the Dynkin extension of Neveu's stopping problem.

2 Preliminaries

In this section, we recall essential notations and several facts, which are crucial for studying Nash and correlated equilibria in nonzero-sum stochastic games with uncountable state spaces. Here we follow preliminaries in Jaśkiewicz and Nowak (2018a). Let $N = \{1, \ldots, n\}$ be the set of n-players. Let X, $A_1, \ldots, A_n$ be Borel spaces. Assume that for each $i \in N$, $x \to A_i(x) \subset A_i$ is a lower measurable compact-valued action correspondence for player i. This is equivalent to saying that the graph of this correspondence is a Borel subset of $X \times A_i$. Let $A := A_1 \times \cdots \times A_n$. We consider a nonzero-sum n-person game parameterized by a state variable $x \in X$. The payoff or utility function for player $i \in N$ is $r_i : X \times A \to \mathbb{R}$ and it is assumed that r_i is bounded, $r_i(\cdot, a)$ is Borel measurable for each $a \in A$, and

$r_i(x,\cdot)$ is continuous on A for each $x \in X$. Assuming that $i \in N$ chooses a mixed strategy $\nu_i \in \Pr(A_i(x))$ and $\nu := (\nu_1, \ldots, \nu_n)$, we denote the expected payoff to player i by

$$P^i(x,\nu) := \int_{A_n(x)} \cdots \int_{A_1(x)} r_i(x, a_1, \ldots, a_n)\nu_1(da_1) \times \cdots \times \nu_n(da_n).$$

A strategy profile $\nu^* = (\nu_1^*, \ldots, \nu_n^*)$ is a *Nash equilibrium* in the game for a given state $x \in X$ if

$$P^i(x,\nu^*) \geq P^i(x,(\nu_i,\nu_{-i}^*))$$

for every $i \in N$ and $\nu_i \in \Pr(A_i(x))$. As usual (ν_i, ν_{-i}^*) denotes ν^* with ν_i^* replaced by ν_i. For any $x \in X$, by $\mathcal{N}(x)$, we denote the set of all Nash equilibria in the considered game. By Glicksberg (1952), $\mathcal{N}(x) \neq \emptyset$. It is easy to see that $\mathcal{N}(x)$ is compact. Let $\mathcal{NP}(x) \subset \mathbb{R}^n$ be the set of payoff vectors corresponding to all Nash equilibria in $\mathcal{N}(x)$. By co, we denote the convex hull operator in $\mathbb{R}^n$.

Proposition 1. *The mappings $x \to \mathcal{N}(x)$, $x \to \mathcal{NP}(x)$ and $x \to co\mathcal{NP}(x)$ are compact-valued and lower measurable.*

For a detailed discussion of these statements, consult Nowak and Raghavan (1992), Himmelberg (1975), and Klein and Thompson (1984). By Kuratowski and Ryll-Nardzewski (1965), every set-valued mapping in Proposition 1 has a Borel measurable selection. Making use of standard results on measurable selections (see Castaing and Valadier 1977) and Carathéodory's theorem, we obtain the following result.

Proposition 2. *Let $b : X \to \mathbb{R}^n$ be a Borel measurable selection of the mapping $x \to co\mathcal{NP}(x)$. Then, there exist Borel measurable selections $b^i : X \to \mathbb{R}^n$ and Borel measurable functions $\lambda^i : X \to [0,1]$ ($i = 1, \ldots, n+1$) such that for each $x \in X$, we have*

$$\sum_{i=1}^{n+1} \lambda^i(x) = 1 \quad \textit{and} \quad b(x) = \sum_{i=1}^{n+1} \lambda^i(x) b^i(x).$$

Similarly as in Nowak and Raghavan (1992), from Filippov's measurable implicit function theorem, we conclude the following facts.

Proposition 3. *Let $p : X \to \mathbb{R}^n$ be a Borel measurable selection of the mapping $x \to \mathcal{NP}(x)$. Then, there exists a Borel measurable selection ψ of the mapping $x \to \mathcal{N}(x)$ such that*

$$p(x) = (P^1(x, \psi(x)), \ldots, P^n(x, \psi(x))) \quad \textit{for all} \quad x \in X.$$

Proposition 4. *If $b : X \to \mathbb{R}^n$ is a Borel measurable selection of the mapping $x \to co\mathcal{NP}(x)$, then there exist Borel measurable selections ψ^i of the mapping $x \to \mathcal{N}(x)$ and Borel measurable functions $\lambda^i : X \to [0,1]$ ($i = 1, \ldots, n+1$) such that for each $x \in X$, we have $\sum_{i=1}^{n+1} \lambda^i(x) = 1$ and*

$$b(x) = \sum_{i=1}^{n+1} \lambda^i(x)(P^1(x, \psi^i(x)), \ldots, P^n(x, \psi^i(x))).$$

The following result plays an important role in studying Nash equilibria in stochastic games with Borel state spaces and can be deduced from Theorem 2 in Mertens (2003) and Filippov's measurable implicit function theorem. It is related to Lyapunov's theorem on the range of nonatomic measures and also has some predecessors in control theory; see Artstein (1989).

Proposition 5. *Let μ be a nonatomic Borel probability measure on X. Assume that q_j ($j = 1, \ldots, l$) are Borel measurable transition probabilities from X to X and for every j and $x \in X$, $q_j(\cdot|x) \ll \mu$, i.e., $q_j(\cdot|x)$ is dominated by μ. Let $w^0 : X \to \mathbb{R}^n$ be a Borel measurable mapping such that $w^0(x) \in co\mathcal{NP}(x)$ for each $x \in X$. Then there exists a Borel measurable mapping $v^0 : X \times X \to \mathbb{R}^n$ such that $v^0(x,y) \in \mathcal{NP}(x)$ for all $x, y \in X$ and*

$$\int_X w^0(y) q_j(dy|x) = \int_X v^0(x,y) q_j(dy|x), \quad j = 1, \ldots, l.$$

Moreover, there exists a Borel measurable mapping $\phi : X \times X \to \Pr(A)$ such that $\phi(x,y) \in \mathcal{N}(x)$ for all $x, y \in X$.

Let L be a lattice contained in Euclidean space $\mathbb{R}^k$ equipped with the component-wise order $\geq$. For any $x, y \in L$, $x \vee y$ ($x \wedge y$) denotes the join (meet) of x and y. A function $\phi : L \to \mathbb{R}$ is supermodular if for any $x, y \in L$, it holds

$$\phi(x \vee y) + \phi(x \wedge y) \geq \phi(x) + \phi(y).$$

Clearly, if $k = 1$, then any function ϕ is supermodular. Let $L_1 \subset \mathbb{R}^k$, $L_2 \subset \mathbb{R}^l$ be lattices. A function $\psi : L_1 \times L_2 \to \mathbb{R}$ has increasing differences in (x, y) if for every $x' \geq x$ in L_1, $\psi(x', y) - \psi(x, y)$ is non-decreasing in y. Let the set A_i of pure strategies of player $i \in N$ be a compact convex subset of Euclidean space $\mathbb{R}^{m_i}$. An element a_i of A_i is denoted by $a_i = (a_{i1}, a_{i2}, \ldots, a_{im_i})$. We consider an n-person game G_0 in which $R_i : A \to \mathbb{R}$ is the payoff function of player $i \in N$ and $A := A_1 \times \cdots \times A_n$. As usual, any strategy profile $a = (a_1, a_2, \ldots, a_n)$ can also be denoted as (a_i, a_{-i}) for $i \in N$.

Assume that every A_i is a lattice. The game G_0 is called *supermodular* if for every player $i \in N$ and a_{-i}, the function $a_i \to R_i(a_i, a_{-i})$ is supermodular and R_i has increasing differences in (a_i, a_{-i}).

It is well known that any supermodular game with continuous utility functions and compact strategy sets A_i has a pure Nash equilibrium; see Topkis (1998) or Theorems 4 and 5 in Milgrom and Roberts (1990).

The game G_0 is called *smooth* if every R_i can be extended from A to an open superset A^o in such a way that its second-order partial derivatives exist and are continuous on A^o.

A game G_0 is called a *smooth supermodular* game if for every player $i \in N$,

(a) A_i is a compact interval in $\mathbb{R}^{m_i}$,
(b) $\frac{\partial^2 R_i}{\partial a_{ij} \partial a_{ik}} \geq 0$ on A for all $1 \leq j < k \leq m_i$,
(c) $\frac{\partial^2 R_i}{\partial a_{ij} \partial a_{kl}} \geq 0$ on A for each $k \neq i$ and all $1 \leq j \leq m_i$, $1 \leq l \leq m_k$.

It is well known that any game satisfying conditions (a)–(c) is supermodular. Conditions (a) and (b) imply that R_i is a supermodular function with respect to a_i for fixed a_{-i}, while conditions (a) and (c) imply that R_i has increasing differences in (a_i, a_{-i}). For a detailed discussion of these issues, see Topkis (1978, 1998) or Theorem 4 in Milgrom and Roberts (1990).

In order to obtain a uniqueness of an equilibrium in a smooth supermodular game G_0, one needs an additional assumption, often called a *strict diagonal dominance condition*, see page 1271 in Milgrom and Roberts (1990) or Rosen (1965). As noted by Curtat (1996), this condition can be described for smooth supermodular games as follows. Let $M^i := \{1, 2, \ldots, m_i\}$.

(C1) For every $i \in N$ and $j \in M^i$,

$$\frac{\partial^2 R_i}{\partial a_{ij}^2} + \sum_{l \in M^i \setminus \{j\}} \frac{\partial^2 R_i}{\partial a_{ij} \partial a_{il}} + \sum_{k \in N \setminus \{i\}} \sum_{l \in M^k} \frac{\partial^2 R_i}{\partial a_{ij} \partial a_{kl}} < 0.$$

From Milgrom and Roberts (1990) and page 187 in Curtat (1996), we obtain the following auxiliary result.

Proposition 6. *Any smooth supermodular game G_0 satisfying condition* (C1) *has a unique pure Nash equilibrium.*

Assume now that the payoff functions R_i are parameterized by τ in some partially ordered set T, i.e., $R_i : A \times T \to \mathbb{R}$. We introduce a new condition.

(C2) $\frac{\partial^2 R_i}{\partial a_{ij} \partial \tau} \geq 0$ for all $1 \leq j \leq m_i$, and $i \in N$.

It is known that the set of Nash equilibria in any supermodular game G_0 is a lattice and has the smallest and the largest elements. The following result follows from Theorem 7 in Milgrom and Roberts (1990).

Proposition 7. *Suppose that a smooth supermodular game satisfies* (C2). *Then, the largest and smallest pure Nash equilibria are non-decreasing functions of* τ.

3 Subgame-Perfect Equilibria in Stochastic Games with General State Space

We consider an n-person nonzero-sum discounted stochastic game G as defined below.

- $(X, \mathcal{B}(X))$ is a nonempty Borel *state space* with its Borel σ-algebra $\mathcal{B}(X)$.
- A_i is a Borel *space of actions* for player $i \in N := \{1, \ldots, n\}$.
- $A_i(x) \subset A_i$ is a *set of actions* available to player $i \in N$ in state $x \in X$. The correspondence $x \to A_i(x)$ is lower measurable and compact-valued. Define

$$A := A_1 \times \ldots \times A_n \quad \text{and} \quad A(x) = A_1(x) \times \ldots \times A_n(x), \qquad x \in X.$$

- $u_i : X \times A \to \mathbb{R}$ is a Borel measurable bounded *utility* (or *payoff*) *function* for player $i \in N$. It is assumed that $u_i(x, \cdot)$ is continuous on A for every $x \in X$.
- $q : X \times A \times \mathcal{B}(X) \to [0, 1]$ is a *transition probability*. We assume that $q(D|x, \cdot)$ is continuous on A for each $x \in X$ and $D \in \mathcal{B}(X)$.
- $\beta \in (0, 1)$ is a *discount factor*.

Every stage of the game begins with a state $x \in X$, and after observing x, the players simultaneously choose their actions $a_i \in A_i(x)$ $(i \in N)$ and obtain payoffs $u_i(x, a)$, where $a = (a_1, \ldots, a_n)$. A new state x' is realized from the distribution $q(\cdot|x, a)$ and new period begins with payoffs discounted by β. Let $H_1 = X$ and H_t be the set of all plays $h_t = (x_1, a^1, \ldots, x_{t-1}, a^{t-1}, x_t)$, where $a^k = (a_1^k, \ldots, a_n^k) \in A(x_k)$, $k = 1, \ldots, t-1$. A *strategy* for player $i \in N$ is a sequence $\pi_i = (\pi_{it})_{t \in \mathbb{N}}$ of Borel measurable transition probabilities from H_t to A_i such that $\pi_{it}(A_i(x_t)) = 1$ for each $h_t \in H_t$. The set of strategies for player $i \in N$ is denoted by Π_i. We let $\Pi := \Pi_1 \times \ldots \times \Pi_n$. Let F_i (F_i^0) be the set of all Borel measurable mappings $f_i : X \times X \to \Pr(A_i)$ $(\phi_i : X \to \Pr(A_i))$ such that $f_i(x^-, x) \in \Pr(A_i(x))$ $(\phi_i(x) \in \Pr(A_i(x)))$ for each $x^-, x \in X$. A *stationary almost Markov strategy* for player $i \in N$ is a constant sequence $(\pi_{it})_{t \in \mathbb{N}}$ where $\pi_{it} = f_i$ for some $f_i \in F_i$ and for all $t \in \mathbb{N}$. If x_t is a state of the game on its t-stage with $t \geq 2$, then player i chooses an action using the mixed strategy $f_i(x_{t-1}, x_t)$. The mixed strategy used at an initial state x_1 is $f_i(x_1, x_1)$. The set of all stationary almost Markov strategies for player $i \in N$ is identified with the set F_i. A stationary Markov strategy for player $i \in N$ is identified with a Borel measurable mapping $f_i \in F_i^0$. We say that $\pi_i = (\pi_{i1}, \pi_{i2}, \ldots) \in \Pi_i$ is a Markov strategy for player i if $\pi_{it} \in F_i^0$ for all $t \in \mathbb{N}$.

Any strategy profile $\pi = (\pi_1, \ldots, \pi_n) \in \Pi$ together with an initial state $x = x_1 \in X$ determines a unique probability measure P_x^π on the space H_∞ of all plays $h_\infty = (x_1, a^1, x_2, a^2, \ldots)$ endowed with the product σ-algebra. The *expected discounted payoff* or *utility function* for player $i \in N$ is

$$J_\beta^{i,T}(x, \pi) = E_x^\pi \left(\sum_{t=1}^{T} \beta^{t-1} u_i(x_t, a^t) \right) \quad \text{where} \quad T \leq \infty.$$

We shall write $J_\beta^i(x, \pi)$, if $T = \infty$.

A profile of strategies $\pi^* \in \Pi$ is called a *Nash equilibrium,* if

$$J_\beta^{i,T}(x, \pi^*) \geq J_\beta^{i,T}(x, (\pi_{-i}^*, \pi_i)) \text{ for all } x \in X,\ \pi_i \in \Pi_i \text{ and } i \in N.$$

A *stationary almost Markov (stationary Markov) perfect equilibrium* is a Nash equilibrium that belongs to the class of strategy profiles $F := F_1 \times \ldots \times F_n$ ($F^0 := F_1^0 \times \ldots \times F_n^0$). A *Markov perfect equilibrium,* on the other hand, is a Nash equilibrium π^*, in which $\pi_{it}^* = f_{it}$ and $f_{it} \in F_i^0$ for every $t \in \mathbb{N}$ and every player $i \in N$. The strategies involved in such an equilibrium are called “markovian,” “state-contigent,” or “payoff-relevant”; see Maskin and Tirole (2001). Clearly, every stationary Markov perfect equilibrium is also a Markov perfect equilibrium.

Let $\pi = (\pi_1, \ldots, \pi_n) \in \Pi$ and $h_t \in H_t$. By $\pi_i[h_t]$ we denote the conditional strategy for player i that can be applied from stage t onward. Let $\pi[h_t] = (\pi_1[h_t], \ldots, \pi_n[h_t])$. Using this notation, one can say that π^* is a *subgame-perfect equilibrium* in the stochastic game if for any $t \in \mathbb{N}$ and every partial history $h_t \in H_t$, $\pi^*[h_t]$ is a Nash equilibrium in the subgame starting at x_t, where x_t is the last coordinate in h_t. This definition refers to the classical idea of Selten (1975).

Let $B(X)$ be the space of all bounded Borel measurable real-valued functions on X and $B^n(X) := B(X) \times \cdots \times B(X)$ (n times). Similarly define $B(X \times X)$ and $B^n(X \times X)$. With any $x \in X$ and $v = (v_1, \ldots, v_n) \in B^n(X)$, we associate the one-shot game $\Gamma_v(x)$ in which the payoff function to player $i \in N$ is

$$U_\beta^i(v_i, x, a) := u_i(x, a) + \beta \int_X v_i(y) q(dy|x, a), \quad a \in A(x). \tag{6.1}$$

If $\nu = (\nu_1, \ldots, \nu_n) \in \Pr(A(x))$, then

$$U_\beta^i(v_i, x, \nu) = \int_{A_n(x)} \ldots \int_{A_1(x)} U_\beta^i(v_i, x, a_1, \ldots, a_n) \nu_1(da_1) \times \cdots \times \nu_n(da_n)$$

and if $f = (f_1, \ldots, f_n) \in F^0$, then $U_\beta^i(v_i, x, f) = U_\beta^i(v_i, x, \nu)$ with $\nu = (f_1(x), \ldots, f_n(x))$. Further, $U_\beta^i(v_i, x, (\mu_i, f_{-i})) = U_\beta^i(v_i, x, \nu)$ with $\nu_i = \mu_i$,

$\nu_j = f_j(x)$ for $j \neq i$. Under our assumptions, $a \to U^i_\beta(v_i, x, a)$ is continuous on $A(x)$ for every $v_i \in B(X)$, $x \in X$, $i \in N$. Let $\mathcal{N}_v(x)$ be the set of all Nash equilibria in the game $\Gamma_v(x)$. By $\mathcal{NP}_v(x)$ we denote the set of payoff vectors corresponding to all Nash equilibria in $\mathcal{N}_v(x)$. Let $\mathcal{M}_v$ be the set of all Borel measurable selections of the set-valued mapping $x \to \mathcal{N}_v(x)$. We know from Proposition 1 that $\mathcal{M}_v \neq \emptyset$.

Consider a T-stage game ($2 \le T < \infty$). Assume that the $(T-1)$-stage subgame starting at any state $x_2 \in X$ has a Markov perfect equilibrium, say π^*_{T-1}. Let v^*_{T-1} be the vector payoff function in $B^n(X)$ determined by π^*_{T-1}. Then we can get some $f^* \in \mathcal{M}_{v^*_{T-1}}$ and define $\pi^*_T := (f^*, \pi^*_{T-1})$. It is obvious that π^*_T is a Markov perfect equilibrium in the T-stage game. This fact was proved by Rieder (1979) and we state it below.

Theorem 1. *Every finite-stage nonzero-sum discounted stochastic game satisfying the above conditions has a subgame-perfect equilibrium. For any $\varepsilon > 0$, there exists an ε-equilibrium π^ε in Markov strategies, i.e.,*

$$J^i_\beta(x, \pi^\varepsilon) + \varepsilon \ge J^i_\beta(x, (\pi_i, \pi^\varepsilon_{-i})) \quad \textit{for all} \quad x \in X,\ \pi_i \in \Pi_i \textit{ and } i \in N.$$

Note that ε-equilibrium in the second part of this theorem has no subgame-perfection property.

We now make an additional assumption.

(A1) The transition probability q is *norm continuous* in actions, i.e., for each $x \in X$, $a^k \to a^0$ in $A(x)$ as $k \to \infty$, it follows that

$$\sup_{D \in \mathcal{B}(X)} |q(D|x, a^k) - q(D|x, a^0)| \to 0.$$

Condition (A1) is quite restrictive, but it is satisfied, if q has a continuous in actions conditional density with respect to some probability measure on X.

Theorem 2. *Every discounted nonzero-sum stochastic game G satisfying* (A1) *has a subgame-perfect equilibrium.*

Theorem 2 was proved in a more general form by Mertens and Parthasarathy (2003), where the payoffs and discount factors may depend on time and the state space is a general measurable space. A special case was considered by Mertens and Parthasarathy (1991), who assumed that the action sets are finite and state independent and transitions are dominated by some probability measure on X. The proofs given in Mertens and Parthasarathy (1991, 2003) are based upon studying a specified fixed point property of an operator defined in the class of measurable selections of compact set-valued mappings from the state space to the payoff space. The fixed point obtained in that class is used to define in a recursive way a subgame-perfect equilibrium that consists of history-dependent strategies (unbounded memory is assumed). For further comments on possible extensions of Theorem 2, the reader is referred to Mertens (2002) and Mertens et al. (2015). A modified proof of their results was provided by Solan (1998), who analyzed accumulation points of ϵ-equilibria (as $\epsilon \to 0$) obtained in Theorem 1.

Assume that $A_i(x) = A_i$ for each $x \in X$ and $i \in N$ and that every space A_i is compact. Let X and $A_1, \ldots, A_n$ be given the discrete topology. According to Maitra and Sudderth (2007), a function $g : H_\infty \to \mathbb{R}$ is *DS-continuous* on H_∞ if it is continuous on H_∞ endowed with the product topology. It is easy to see that g is DS-continuous on H_∞ if and only if, for any $\epsilon > 0$ and $y = (y_1, y_2, \ldots) \in H_\infty$, there exists m such that $|g(y) - g(y')| < \epsilon$ for each $y' = (y'_1, y'_2, \ldots) \in H_\infty$ such that $y_l = y'_l$ for $1 \leq l \leq m$. Suppose that $g_i : H_\infty \to \mathbb{R}$ is a bounded Borel measurable payoff function for player $i \in N$. For any strategy profile $\pi \in \Pi$ and every initial state $x = x_1$, the expected payoff to player i is $E_x^\pi(g_i)$. The subgame-perfect equilibrium can be defined for this game in the usual way. Maitra and Sudderth (2007) (see Theorem 1.2) obtained a general theorem on the existence of subgame-perfect equilibria for stochastic games.

Theorem 3. *Let the payoff functions g_i, $i \in N$ be bounded, Borel measurable, and DS-continuous on H_∞ and let the action spaces A_i, $i \in N$ be finite. Then the game has a subgame-perfect equilibrium.*

The proof of Theorem 3 applies some techniques from gambling theory described in Dubins and Savage (2014), i.e., approximations of DS-continuous functions by "finitary functions". Theorem 3 extends a result due to Fudenberg and Levine (1983). An example given in Harris et al. (1995) shows that Theorem 3 is false, if the action spaces are compact metric and the transition probability q is *weakly continuous*.

The next result was proved by Maitra and Sudderth (2007) (see Theorem 1.3) for "additive games" and sounds as follows.

Theorem 4. *Assume that every action space is compact and the transition probability satisfies* (A1). *Assume that $g_i(h_\infty) = \sum_{t=1}^{\infty} r_{it}(x_t, a^t)$ and this series converges uniformly on H_∞. If, in addition, every function r_{it} is bounded, $r_{it}(\cdot, a)$ is Borel measurable on X for each $a \in A := A_1 \times \cdots \times A_n$ and $r_{it}(x, \cdot)$ is continuous on A for each $x \in X$, then the game has a subgame-perfect equilibrium.*

It is worth to emphasize that stationary Markov perfect equilibria may not exist in games considered in this section. Namely, Levy (2013) gave a counterexample of a discounted game with uncountable state space, finite action sets and deterministic transitions. Then, Levy and McLennan (2015) showed that stationary Markov perfect equilibria may not exist even if the action spaces are finite, $X = [0, 1]$ and the transition probability has a density function with respect to some measure $\mu \in \Pr(X)$. A simple modification of the example given in Levy and McLennan

(2015) shows that a new game (with $X = [0, 2]$) need not have a stationary Markov perfect equilibrium, when the measure μ (dominating the transition probability q) is nonatomic.

4 Correlated Equilibria with Public Signals in Games with Borel State Spaces

Correlated equilibria for normal form games were first studied by Aumann (1974, 1987). In this section we describe an extensive-form correlated equilibrium with public randomization inspired by the work of Forges (1986). A further discussion of correlated equilibria and communication in games can be found in Forges (2009). The sets of all equilibrium payoffs in extended form games that include a general communication device are characterized by Solan (2001).

We now extend the sets of strategies available to the players in the sense that we allow them to correlate their choices in some natural way. Suppose that $(\xi_t)_{t\in\mathbb{N}}$ is a sequence of so-called signals, drawn independently from $[0, 1]$ according to the uniform distribution. Suppose that at the beginning of each period t of the game the players are informed not only of the outcome of the preceding period and the current state x_t, but also of ξ_t. Then, the information available to them is a vector $h^t = (x_1, \xi_1, a^1, \ldots, x_{t-1}, \xi_{t-1}, a^{t-1}, x_t, \xi_t)$, where $x_\tau \in X$, $\xi_\tau \in [0, 1]$, $a^\tau \in A(x_\tau)$, $1 \le \tau \le t-1$. We denote the set of such vectors by H^t. An extended strategy for player i is a sequence $\pi_i = (\pi_{it})_{t\in\mathbb{N}}$, where π_{it} is a Borel measurable transition probability from H^t to A_i such that $\pi_{it}(A_i(x_t)|h^t) = 1$ for each $h^t \in H^t$. An extended stationary strategy for player $i \in N$ can be identified with a Borel measurable mapping $f : X \times [0, 1] \to \Pr(A_i)$ such that $f(A_i(x)|x, \xi) = 1$ for all $(x, \xi) \in X \times [0, 1]$. Assuming that the players use extended strategies we actually assume that they play the stochastic game with the extended state space $X \times [0, 1]$. The law of motion, say p, in the extended state space model is obviously the product of the original law of motion q and the uniform distribution η on $[0, 1]$. More precisely, for any $x \in X$, $\xi \in [0, 1]$, $a \in A(x)$, Borel sets $C \subset X$ and $D \subset [0, 1]$, $p(C \times D|x, \xi, a) = q(C|x, a)\eta(D)$. For any profile of extended strategies $\pi = (\pi_1, \ldots, \pi_n)$ of the players, the expected discounted payoff to player $i \in N$ is a function of the initial state $x_1 = x$ and the first signal $\xi_1 = \xi$ and is denoted by $J^i_\beta(x, \xi, \pi)$. We say that $f^* = (f^*_1, \ldots, f^*_n)$ is a Nash equilibrium in the β-discounted stochastic game in the class of extended strategies if for each initial state $x_1 = x$, $i \in N$ and every extended strategy π_i of player i, we have

$$\int_0^1 J^i_\beta(x, \xi, f^*)d\xi \ge \int_0^1 J^i_\beta(x, \xi, (\pi_i, f^*_{-i}))d\xi.$$

The Nash equilibrium in extended strategies is also called a *correlated equilibrium with public signals*. The reason is that after the outcome of any period of the game, the players can coordinate their next choices by exploiting the next (known to all of them, i.e., public) signal and using some coordination mechanism telling which (pure or mixed) action is to be played by everyone. In many applications, we are particularly interested in stationary equilibria. In such a case the coordination mechanism can be represented by a family of $n+1$ Borel measurable functions $\lambda^j : X \to [0,1]$ such that $\sum_{j=1}^{n+1} \lambda^j(x) = 1$ for each $x \in X$. A stationary correlated equilibrium can be constructed then by using a family of $n+1$ stationary strategies $f_i^1, \ldots, f_i^{n+1}$ given for every player i, and the following coordination rule. If the game is in state $x_t = x$ on stage t and a random number $\xi_t = \xi$ is selected, then player $i \in N$ is suggested to use $f_i^k(\cdot|x)$ where k is the least index for which $\sum_{j=1}^{k} \lambda^j(x) \geq \xi$. The functions λ^j and f_i^j induce an extended stationary strategy f_i^* for every player i as follows

$$f_i^*(\cdot|x,\xi) := f_i^1(\cdot|x) \quad \text{if} \quad \xi \leq \lambda^1(x), \;\; x \in X,$$

and

$$f_i^*(\cdot|x,\xi) := f_i^k(\cdot|x) \quad \text{if} \quad \sum_{j=1}^{k-1} \lambda^j(x) < \xi \leq \sum_{j=1}^{k} \lambda^j(x)$$

for $x \in X$, $2 \leq k \leq n+1$. Because the signals are independent and uniformly distributed in $[0,1]$, it follows that at any period of the game and for any current state x, the number $\lambda_j(x)$ can be interpreted as the probability that player i is suggested to use $f_i^j(\cdot|x)$ as a mixed action.

(A2) Let $\mu \in \Pr(X)$. There exists a conditional density function ρ for q with respect to μ such that if $a^k \to a^0$ in $A(x)$, $x \in X$, as $k \to \infty$, then

$$\lim_{k\to\infty} \int_X |\rho(x,a^k,y) - \rho(x,a^0,y)|\mu(dy) = 0.$$

Theorem 5. *Any discounted stochastic game G satisfying* (A2) *has a stationary correlated equilibrium with public signals.*

Theorem 5 was proved by Nowak and Raghavan (1992). First it is shown by making use of a theorem in Glicksberg (1952) that the correspondence $v \to \mathcal{M}_v$ has a fixed point, i.e., there exists $w^* \in B^n(X)$ such that $w^*(x) \in co\mathcal{NP}_{w^*}(x)$ for all $x \in X$. Then, applying Propositions 2 and 4, one can prove the existence of a stationary correlated equilibrium with public signals for the game with the payoff functions $U_\beta^i(w_i^*, x, a)$ defined in (6.1). A verification that f^* obtained in this way is indeed a Nash equilibrium in the game with the extended state space $X \times [0,1]$ relies on using standard Bellman equations for discounted dynamic programming; see Blackwell (1965) or Puterman (1994). Observe also that the set of all atoms D_a for μ is countable. A refinement of the above result is Theorem 2 in Jaśkiewicz and Nowak (2016), where it is shown that public signals are important only in states belonging to the set $X \setminus D_a$. A similar result on correlated equilibria was given in

Nowak and Jaśkiewicz (2005) for semi-Markov games with Borel state spaces and the expected average payoffs. This result, in turn, was proved under geometric drift conditions (GE1)–(GE3) formulated in Sect. 5 in Jaśkiewicz and Nowak (2018a).

Condition (A2) can be replaced in the proof (with minor changes) by assumption (A1) on norm continuity of q with respect to actions. A similar result to Theorem 5 was given by Duffie et al. (1994), where it was assumed that for any $x, x' \in X$, $a \in A(x)$, $a' \in A(x')$, we have

$$q(\cdot|x,a) \ll q(\cdot|x',a') \quad \text{and} \quad q(\cdot|x',a') \ll q(\cdot|x,a).$$

In addition, Duffie et al. (1994) required the continuity of the payoffs and transitions with respect to actions. Thus, the result in Duffie et al. (1994) is weaker than Theorem 5. Moreover, they also established the ergodicity of the Markov chain induced by a stationary correlated equilibrium. Their proof is different from that of Nowak and Raghavan (1992). Subgame-perfect correlated equilibria were also studied by Harris et al. (1995) for games with weakly continuous transitions and general continuous payoff functions on the space of infinite plays endowed with the product topology. Harris et al. (1995) gave an example showing that public signals play an important role. They proved that the subgame-perfect equilibrium path correspondence is upper hemicontinuous. Later, Reny and Robson (2002) provided a shorter and simpler proof of existence that focuses on considerations of equilibrium payoffs rather than paths. Some comments on correlated equilibria for games with finitely many states or different payoff evaluation will be given in the sequel.

5 Stationary Equilibria in Discounted Stochastic Games with Borel State Spaces

In this section, we introduce the following condition.

(A3) There exist l Carathéodory functions $\alpha_j : X \times A \to [0,1]$ such that $\sum_{j=1}^{l} \alpha_j(x,a) = 1$ for every $(x,a) \in X \times A$ and Borel measurable transition probabilities $q_j : X \times \mathcal{B}(X) \to [0,1]$ such that

$$q(\cdot|x,a) = \sum_{j=1}^{l} \alpha_j(x,a) q_j(\cdot|x), \quad (x,a) \in X \times A.$$

Additionally, every $q_j(\cdot|x)$ is dominated by some $\mu \in \Pr(X)$.

We can now state a result due to Jaśkiewicz and Nowak (2016).

Theorem 6. *Assume that game G satisfies* (A3). *Then, G has a stationary almost Markov perfect equilibrium.*

We *outline the proof* of Theorem 6 for nonatomic measure μ. The general case needs an additional notation. First, we show that there exists a Borel measurable mapping $w^* \in B^n(X)$ such that $w^*(x) \in co\mathcal{NP}_{w^*}(x)$ for all $x \in X$. This result is obtained by applying a generalization of the Kakutani fixed point theorem due to Glicksberg (1952). (Note that closed balls in $B^n(X)$ are compact in the weak-star topology due to Banach-Alaoglu's theorem.) Second, applying Proposition 5 we conclude that there exists some $v^* \in B^n(X \times X)$ such that

$$\int_X w^*(y)q_j(dy|x) = \int_X v^*(x,y)q_j(dy|x), \quad j = 1, \ldots, l.$$

Hence, by (A3) we infer that

$$\int_X w^*(y)q(dy|x,a) = \int_X v^*(x,y)q(dy|x,a), \quad (x,a) \in X \times A.$$

Moreover, we know that $v^*(x,y) \in \mathcal{NP}_{v^*}(y)$ for all states x and y. Furthermore, making use of Filippov's measurable implicit function theorem (as in Proposition 5), we claim that $v^*(x,y)$ is the vector of equilibrium payoffs corresponding to some stationary almost Markov strategy profile. Finally, we utilize the system of n Bellman equations to provide a characterization of stationary equilibrium and to deduce that this profile is indeed a stationary almost Markov perfect equilibrium. For details the reader is referred to Jaśkiewicz and Nowak (2016).

Corollary 1. *Consider a game where the set A is finite and the transition probability q is Borel measurable. Then, the game has a stationary almost Markov perfect equilibrium.*

Proof. We show that the game meets (A3). Let $m \in \mathbb{N}$ be such that $A = \{a^1, \ldots, a^m\}$. Now, for $j = 1, \ldots, m$, define

$$\alpha_j(s,a) := \begin{cases} 1, \text{ if } \; a \in A(x), \; a = a^j \\ 0, \text{ otherwise,} \end{cases} \qquad \text{and}$$

$$q_j(\cdot|x) := \begin{cases} q(\cdot|x,a), \text{ if } \; a \in A(x), \; a = a^j \\ \mu(\cdot), \qquad \text{otherwise.} \end{cases}$$

Then, $q(\cdot|s,a) = \sum_{j=1}^{l} g_j(s,a)q_j(\cdot|s)$ for $l = m$ and the conclusion follows from Theorem 6.

Remark 1. Corollary 1 extends the result of Mertens and Parthasarathy (1991), where it is additionally assumed that $A_i(x) = A_i$ for all $x \in X$, $i \in N$ and that μ is nonatomic; see Comment on page 147 in Mertens and Parthasarathy (1991) or Theorem VII.1.8 on page 398 in Mertens et al. (2015). If μ admits some atoms, then they proved the existence of a subgame-perfect equilibrium in which the strategy of player $i \in N$ is of the form $(f_{i1}, f_{i2}, \ldots)$ with $f_{it} \in F_i^0$ for each $t \in \mathbb{N}$. Thus, the equilibrium strategy of player $i \in N$ is stage-dependent.

Remark 2. It is worth to emphasize that equilibria established in Theorem 6 are subgame-perfect. A related result to Theorem 6 is given in Barelli and Duggan (2014). The assumption imposed on the transition probability in their paper is weaker, but an equilibrium is shown to exist in the class of stationary semi-Markov strategies, where the players take into account the current state, the previous state and the actions chosen by the players in the previous state.

Remark 3. As already mentioned in Sect. 3, Levy and McLennan (2015) constructed a stochastic game that does not have a stationary Markov perfect equilibrium. In their model, each set A_i is finite, $A_i(x) = A_i$ for every $i \in N$, $x \in X$, and the transition law is a convex combination of a probability measure (depending on the current state) and the Dirac measure concentrated at some state. Such a model satisfies the absolute continuity condition. Hence, their example confirms that one cannot expect to obtain an equilibrium in stationary Markov strategies even for games with finite action spaces. Therefore, Corollary 1 is meaningful.

Remark 4. By Urysohn's metrization theorem (see Theorem 3.40 in Aliprantis and Border 2006), every action space A_i can be embedded homeomorphically in the Hilbert cube. The action correspondences remain measurable and compact-valued after the embedding. Therefore, one can assume without loss of generality as in Jaśkiewicz and Nowak (2016) that the action spaces are compact metric.

A stochastic game with *additive reward and additive transitions* (ARAT for short) satisfies some separability condition for the actions of the players. To simplify presentation, we assume that $N = \{1, 2\}$. The payoff function for player $i \in N$ is of the form

$$u_i(x, a_1, a_2) = u_{i1}(x, a_1) + u_{i2}(x, a_2),$$

where $x \in X,\ a_1 \in A_1(x),\ a_2 \in A_2(x)$ and similarly

$$q(\cdot|x, a_1, a_2) = q_1(\cdot|x, a_1) + q_2(\cdot|x, a_2),$$

where q_1 and q_2 are Borel measurable subtransition probabilities dominated by some $\mu \in \Pr(X)$.

The following result was proved in Jaśkiewicz and Nowak (2015a).

Theorem 7. *If μ is a nonatomic probability measure and the action sets A_1 and A_2 are finite, then the ARAT stochastic game has a Nash equilibrium in pure stationary almost Markov strategies.*

The separability condition as in ARAT games can be easily generalized to n-person case. Assumptions of similar type are often used in differential games; see Başar and Olsder (1995). ARAT stochastic games with Borel state and finite action spaces were first studied by Himmelberg et al. (1976), who showed the existence of stationary Markov equilibria for μ-almost all initial states with $\mu \in \Pr(X)$. Their result was strengthened by Nowak (1987), who considered compact metric action spaces and obtained stationary equilibria for all initial states. Pure stationary Markov perfect equilibria may not exist in ARAT stochastic games if μ has atoms; see Example 3.1 (a game with 4 states) in Raghavan et al. (1985) or counterexample (a game with 2 states) in Jaśkiewicz and Nowak (2015a). Küenle (1999) studied ARAT stochastic games with a Borel state space and compact metric action spaces and established the existence of non-stationary history-dependent pure Nash equilibria. In order to construct subgame-perfect equilibria, he used the well-known idea of threats (frequently used in repeated games). The result of Küenle (1999) is stated for two-person games only. Theorem 7 can also be proved for n-person games under a similar additivity assumption. An almost Markov equilibrium is obviously subgame-perfect.

Stationary Markov perfect equilibria exist in discounted stochastic games with state-independent transitions (SIT games) studied by Parthasarathy and Sinha (1989). They assumed that $A_i(x) = A_i$ for all $x \in X$ and $i \in N$, the action sets A_i are finite, and $q(\cdot|x, a) = q(\cdot|a)$ are nonatomic for all $a \in A$. A more general class of games with additive transitions satisfying (A3) but with all q_j independent of state $x \in X$ (AT games) was examined by Nowak (2003b). A stationary Markov perfect equilibrium $f^* \in F^0$ was shown to exist in that class of stochastic games. Additional special classes of discounted stochastic games with uncountable state space having stationary Markov perfect equilibrium are described in Krishnamurthy et al. (2012). Some of them are related to AT games studied by Nowak (2003b).

Let $X = Y \times Z$ where Y and Z are Borel spaces. In a *noisy stochastic game* considered by Duggan (2012), the states are of the form $x = (y, z) \in X$, where z is called a noise variable. The payoffs depend measurably on $x = (y, z)$ and are continuous in actions of the players. The transition probability q is defined as follows:

$$q(D|x, a) = \int_Y \int_Z 1_D(y', z') q_2(dz'|y') q_1(dy'|x, a), \quad a \in A(x),\ D \in \mathcal{B}(Y \times Z).$$

Moreover, it is assumed that q_1 is dominated by some $\mu_1 \in \Pr(Y)$ and q_2 is absolutely continuous with respect to some *nonatomic measure* $\mu_2 \in \Pr(Z)$. Additionally, $q_1(\cdot|x, a)$ is norm continuous in actions $a \in A$, for each $x \in X$. This form of q implies that conditional on y' the next *shock* z' is independent of the current state and actions. In applications, (y, z) may represent a pair: the price of some good and the realization of random demand. By choosing actions, the players can determine (stochastically) the next period price y', which in turn, has some influence on the next demand shock. Other applications are discussed by Duggan (2012), who obtained the following result.

Theorem 8. *Every noisy stochastic game has a stationary Markov perfect equilibrium.*

Let X be a Borel space, $\mu \in \Pr(X)$ and let $\mathcal{G} \subset \mathcal{B}(X)$ be a sub-σ-algebra. A set $D \in \mathcal{B}(X)$ is said to be a (conditional) $\mathcal{G}$-*atom* if $\mu(D) > 0$ and for any Borel set $B \subset D$, there exists some $D_0 \in \mathcal{G}$ such that $\mu(B \triangle (D \cap D_0)) = 0$. Assume that the transition probability q is dominated by some probability measure μ, and ρ denotes a conditional density function. Following He and Sun (2017), we say that a discounted stochastic game has a *decomposable coarser transition kernel* if there exists a sub-σ-algebra $\mathcal{G} \subset \mathcal{B}(X)$ such that $\mathcal{B}(X)$ has no $\mathcal{G}$-atom and there exist Borel measurable nonnegative functions ρ_j and d_j $(j = 1, \ldots, l)$ such that, for every $x \in X$, $a \in A$, each function $\rho_j(\cdot, x, a)$ is $\mathcal{G}$-measurable and the transition probability density ρ is of the form

$$\rho(y, x, a) = \sum_{j=1}^{l} \rho_j(y, x, a) d_j(y), \quad x,\ y \in X,\ a \in A.$$

Using a theorem of Dynkin and Evstigneev (1977) on conditional expectations of measurable correspondences and a fixed point property proved in Nowak and Raghavan (1992), He and Sun (2017) established the following result.

Theorem 9. *Every discounted stochastic game having decomposable coarser transition kernel with respect to a nonatomic probability measure μ on X has a stationary Markov perfect equilibrium.*

A slight extension of the above theorem, given in He and Sun (2017), contains as special cases the results proved by Parthasarathy and Sinha (1989), Nowak (2003b), Nowak and Raghavan (1992). However, the form of the equilibrium strategy obtained by Nowak and Raghavan (1992) does not follow from He and Sun (2017). The result of He and Sun (2017) also covers the class of noisy stochastic games examined in Duggan (2012). In this case, it suffices to take $\mathcal{G} \subset \mathcal{B}(Y \times Z)$ that consists of the sets $D \times Z$, $D \in \mathcal{B}(Y)$. Finally, we wish to point out that ARAT discounted stochastic games as well as games considered in Jaśkiewicz and Nowak (2016) (see Theorem 6) are not included in the class of models mentioned in Theorem 9.

Remark 5. A key tool in proving the existence of a stationary Markov perfect equilibrium in a discounted stochastic game that has no ARAT structure is Lyapunov's theorem on the range of a nonatomic vector measure. Since the Lyapunov theorem is false for infinitely many measures, the counterexample of Levy's eight-person game is of some importance (see Levy and McLennan 2015). There is another reason for which the existence of an equilibrium in the class of stationary Markov strategies F^0 is difficult to obtain. One can recognize strategies from the sets

F_i^0 as "Young measures" and consider natural in that class weak-star topology; see Valadier (1994). Young measures are often called relaxed controls in control theory. With the help of Example 3.16 from Elliott et al. (1973), one can easily construct a stochastic game with $X = [0,1]$, finite action spaces and trivial transition probability q being a Lebesgue measure on X, where the expected discounted payoffs $J_\beta^i(x, f)$ are discontinuous on F^0 endowed with the product topology. The continuity of $f \to J_\beta^i(x, f)$ (for fixed initial state) can only be proved for ARAT games. Generally, it is difficult to obtain compact families of continuous strategies. This property requires very strong conditions in order to get, for instance, equicontinuous family of functions (see Sect. 6).

6 Special Classes of Stochastic Games with Uncountable State Space and Their Applications in Economics

In a number of applications of discrete-time dynamic games in economics, the state space is an interval in Euclidean space. An illustrative example is the "fish war game" studied by Levhari and Mirman (1980), where the state space $X = [0,1]$, $A_i(x) = [0, x/n]$ for each $i \in N$. Usually, X is interpreted as the set of common property renewable resources. If x_t is a resource stock at the beginning of period $t \in \mathbb{N}$ and player $i \in N$ extracts $a_{it} \in A_i(x_t)$ for consumption, then the new state is $x_{t+1} = \left(x_t - \sum_{j=1}^n a_{jt}\right)^\alpha$ with $\alpha \in (0,1)$. The game is symmetric in the sense that the utility function of player $i \in N$ is: $u_i(x,a) := \ln a_i$ with $a = (a_1, \ldots, a_n)$ being a pure strategy profile chosen by the players in state $x \in X$. Levhari and Mirman (1980) constructed a symmetric stationary Markov perfect equilibrium for the *two*-player β-discounted game that consists of linear strategies. For the arbitrary n-player case, the equilibrium strategy profile is $f^\beta = (f_1^\beta, \ldots, f_n^\beta)$ where $f_i^\beta(x) = \frac{(1-\alpha\beta)x}{n+(1-n)\alpha\beta}$, $x \in X$, $i \in N$; see Nowak (2006c). Levhari and Mirman (1980) concluded that, in equilibrium, the fish population will be smaller than the population that would have resulted if the players had cooperated and had maximized their joint utility. The phenomenon of overexploitation of a common property resource is known in economics as the "tragedy of the commons." Dutta and Sundaram (1993) showed that there may exist equilibria (that consist of discontinuous consumption functions), in which the common resource is underexploited, so that the tragedy of the commons need not occur. A characterization of the set of equilibria in this model has been given by Chiarella et al. (1984). If $\beta \to 1$, then $f^\beta \to f^* = (f_1^*, \ldots, f_n^*)$ where $f_i^*(x) = \frac{(1-\alpha)x}{n+(1-n)\alpha}$, $x \in X$, $i \in N$. As shown in Nowak (2008), f^* is a Nash equilibrium in the class of all strategies of the players in the fish war game under the *overtaking optimality criterion*. Such a criterion was examined in economics by Ramsey (1928), von Weizsäcker (1965), and Gale (1967), and its application to repeated games was pointed out by Rubinstein (1979). Generally, finding an equilibrium under the overtaking optimality criterion in the class of all strategies is a difficult task; see Carlson and Haurie (1996).

Dutta and Sundaram (1992) considered a stochastic game of resource extraction with state space $X = [0, \infty)$, $A_i(x) = [0, x/n]$ for each $i \in N$, $x \in X$ and the same nonnegative utility function u for each player. Their model includes both the dynamic game with deterministic transitions studied by Sundaram (1989a,b) and the stochastic game with nonatomic transition probabilities considered by Majumdar and Sundaram (1991). Now we list the assumptions used by Dutta and Sundaram (1992). For any $y, z \in X$, let $Q(y|z) := q([0, y]|z)$ and for any $y > 0$ set $Q(y^-|z) := \lim_{y' \uparrow y} Q(y'|z)$.

(D1) For any $x \in X$, $a = (a_1, \ldots, a_n) \in A(x)$ and $i \in N$, $u_i(x, a) = u(a_i) \geq 0$. The utility function u is strictly concave, increasing and continuously differentiable. Moreover, $\lim_{a \downarrow 0} u'(a) = \infty$.
(D2) $Q(0|0) = 1$ and for each $z > 0$, there exists a compact interval $I(z) \subset (0, \infty)$ such that $q(I(z)|z) = 1$.
(D3) There exists $z_1 > 0$ such that if $0 < z < z_1$, then $Q(z^-|z) = 0$, i.e., $q([z, \infty)|z) = 1$.
(D4) There exists $\hat{z} > 0$ such that for each $z \geq \hat{z}$, $Q(z|z) = 1$, i.e., $q([0, z]|z) = 1$.
(D5) If $z_m \to z$ as $m \to \infty$, then $q(\cdot|z_m) \to q(\cdot|z)$ in the weak topology on $\Pr(X)$.
(D6) If $z < z'$, then for each $y > 0$, $Q(y^-|z) \geq Q(y|z')$.

Assumption (D6) is a "strong stochastic dominance" condition that requires larger investments to obtain probabilistically higher stock levels. This assumption and the fact that the players have identical utility functions play a crucial role in the proof of Theorem 1 in Dutta and Sundaram (1992) that can be stated as follows.

Theorem 10. *Every discounted stochastic game satisfying conditions* (D1)–(D6) *has a pure stationary Markov perfect equilibrium.*

Remark 6. The equilibrium strategies obtained by Dutta and Sundaram (1992) are identical for all the players, and the corresponding equilibrium functions are nondecreasing and upper semicontinuous on X. One can observe that the assumptions on the transition probability functions include the usual deterministic case with an increasing continuous production function. A slightly more general model was recently studied by Jaśkiewicz and Nowak (2018b), who dealt with unbounded utilities.

Transition probabilities considered in other papers on equilibria in stochastic games are assumed to satisfy much stronger continuity conditions, e.g., the norm continuity in actions.

The problem of proving the existence of a Nash equilibrium in a stochastic game of resource extraction with different utility functions for the players seems to be difficult. Partial results were reported by Amir (1996a), Nowak (2003b), Balbus and Nowak (2008) and Jaśkiewicz and Nowak (2015b), where specific transition structures were assumed. Below we give an example, where the assumptions are relatively simple to list.

(S1) $X = [0, \infty)$ and $A_i(x) = [0, b_i(x)]$ with $\sum_{j=1}^{n} b_j(x) \leq x$ for all $x \in X$, where each b_j is a continuous increasing function.

(S2) $u_i : [0, \infty) \to \mathbb{R}$ is a nonnegative increasing twice differentiable utility function for player $i \in N$ such that $u_i(0) = 0$.

(S3) If $a = (a_1, \ldots, a_n) \in A(x)$ and $s(a) = \sum_{i=1}^{n} a_i$, then

$$q(\cdot|x, a) = h(x - s(a))q_0(\cdot|x) + (1 - h(x - s(a)))\delta_0(\cdot),$$

where $h : X \to [0, 1]$ is an increasing twice differentiable function such that $h'' < 0$ and $h(0) = 0$, δ_0 is the Dirac measure concentrated at the point $0 \in X$. Moreover, $q_0((0, \infty)|x) = 1$ for each $x > 0$, $q_0(\{0\}|0) = 1$ and $q_0(\cdot|x)$ has a density function $\rho(x, \cdot)$ with respect to a σ-finite measure μ defined on X. The function $x \to \rho(x, y)$ is continuous for each $y \in X$.

The following result is a special case of Theorem 2 in Jaśkiewicz and Nowak (2015b).

Theorem 11. *Every discounted stochastic game satisfying assumptions* (S1)–(S3) *has a pure stationary Markov perfect equilibrium.*

The proof of Theorem 11 uses the fact that the auxiliary game $\Gamma_v(x)$ has a unique Nash equilibrium for any vector $v = (v_1, \ldots, v_n)$ of nonnegative continuation payoffs v_i such that $v_i(0) = 0$. The uniqueness follows from page 1476 in Balbus and Nowak (2008) or can be deduced from the classical theorem of Rosen (1965) (see also Theorem 3.6 in Haurie et al. 2012). The game $\Gamma_v(x)$ is not supermodular since for increasing continuation payoffs v_i such that $v_i(0) = 0$, we have $\frac{\partial^2 U^i_\beta(v_i, x, a)}{\partial a_i \partial a_j} < 0$, for $i \neq j$. A stronger version of Theorem 11 and related results can be found in Jaśkiewicz and Nowak (2015b).

Transition probabilities presented in (S3) were first used in Balbus and Nowak (2004). They dealt with the symmetric discounted stochastic games of resource extraction and proved that the sequence of Nash equilibrium payoffs in the n-stage games converges monotonically as $n \to \infty$. Stochastic games of resource extraction without the symmetry condition were first examined by Amir (1996a), who considered so-called convex transitions. More precisely, he assumed that the conditional cumulative distribution function $Q(y|z)$ is strictly convex with respect to $z \in X$ for every fixed $y > 0$. He proved the existence of pure stationary Markov perfect equilibria, which are Lipschitz continuous functions in the state variable. Although the result obtained is strong, a careful analysis of various examples suggests that the convexity assumption imposed by Amir (1996a) is satisfied very rarely. Usually, the cumulative distribution $Q(y|z)$ is neither convex nor concave with respect to z. A further discussion on this condition is provided in Remarks 7–8 in Jaśkiewicz and Nowak (2015b). The function $Q(y|z)$ induced by the transition probability q of the form considered in (S3) is strictly concave in $z = x - s(a)$ only when q_0 is independent of $x \in X$. Such transition probabilities that are "mixtures" of finitely many probability measures on X were considered in Nowak (2003b) and Balbus and Nowak (2008). A survey of various game theoretic approaches to resource extraction models can be found in Van Long (2011).

In many other examples, the one-shot game $\Gamma_v(x)$ has also nonempty compact set of pure Nash equilibria. Therefore, a counterpart of Theorem 6 can be formulated for the class of pure strategies of the players. We now describe some examples presented in Jaśkiewicz and Nowak (2016).

Example 1 (Dynamic Cournot oligopoly). Let $X = [0, \bar{x}]$ and $x \in X$ represent a realization of a random *demand shock* that is modified at each period of the game. Player $i \in N$ (oligopolist) sets a production quantity $a_i \in A_i(x) = [0, 1]$. If $P\left(x, \sum_{j=1}^n a_j\right)$ is the inverse demand function, and $c_i(x, a_i)$ is the cost function for player i, then

$$u_i(x, a) := a_i P\left(x, \sum_{j=1}^{n} a_j\right) - c_i(x, a_i), \quad a = (a_1, \ldots, a_n).$$

A simple example of the inverse demand function is

$$P\left(x, \sum_{j=1}^{n} a_j\right) = x\left(n - \sum_{j=1}^{n} a_j\right).$$

The function $a_i \to a_i P\left(x, \sum_{j=1}^n a_j\right)$ is usually concave and $a_i \to c_i(x, a_i)$ is often convex. Assume that

$$q(\cdot|x, a) = (1 - \overline{a})q_1(\cdot|x) + \overline{a}q_2(\cdot|x), \quad \overline{a} := \frac{1}{n}\sum_{j=1}^{n} a_j,$$

where $q_1(\cdot|x)$ and $q_2(\cdot|x)$ are dominated by some probability measure μ on X for all $x \in X$. In order to provide an interpretation of q, we observe that

$$q(\cdot|x, a) = q_1(\cdot|x) + \overline{a}(q_2(\cdot|x) - q_1(\cdot|x)). \tag{6.2}$$

Let

$$E_q(x, a) := \int_X y q(dy|x, a) \quad \text{and} \quad E_{q_j}(x) := \int_X y q_j(dy|x)$$

be the mean values of the distributions $q(\cdot|x,a)$ and $q_j(\cdot|x)$, respectively. By (6.2), we have $E_q(x,a) := E_{q_1}(x) + \overline{a}(E_{q_2}(x) - E_{q_1}(x))$. Assume that $E_{q_1}(x) \geq x \geq E_{q_2}(x)$. This condition implies that

$$E_{q_2}(x) - E_{q_1}(x) \leq 0.$$

Thus, the expectation of the next demand shock $E_q(x,a)$ decreases if the total sale $n\overline{a}$ in the current state $x \in X$ increases. Observe that the game $\Gamma_v(x)$ is concave. From Nash (1950), it follows that the game $\Gamma_v(x)$ has a pure equilibrium point. However, the set of Nash equilibria in $\Gamma_v(x)$ may contain many points. A modification of the proof of Theorem 1 given in Jaśkiewicz and Nowak (2016) implies that this game has a pure stationary almost Markov perfect equilibrium.

Example 2 (Cournot competition with substituting goods in differentiated markets). This model is inspired by a dynamic game with complementary goods studied by Curtat (1996). Related static games were already discussed in Spence (1976) and Vives (1990). There are n firms on the market and firm $i \in N$ produces a quantity $a_i \in A_i(x) = [0,1]$ of a differentiated product. The inverse demand function is given by a twice differentiable function $P_i(a)$, where $a = (a_1,\ldots,a_n)$. The goods are *substitutes*, i.e., $\frac{\partial P_i(a)}{\partial a_j} < 0$ for all $i, j \in N$, see Spence (1976). In other words, consumption of one good will decrease consumption of the others. We assume that $X = [0,1]^n$, where i-th coordinate $x_i \in [0,1]$ is a measure of the cumulative experience of firm $i \in N$. If $c_i(x_i)$ is the marginal cost for firm $i \in N$, then

$$u_i(x,a) := a_i\left[P_i(a) - c_i(x_i)\right], \quad a = (a_1,\ldots,a_n), \quad x = (x_1,\ldots,x_n) \in X. \tag{6.3}$$

The transition probability of the next state (experience vector) is of the form:

$$q(\cdot|x,a) = h\left(\sum_{j=1}^{n}(x_j + a_j)\right) q_2(\cdot|x) + \left(1 - h\left(\sum_{j=1}^{n}(x_j + a_j)\right)\right) q_1(\cdot|x), \tag{6.4}$$

where

$$h\left(\sum_{j=1}^{n}(x_j + a_j)\right) = \frac{\sum_{j=1}^{n} x_j + \sum_{j=1}^{n} a_j}{2n} \tag{6.5}$$

and $q_1(\cdot|x)$, $q_2(\cdot|x)$ are dominated by some probability measure μ on X for all $x \in X$. In Curtat (1996) it is assumed that q_1 and q_2 are independent of $x \in X$ and also that q_2 stochastically dominates q_1. Then, the underlying Markov process governed by q captures the ideas of learning-by-doing and spillover (see page 197 in Curtat 1996). Here, this stochastic dominance condition can be dropped, although it is quite natural. It is easy to see that the game $\Gamma_v(x)$ is concave, if $u_i(x,(\cdot,a_{-i}))$ is concave on $[0,1]$. Clearly, this is satisfied, if for each $i \in N$, we have

$$2\frac{\partial P_i(a)}{\partial a_i} + \frac{\partial^2 P_i(a)}{\partial a_i^2} a_i < 0.$$

If the goods are substitutes, this condition holds, when $\frac{\partial^2 P_i(a)}{\partial a_i^2} \leq 0$ for all $i \in N$. The game $\Gamma_v(x)$ may have multiple pure Nash equilibria. Using the methods from Jaśkiewicz and Nowak (2016), one can show that any concave game discussed here has a pure stationary almost Markov perfect equilibrium.

Supermodular static games were extensively studied by Milgrom and Roberts (1990) and Topkis (1998). This class of games finds applications in dynamic economic models with complementarities. Our next illustration refers to Example 2, but with products that are *complements*. The state space and action spaces for firms are the same as in Example 2. We endow both X and $A = [0, 1]^n$ with the usual component-wise ordering. Then, X and A are complete lattices. We assume that the transition probability is defined as in (6.4) and $q_1(\cdot|x)$ and $q_2(\cdot|x)$ are for all $x \in X$ dominated by some probability measure μ on S. The payoff function for every firm is given in (6.3).

Example 3 (Cournot oligopoly with complementary goods in differentiated markets). Let h be defined as in (6.5). Suppose that the payoff function in the game $\Gamma_v(x)$ satisfies the following condition:

$$\frac{\partial^2 U^i_\beta(v_i, x, a)}{\partial a_i \partial a_j} \geq 0 \quad \text{for} \quad j \neq i.$$

Then, by Theorem 4 in Milgrom and Roberts (1990), the game $\Gamma_v(x)$ is supermodular. Note that within our framework, it is sufficient to prove that for $u_i(x, a)$, defined in (6.3), it holds $\frac{\partial^2 u_i(s,a)}{\partial a_i \partial a_j} \geq 0$, $j \neq i$. But

$$\frac{\partial^2 u_i(x, a)}{\partial a_i \partial a_j} = a_i \frac{\partial^2 P_i(a)}{\partial a_i \partial a_j} + \frac{\partial P_i(a)}{\partial a_j}, \quad j \neq i$$

and $\frac{\partial^2 u_i(x,a)}{\partial a_i \partial a_j}$ are likely to be nonnegative, if the goods are complements, i.e., $\frac{\partial P_i(a)}{\partial a_j} \geq 0$ for $j \neq i$; see Vives (1990). From Theorem 5 in Milgrom and Roberts (1990), it follows that the game $\Gamma_v(x)$ has a pure Nash equilibrium. Therefore, the arguments used in Jaśkiewicz and Nowak (2016) imply that the stochastic game has a pure stationary almost Markov perfect equilibrium.

Remark 7. The game described in Example 3 is also studied in Curtat (1996), but with additional restrictive assumptions that q_1 and q_2 are independent of $x \in X$. Then, the transition probability q has so-called increasing differences in (x, a). This fact implies that the functions $U^i_\beta(v_i, \cdot, \cdot)$ satisfy the assumptions in

Proposition 7. Other condition imposed by Curtat (1996) states that the payoff functions $u_i(x, a)$ are increasing in a_{-i} and, more importantly, satisfy the so-called strict diagonal dominance condition for each $x \in X$. For details the reader is referred to Curtat (1996) and Rosen (1965). This additional condition entails the uniqueness of a pure Nash equilibrium in every auxiliary game $\Gamma_v(x)$ under consideration; see Proposition 6. The advantage is that Curtat (1996) can directly work with Lipschitz continuous strategies for the players and find a stationary Markov perfect equilibrium in that class using Schauder's fixed point theorem. Without the strict diagonal dominance condition, $\Gamma_v(x)$ may have many pure Nash equilibria and Curtat's approach cannot be applied. The coefficients of the convex combination in (6.4) are affine functions of $a \in A$. This requirement can slightly be generalized; see, for instance, Example 4 in Jaśkiewicz and Nowak (2016). If q_1 or q_2 depends on $x \in X$, then the increasing differences property of q does not hold and the method of Curtat (1996) does not work. Additional comments on supermodular stochastic games can be found in Amir (2003).

The result in Curtat (1996) on the existence of stationary Markov perfect equilibria for supermodular discounted stochastic games is based upon the lattice theoretic arguments and on complementarity and monotonicity assumptions. The state and action spaces are assumed to be compact intervals in Euclidean space, and the transition probability is assumed to be norm continuous in state and action variables. Moreover, the strict diagonal dominance condition (see (C1) in Sect. 2) applied to the auxiliary one-shot games $\Gamma_v(x)$ for any increasing Lipschitz continuous continuation vector payoff v plays a crucial role. Namely, this assumption together with others implies that $\mathcal{NP}_v(x)$ is a singleton. In addition, the function $x \to \mathcal{NP}_v(x)$ is increasing and Lipschitz continuous. Thus, his proof is comprised of two steps. First, he shows that there exists an increasing Lipschitz continuous vector payoff function v^* such that $v^*(x) = \mathcal{NP}_{v^*}(x)$ for all $x \in X$. Second, he makes use of a theorem on the Lipschitz property of the unique equilibrium in Γ_{v^*}.

Horst (2005) provided a different and more unified approach to stationary Markov perfect equilibria that can be applied beyond the setting of supermodular games. Instead of imposing monotonicity conditions on the players' utility functions, he considered stochastic games in which the interaction between different players is sufficiently weak. For instance, certain "production games" satisfy this property. The method of his proof is based on a selection theorem of Montrucchio (1987) and the Schauder fixed point theorem applied to the space of Lipschitz continuous strategy profiles of the players. The assumptions imposed by Horst (2005) are rather complicated. For example, they may enforce a number of players in the game or the upper bound for a discount factor. Such limitations do not occur in the approach of Curtat (1996).

Balbus et al. (2014a) considered supermodular stochastic games with an absorbing state and the transition probabilities of the form $q(\cdot|x, a) = g(x, a)q_0(\cdot|x)+(1-g(x, a))\delta_0(\cdot)$. Under some strong monotonicity conditions on the utility functions and transitions, they showed that the Nash equilibrium payoffs in the n-stage games monotonically converge as $n \to \infty$. This fact yields the existence of pure stationary Markov perfect equilibrium. A related result is given in Balbus et al. (2013b) for a similar class of dynamic games. The state space X in Balbus et al. (2014a) is one-dimensional, and their results do not apply to the games of resource extraction discussed earlier. If, on the other hand, the transition probability is a "mixture" of finitely many probability measures, then a stationary Markov perfect equilibrium can be obtained, in certain models, by solving a system of non-linear equations. This method was discussed in Sect. 5 of Nowak (2007). The next example is not a supermodular game in the sense of Balbus et al. (2014a), but it belongs to the class of production games examined by Horst (2005). Generally, there are only few examples of games with continuum states, for which Nash equilibria can be given in a closed form.

Example 4. Let $X = [0, 1]$, $A_i(x) = [0, 1]$ for all $x \in X$ and $i \in N = \{1, 2\}$. We consider the symmetric game where the stage utility of player i is

$$u_i(x, a_1, a_2) = a_1 + a_2 + 2xa_1a_2 - a_i^2. \tag{6.6}$$

The state variable x in (6.6) is a complementarity coefficient of the players' actions. The transition probabilities are of the form

$$q(\cdot|x, a_1, a_2) := \frac{x + a_1 + a_2}{3}\mu_1(\cdot) + \frac{3 - x - a_1 - a_2}{3}\mu_2(\cdot).$$

We assume that μ_1 has the density $\rho_1(y) = 2y$ and μ_2 has the density $\rho_2(y) = 2 - 2y$, $y \in X$. Note that μ_1 stochastically dominates μ_2. From the definition of q, it follows that higher states $x \in X$ or high actions a_1, a_2 (efforts) of the players induce a distribution of the next state having a higher mean value. Assume that $v^* = (v_1^*, v_2^*)$ is an equilibrium payoff vector in the β-discounted stochastic game. As shown in Example 1 of Nowak (2007), it is possible to construct a system of non-linear equations with unknown z_1 and z_2, whose solution z_1^*, z_2^* is $z_i^* = \int_X v_i^*(y)\mu_i(dy)$. This fact, in turn, gives the possibility finding of a symmetric stationary Markov perfect equilibrium (f_1^*, f_2^*) and $v_1^* = v_2^*$. It is of the form $f_i^*(x) = \frac{1+z^*}{4-2x}$. $x \in X$, $i \in N$, where

$$z^* = \frac{-8 - 6p(\beta - 1) - \sqrt{(8 + 6p(\beta - 1))^2 - 36}}{6} \text{ and } p = \frac{9 + 2\beta \ln 2 - 2\beta}{\beta(1 - \beta)(6 \ln 2 - 3)}.$$

Moreover, we have

$$v_i^*(x) = (p\beta + x)z^* + \frac{(1 + z^*)^2(3 - x)}{2(2 - x)^2}.$$

Ericson and Pakes (1995) provided a model of firm and industry dynamics that allows for entry, exit and uncertainty generating variability in the fortunes of firms. They considered the ergodicity of the stochastic process resulting from a Markov perfect industry equilibrium. A dynamic competition in an oligopolistic industry with investment, entry and exit was also extensively studied by Doraszelski and Satterthwaite (2010). Computational methods for the class of games studied by Ericson and Pakes (1995) are presented in Doraszelski and Pakes (2007). Further applications of discounted stochastic games with countably many states to models in industrial organization including models of industry dynamics are given in Escobar (2013).

Shubik and Whitt (1973) considered a non-stochastic model of sequential strategic market game, where the state includes a current stocks of capital. At each period of the game, one unit of a consumer good is put up for sale, and players bid some amounts of fiat money for it. A stochastic counterpart of this game was first presented in Secchi and Sudderth (2005). Więcek (2009), on the other hand, obtained a general structure of equilibrium policies in two-person games, where bids gradually decrease with increase of the discount factor. Moreover, Więcek (2012) proved that a Nash equilibrium, where all the players use "aggressive strategies", emerges in the game for any value of the discount factor as the number of players is sufficiently large. This fact corresponds to a similar result for a deterministic economy given in Shubik and Whitt (1973) as well as being consistent with existing results about economies with continuum of players. Other applications of nonzero-sum stochastic games to economic models can be found in Duggan (2012) and He and Sun (2017). Although the concept of mean field equilibrium in dynamic games is not directly inspired by Nash, the influence of the theory of non-cooperative stochastic games on this area of research is obvious. Also the notion of supermodularity is used in studying the mean field equilibria in dynamic games. The reader is referred to Adlakha and Johari (2013) where some applications to computer science and operations research are given.

7 Special Classes of Stochastic Games with Countably Many States

Assume that the state space X is countable. Then every F_i^0 can be recognized as a compact convex subset of a linear topological space. A sequence $(f_i^k)_{k\in\mathbb{N}}$ converges to $f_i \in F_i^0$ if for every $x \in X$ $f_i^k(\cdot|x) \to f_i(\cdot|x)$ in the weak-star topology on the space of probability measures on $A_i(x)$. The weak or weak-star convergence of probability measures on metric spaces is fully described in Aliprantis and Border (2006) or Billingsley (1968). Since X is countable, every space F_i^0 is sequentially compact (it suffices to use the standard diagonal method for selecting convergent subsequences) and, therefore, F^0 is sequentially compact when endowed with the product topology. If X is finite and the sets of actions are finite, then F^0 can actually be viewed as a convex compact subset of Euclidean space. In the finite state space case, it is easy to prove that the discounted payoffs $J^i_\beta(x, f)$ are continuous on F^0. If X is countable and the payoff functions are uniformly bounded, and $q(y|x, a)$ is continuous in $a \in A(x)$ for all $x, y \in X$, then showing the continuity of $J^i_\beta(x, f)$ on F^0 requires a little more work; see Federgruen (1978). From the Bellman equation in discounted dynamic programming (see Puterman 1994), it follows that $f^* = (f^*_1, \ldots, f^*_n)$ is a stationary Markov perfect equilibrium in the discounted stochastic game if and only if there exist bounded functions $v^*_i : X \to \mathbb{R}$ such that for each $x \in X$ and $i \in N$ we have

$$v^*_i(x) = \max_{\nu_i \in \Pr(A_i(x))} U^i_\beta(v^*_i, x, (\nu_i, f^*_{-i})) = U^i_\beta(v^*_i, x, f^*). \tag{6.7}$$

From (6.7), it follows that $v^*_i(x) = J^i_\beta(x, f^*)$. Using the continuity of the expected discounted payoffs in $f \in F^0$ and (6.7), one can define the best response correspondence in the space of strategies, show its upper semicontinuity and conclude from the fixed point theorem due to Glicksberg (1952) (or due to Kakutani (1941) in the case of finite state and action space) that the game has a stationary Markov perfect equilibrium $f^* \in F^0$. This fact was proved for finite state space discounted stochastic games by Fink (1964) and Takahashi (1964). An extension to games with countable state spaces was reported in Parthasarathy (1973) and Federgruen (1978). Some results for a class of discounted games with discontinuous payoff functions can be found in Nowak and Więcek (2007).

The fundamental results in the theory of regular Nash equilibria in normal form games concerning genericity (see Harsanyi 1973a) and purification (see Harsanyi 1973b) were extended to dynamic games by Doraszelski and Escobar (2010). A discounted stochastic game possessing equilibria that are all regular in the sense of Doraszelski and Escobar (2010) has a compact equilibrium set that consists of isolated points. Hence, it follows that the equilibrium set is finite. They proved that the set of discounted stochastic games (with finite sets of states and actions) having Markov perfect equilibria that all are regular is open and has full Lebesgue measure. Related results were given by Haller and Lagunoff (2000), but their definition of regular equilibrium is different and may not be purifiable.

The payoff function for player $i \in N$ in the limit-average stochastic game can be defined as follows:

$$\bar{J}^i(x, \pi) := \liminf_{T \to \infty} E^\pi_x \left(\frac{1}{T} \sum_{t=1}^{T} u_i(x_t, a^t) \right), \quad x \in X, \ \pi \in \Pi.$$

The equilibrium solutions for this class of games are defined similarly as in the discounted case. The existence of stationary Markov perfect equilibria for games with finite state and action spaces and the limit-average payoffs was proved independently by Rogers (1969) and Sobel (1971). They assumed that the Markov chain induced by any stationary strategy profile and the transition probability q is *irreducible* . It is shown under the irreducibility condition that the equilibrium payoffs w^*_i of the players are independent of the initial state. Moreover, it is proved that there exists a sequence of equilibria $(f^k)_{k\in\mathbb{N}}$ in β_k-discounted games (with $\beta_k \to 1$ as $k \to \infty$) such that $w_i^* = \lim_{k\to\infty}(1-\beta_k)J^i_{\beta_k}(x, f^k)$. Later, Federgruen (1978) extended these results to limit-average stochastic games with countably many states satisfying some uniform ergodicity conditions. Other cases of similar type were mentioned by Nowak (2003a). Below we provide a result due to Altman et al. (1997), which has some potential for applications in queueing models. The stage payoffs in their approach may be unbounded. We start with a formulation of their assumptions.

Let $m : X \to [1, \infty)$ be a function for which the following conditions hold.

(A4) For each $x, y \in X$, $i \in N$, the functions $u_i(x, \cdot)$ and $q(y|x, \cdot)$ are continuous on $A(x)$. Moreover,

$$\sup_{x\in X}\max_{a\in A(x)} |u_i(x,a)|/m(x) < \infty \quad \text{and}$$

$$\lim_{k\to\infty}\sum_{y\in X} |q(y|x,a^k) - q(y|x,a)|m(y) = 0$$

for any $a^k \to a \in A(x)$.

(A5) There exist a finite set $Y \subset X$ and $\gamma \in (0, 1)$ such that

$$\sum_{y\in X\setminus Y} q(y|x,a)m(y) \le \gamma m(x) \quad \text{for all} \quad x \in X,\ a \in A(x).$$

(A6) The function $f \to n(f)$ is continuous with respect to stationary strategy profiles $f \in F^0$, where $n(f)$ denotes the number of closed classes in the Markov chain induced by the transition probability $q(y|x, f)$, $x, y \in X$.

Property (A5) is called m-uniform geometric recurrence; see Altman et al. (1997). Condition (A6) is quite restrictive and implies that the number of positive recurrent classes is a constant function of the stationary strategies. If the Markov chains resulting from the stationary policies are all unichain, the limit-average payoff functions are constant, i.e., independent of the initial state. For a detailed discussion, we refer the reader to Altman et al. (1997) and the references cited therein.

Theorem 12. *If conditions* (A4)–(A6) *are satisfied, then the limit-average payoff n-person stochastic game has a stationary Markov perfect equilibrium.*

The above result follows from Theorem 2.6 in Altman et al. (1997), where it is also shown that under (A4)–(A6) any limit of stationary Markov equilibria in β-discounted games (as $\beta \to 1$) is an equilibrium in the limit-average game. A related result was established by Borkar and Ghosh (1993) under a stochastic stability condition. More precisely, they assumed that the Markov chain induced by any stationary strategy profile is unichain and the transition probability from any fixed state has a finite support.

Stochastic games with countably many states are usually studied under some recurrence or ergodicity conditions. Without these conditions n-person nonzero-sum limit-average payoff stochastic games with countable state spaces are very difficult to deal with. Nevertheless, the results obtained in the literature have some interesting applications, especially to queueing systems; see, for example, Altman (1996) and Altman et al. (1997).

Now assume that X is a Borel space and μ is a probability measure on X. Consider an n-person discounted stochastic game G, where $A_i(x) = A_i$ for all $i \in N$ and $x \in X$, the payoff functions are uniformly bounded and continuous in actions.

(A7) The transition probability q has a conditional density function ρ, which is continuous in actions and such that

$$\int_X \max_{a \in A} \rho(x, a, y)\mu(dy) < \infty.$$

Let $C(A)$ be the Banach space of all real-valued continuous functions on the compact space A endowed with the supremum norm $\| \cdot \|_\infty$. By $L_1(X, C(A))$ we denote the Banach space of all $C(A)$-valued measurable functions ϕ on X such that $\|\phi\|_1 := \int_X \|\phi(y)\|_\infty \mu(dy) < \infty$. Let $\{X_k\}_{k \in \mathbb{N}_0}$ be a measurable partition of the state space ($\mathbb{N}_0 \subset \mathbb{N}$), $\{u_{i,k}\}_{k \in \mathbb{N}_0}$ be a family of functions $u_{i,k} \in C(A)$ and $\{\rho_k\}_{k \in \mathbb{N}_0}$ be a family of functions $\rho_k \in L_1(X, C(A))$ such that $\rho_k(x)(a, y) \geq 0$ and $\int_X \rho_k(x)(a, y)\mu(dy) = 1$ for each $k \in \mathbb{N}_0$, $a \in A$. Consider a game $\tilde{G}$ where the payoff function for player i is $\tilde{u}_i(x, a) = u_k(a)$ if $x \in X_k$. The transition density is $\tilde{\rho}(x, a, y) = \rho_k(x)(a, y)$ if $x \in X_k$. Let $\tilde{F}_i^0$ be the set of all $f_i \in F_i^0$ that are constant on every set X_k. Let $\tilde{F}^0 := \tilde{F}_1^0 \times \cdots \times \tilde{F}_n^0$. The game $\tilde{G}$ resembles a game with countably many states and if the payoff functions $\tilde{u}_i$ are uniformly bounded, then $\tilde{G}$ with the discounted payoff criterion has an equilibrium in $\tilde{F}^0$. Denote by $\tilde{J}_\beta^i(x, \pi)$ the discounted expected payoff to player $i \in N$ in the game $\tilde{G}$. It is well known that $C(A)$ is separable. The Banach space $L_1(X, C(A))$ is also separable. Note that $x \to u_i(x, \cdot)$ is a measurable mapping from X to $C(A)$. By, (A7) the mapping $x \to \rho(x, \cdot, \cdot)$ from X to $L_1(X, C(A))$ is also measurable. Using these facts Nowak (1985) proved the following result.

Theorem 13. *Assume that G satisfies* (A7). *For any $\epsilon > 0$, there exists a game $\tilde{G}$ such that $|J_\beta^i(x, \pi) - \tilde{J}_\beta^i(x, \pi)| < \epsilon/2$ for all $x \in X$, $i \in N$ and $\pi \in \Pi$. Moreover, the game G has a stationary Markov ϵ-equilibrium.*

A related result on approximation of discounted nonzero-sum games and existence of ϵ-equilibria was given by Whitt (1980), who used stronger uniform continuity conditions and used a different technique. Approximations of discounted and also limit-average stochastic games with general state spaces and unbounded stage functions were studied in Nowak and Altman (2002). They used the weighted norm approach and imposed some geometric ergodicity conditions while examining the limit-average case. An extension with simpler and more transparent proof for semi-Markov games satisfying a geometric drift condition and a majorization property, similar to (GE1)–(GE3) in Sect. 5 in Jaśkiewicz and Nowak (2018a), was given in Jaśkiewicz and Nowak (2006).

8 Algorithms for Nonzero-Sum Stochastic Games

In this section, we assume that the state space X and the sets of actions A_i are finite. In the 2-player case, we let for notational convenience $A_1(x) = A_1$, $A_2(x) = A_2$ and $a = a_1 \in A_1$, $b = a_2 \in A_2$. Further, for any $f_i \in F_i^0$, $i = 1, 2$, we set

$$q(y|x, f_1, f_2) := \sum_{a\in A_1} \sum_{b\in A_2} q(y|x, a, b) f_1(a|x) f_2(b|x),$$

$$q(y|x, f_1, b) := \sum_{a\in A_1} q(y|x, a, b) f_1(a|x),$$

$$u_i(x, f_1, f_2) := \sum_{a\in A_1} \sum_{b\in A_2} u_i(x, a, b) f_1(a|x) f_2(b|x),$$

$$u_i(x, f_1, b) := \sum_{a\in A_1} u_i(x, a, b) f_1(a|x).$$

In a similar way, we define $q(y|x, a, f_2)$ and $u_i(x, a, f_2)$. Note that every $f_i \in F_i^0$ can be recognized as a compact convex subset of Euclidean space. Also every function $\phi : X \to \mathbb{R}$ can be viewed as a vector in Euclidean space. Below we describe two results of Filar et al. (1991) about characterization of stationary equilibria in stochastic games by *constrained nonlinear programming*. However, due to the fact that the constraint sets are not convex, the results are not straightforward in numerical implementation. Although it is common in mathematical programming to use matrix notation, we follow the one introduced in previous sections.

Let $c = (v_1, v_2, f_1, f_2)$. Consider the following problem (OP_β):

$$\min\ O_1(c) := \sum_{i=1}^{2} \sum_{x\in X} \left(v_i(x) - u_i(x, f_1, f_2) - \beta \sum_{y\in X} v_i(y) q(y|x, f_1, f_2) \right)$$

subject to $(f_1, f_2) \in F_1^0 \times F_2^0$ and

$$u_1(x, a, f_2) + \beta \sum_{x\in X} v_1(y) q(y|x, a, f_2) \le v_1(x),\ \text{for all } x \in X,\ a \in A_1,$$

and

$$u_2(x, f_1, b) + \beta \sum_{x \in X} v_2(y) q(y|x, f_1, b) \le v_2(x), \text{ for all } x \in X,\ b \in A_2.$$

Theorem 14. *Consider a feasible point* $c^* = (v_1^*, v_2^*, f_1^*, f_2^*)$ *in* (OP_β)*. Then* $(f_1^*, f_2^*) \in F_1^0 \times F_2^0$ *is a stationary Nash equilibrium in the discounted stochastic game if and only if* c^* *is a solution to problem* (OP_β) *with* $O_1(c^*) = 0$.

Let $c = (z_1, v_1, w_1, f_2, z_2, v_2, w_2, f_1)$. Now consider the following problem (OP_a):

$$\min\ O_2(c) := \sum_{i=1}^{2} \sum_{x \in X} \left(v_i(x) - \sum_{y \in X} v_i(y) q(y|x, f_1, f_2) \right)$$

subject to $(f_1, f_2) \in F_1^0 \times F_2^0$ and

$$\sum_{y \in X} v_1(y) q(y|x, a, f_2) \le v_1(x),$$

$$u_1(x, a, f_2) + \sum_{y \in X} z_1(y) q(y|x, a, f_2) \le v_1(x) + z_1(x)$$

for all $x \in X,\ a \in A_1$ and

$$\sum_{y \in X} v_2(y) q(y|x, f_1, b) \le v_2(x),$$

$$u_2(x, f_1, b) + \sum_{y \in X} z_2(y) q(y|x, f_1, b) \le v_2(x) + z_2(x)$$

for all $x \in X,\ b \in A_2$ and

$$u_i(x, f_1, f_2) + \sum_{y \in X} w_i(y) q(y|x, f_1, f_2) = v_i(x) + w_i(x)$$

for all $x \in X$ and $i = 1, 2$.

Theorem 15. *Consider a feasible point* $c^* = (z_1^*, v_1^*, w_1^*, f_2^*, z_2^*, v_2^*, w_2^*, f_1^*)$ *in* (OP_a)*. Then* $(f_1^*, f_2^*) \in F_1^0 \times F_2^0$ *is a stationary Nash equilibrium in the limit-average payoff stochastic game if and only if* c^* *is a solution to problem* (OP_a) *with* $O_2(c^*) = 0$.

Theorems 14 and 15 were stated and proved in Filar et al. (1991); see also Theorems 3.8.2 and 3.8.4 in Filar and Vrieze (1997).

As in the zero-sum case, when one player controls the transitions, it is possible to construct finite-step algorithms to compute Nash equilibria. The *linear complementarity problem* (LCP) is defined as follows. Given a square matrix $\mathbb{M}$ of order m and a (column) vector $\overline{Q} \in \mathbb{R}^m$, we find two vectors $\overline{Z} = [z_1, \ldots, z_m]^T \in \mathbb{R}^m$ and $\overline{W} = [w_1, \ldots, w_m]^T \in \mathbb{R}^m$ such that

$$\mathbb{M}\overline{Z} + \overline{Q} = \overline{W} \quad \text{and} \quad w_j \geq 0,\ z_j \geq 0,\ z_j w_j = 0 \quad \text{for all} \quad j = 1, \ldots, m.$$

Lemke (1965) proposed some pivoting finite-step algorithms to solve the LCP for a large class of matrices $\mathbb{M}$ and vectors $\overline{Q}$. Further research on the LCP can be found in Cottle et al. (1992).

Finding a Nash equilibrium in any bimatrix game $(\mathbb{A}, \mathbb{B})$ is equivalent to solving the LCP with

$$\mathbb{M} = \begin{bmatrix} \mathbb{B}^T & \mathbb{O} \\ \mathbb{O} & \mathbb{A} \end{bmatrix} \text{ where } \mathbb{O} \text{ is the matrix with zero entries, } \overline{Q} = [-1, \ldots, -1]^T.$$

A finite-step algorithm for this LCP was given by Lemke and Howson (1964). If $\overline{Z}^* = [\overline{Z}_1^*, \overline{Z}_2^*]$ is a part of the solution of the above LCP, then the normalization of $\overline{Z}_i^*$ is an equilibrium strategy for player i.

Suppose that only player 2 controls the transitions in a *discounted stochastic game*, i.e., $q(y|x, a, b)$ is independent of $a \in A$. Let $\{f_1, \ldots, f_{m_1}\}$ and $\{g_1, \ldots, g_{m_2}\}$ be the families of all *pure stationary strategies* for players 1 and 2, respectively. Consider the bimatrix game $(\mathbb{A}, \mathbb{B})$, where the entries a_{ij} of $\mathbb{A}$ and b_{ij} of $\mathbb{B}$ are

$$a_{ij} := \sum_{x \in X} u_1(x, f_i(x), g_j(x)) \quad \text{and} \quad b_{ij} := \sum_{x \in X} J_\beta^2(x, f_i, g_j).$$

Then, making use of the Lemke-Howson algorithm, Nowak and Raghavan (1993) proved the following result.

Theorem 16. *Let* $\xi^* = (\xi_1^*, \ldots, \xi_{m_1}^*)$ *and* $\zeta^* = (\zeta_1^*, \ldots, \zeta_{m_2}^*)$ *and assume that* (ξ^*, ζ^*) *is a Nash equilibrium in the bimatrix game* $(\mathbb{A}, \mathbb{B})$ *defined above. Then the stationary strategies*

$$f^*(x) = \sum_{j=1}^{m_1} \xi_j^* \delta_{f_j(x)} \quad \textit{and} \quad g^*(x) = \sum_{j=1}^{m_2} \zeta_j^* \delta_{g_j(x)}$$

form a Nash equilibrium in the discounted stochastic game.

It should be noted that a similar result does not hold for stochastic games with the limit-average payoffs. Note that the entries of the matrix $\mathbb{B}$ can be computed in finitely many steps, but the order of the associated LCP is very high. Therefore, a natural question arises as to whether the single-controller stochastic game can be solved with the help of LCP formulation with appropriately defined matrix $\mathbb{M}$ (with lower dimension) and vector $\overline{Q}$. Since the payoffs and transitions depend on states and stationary equilibria which are characterized by the systems of Bellman equations, the dimension of the LCP must be high. However, it should be essentially smaller than in the case of Theorem 16. Such an LCP formulation for discounted single-controller stochastic games was given by Mohan et al. (1997) and further developed in Mohan et al. (2001). In the case of the limit-average payoff and single-controller stochastic game, Raghavan and Syed (2002) provided an analogous algorithm. Further studies on specific classes of stochastic games (acyclic 3-person switching control games, polystochastic games) can be found in Krishnamurthy et al. (2012).

Let us recall that a Nash equilibrium in an n-person game is a fixed point of some mapping. A fixed point theorem of certain deformations of continuous mappings proved by Browder (1960) turned out to be basic for developing so-called homotopy methods in computing equilibria in nonzero-sum games. It reads as follows.

Theorem 17. *Assume that $C \subset \mathbb{R}^d$ is a nonempty compact convex set. Let $\Psi : [0,1] \times C \to C$ be a continuous mapping and $F(\Psi) := \{(t,c) \in [0,1] \times C : c = \Psi(t,c)\}$. Then $F(\Psi)$ contains a connected subset $F_c(\Psi)$ such that $F_c(\Psi) \cap (\{0\} \times C) \neq \emptyset$ and $F_c(\Psi) \cap (\{1\} \times C) \neq \emptyset$.*

This result was extended to upper semicontinuous correspondences by Mas-Colell (1974). Consider an n-person game and assume that Ψ_1 is a continuous mapping whose fixed points in the set C of strategy profiles correspond to Nash equilibria in this game. The basic idea in the homotopy methods is to define a "deformation" Ψ of Ψ_1 such that $\Psi(1,c) = \Psi_1(c)$ for all $c \in C$ and such that $\Psi(0,c)$ has a unique fixed point, say c_0^*, that is relatively simple to find. By Theorem 17, $F_c(\Psi)$ is a connected set. Thus, c_0^* is connected via $F_c(\Psi)$ with a fixed point c_1^* of Ψ_1. Hence, the idea is to consider the connected set $F_c(\Psi)$. Since the dimension of the domain of Ψ is one higher than the dimension of its range, one can formulate regularity conditions under which the approximation path is a compact, piecewise differentiable one-dimensional manifold, i.e., it is a finite collection of arcs and loops. In the case of bimatrix games, a nondegeneracy condition is sufficient to guarantee that the aforementioned properties are satisfied. A comprehensive discussion of the homotopy algorithms applied to n-person games is provided by Herings and Peeters (2010) and references cited therein. According to the authors, "advantages of homotopy algorithms include their numerical stability, their ability to locate multiple solutions, and the insight they provide in the properties of solutions". Various examples show that implementation of homotopy methods is rather straightforward with the aid of available professional software. It is worth recalling the known fact that the Lemke-Howson algorithm can be applied to bimatrix games only. An issue of finding Nash equilibria in concave n-person games comprises a non-linear complementarity problem. Therefore, one can only expect to obtain approximate equilibria by different numerical methods.

The homotopy methods, as noted by Herings and Peeters (2004), are also useful in the study of stationary equilibria, their structure and computation in nonzero-sum stochastic games. Their results can be applied to n-person discounted stochastic games with finite state and action spaces.

Recently, Govindan and Wilson (2003) proposed a new algorithm to compute Nash equilibria in finite games. Their algorithm combines the global Newton method (see Smale 1976)) and a homotopy method for finding fixed points of continuous mappings developed by Eaves (1972, 1984). In the construction of a Nash equilibrium, a fundamental topological property of the graph of the Nash equilibrium correspondence discovered by Kohlberg and Mertens (1986) plays an important role. Being more precise, the authors show that making use of the global Newton method, it is possible to trace the path of the homotopy by a dynamical system. The same method can be applied to a construction of an algorithm for n-person discounted stochastic games with finite action and state sets; see Govindan and Wilson (2009). Strategic n-person games with a potential function having pure Nash equilibria were considered by Monderer and Shapley (1996). Potters et al. (2009), on the other hand, examined certain classes of discounted stochastic games via the potential function approach and constructed pure stationary Nash equilibria by solving a finite number of finite strategic games.

Solan and Vieille (2010) pointed out that the methods based on formal logic, successfully applied to zero-sum games, are also useful in the examination of certain classes of nonzero-sum stochastic games with the limit-average payoff criterion.

9 Uniform Equilibrium, Subgame Perfection, and Correlation in Stochastic Games with Finite State and Action Spaces

In this section, we consider stochastic games with *finite state space* $X = \{1, \dots, s\}$ and *finite sets of actions*. We deal with "normalized discounted payoffs" and use notation which is more consistent with the surveyed literature. We let $\beta = 1 - \lambda$ and multiply all current payoffs by $\lambda \in (0, 1)$. Thus, we consider

$$J_\lambda^i(x, \pi) := E_x^\pi \left(\sum_{t=1}^{\infty} \lambda (1-\lambda)^{t-1} u_i(x_t, a^t) \right), \quad x = x_1 \in X,\ \pi \in \Pi,\ i \in N.$$

For any $T \in \mathbb{N}$ and $x = x_1 \in X$, $\pi \in \Pi$, the T-stage average payoff for player $i \in N$ is

$$J_T^i(x, \pi) := E_x^\pi \left(\frac{1}{T} \sum_{t=1}^{T} u_i(x_t, a^t) \right).$$

A vector $\bar{g} \in \mathbb{R}^n$ is called a *uniform equilibrium payoff* at the initial state $x \in X$ if for every $\epsilon > 0$ there exist $\lambda^0 \in (0, 1]$, $T^0 \in \mathbb{N}$ and a strategy profile $\pi^0 \in \Pi$

such that for every player $i \in N$ and every strategy $\pi_i \in \Pi_i$, we have

$$\bar{g}^i + \epsilon \geq J^i_\lambda(x, \pi^0) \geq \bar{g}^i - \epsilon \geq J^i_\lambda(x, (\pi_i, \pi^0_{-i})) - 2\epsilon \quad \text{for} \quad \lambda \in (0, \lambda^0]$$

and

$$\bar{g}^i + \epsilon \geq J^i_T(x, \pi^0) \geq \bar{g}^i - \epsilon \geq J^i_T(x, (\pi_i, \pi^0_{-i})) - 2\epsilon \quad \text{for} \quad T \geq T^0.$$

Any profile π^0 that has the above two properties is a called a *uniform ϵ-equilibrium*. In other words, the game has a uniform equilibrium payoff if for every $\epsilon > 0$ there is a strategy profile π^0 which is an ϵ-equilibrium in every discounted game with a sufficiently small discount factor λ and in every finite-stage game with sufficiently long time horizon.

A stochastic game is called *absorbing* if all states but one are absorbing. Assume that $X = \{1, 2, 3\}$ and only state $x = 1$ is nonabsorbing. Let E^0 denote the set of all uniform equilibrium payoffs. Since the payoffs are determined in states 2 and 3, in a *two*-person game, the set E^0 can be viewed as a subset of $\mathbb{R}^2$. Let $\lambda_k \to 0$ as $k \to \infty$, and let f^*_k be a stationary Markov perfect equilibrium in the λ_k-discounted *two*-person game. A question arises as to whether the sequence $(J^1_{\lambda_k}(x, f^*_k), J^2_{\lambda_k}(x, f^*_k))_{k \in \mathbb{N}}$ with $x = 1$ has an accumulation point $\bar{g} \in E^0$. That is the case in the zero-sum case (see Mertens and Neyman 1981). Sorin (1986) provided a nonzero-sum modification of the "Big Match," where only state $x = 1$ is nonabsorbing in which $\lim_{k \to \infty}(J^1_{\lambda_k}(1, f^*_k), J^2_{\lambda_k}(1, f^*_k)) \notin E^0$. A similar phenomenon occurs for the limit of T-stage equilibrium payoffs. Sorin (1986) gave a full description of the set E^0 in his example. His observations were generalized by Vrieze and Thuijsman (1989) to all 2-person absorbing games. They proved the following result.

Theorem 18. *Any two-person absorbing stochastic game has a uniform equilibrium payoff.*

We now state the fundamental result of Vieille (2000a,b).

Theorem 19. *Every two-person stochastic game has a uniform equilibrium payoff.*

The proof of Vrieze and Thuijsman (1989) is based on the "vanishing discount factor approach" combined with the idea of "punishment" successfully used in repeated games. The assumption that there are only two players is important in the proof. The ϵ-equilibrium strategies that they construct need unbounded memory. The proof of Vieille (2000a,b) is involved. One of the reasons is that the ergodic classes do not depend continuously on strategy profiles. Following Vieille (2002) one can briefly say that "the basic idea is to devise an ϵ-equilibrium profile that takes the form of a stationary-like strategy vector, supplemented by threats of indefinite punishment". The construction of uniform equilibrium payoff consists of two independent steps. First, a class of solvable states is recognized and some controlled sets are considered. Second, the problem is reduced to the existence of equilibria in a class of recursive games. The punishment component is crucial in the construction and therefore the fact that the game is 2-person is important. Neither of the two parts of the proof can be extended to games with more than two players. The ϵ-equilibrium profiles have no subgame-perfection property and require unbounded memory for the players. For a heuristic description of the proof, the reader is referred to Vieille (2002). In a recent paper Solan (2017) proposed a new solution concept for multiplayer stochastic games called acceptable strategy profiles. It is relatively simpler than uniform equilibrium and has some interesting properties. A suitable adaptation of the notion of uniform equilibrium is studied by Neyman (2017) in the class of continuous-time stochastic games with a small imprecision in the specification of players' evaluations of streams of payoffs.

Flesch et al. (1997) proposed a *three*-person game with absorbing states where only a cyclic Markov equilibrium exists. No examples of this type were found in the 2-person case. This example inspired Solan (1999), who making use of certain arguments from Vrieze and Thuijsman (1989), proved the following result.

Theorem 20. *Every* 3*-person absorbing stochastic game has a uniform equilibrium payoff.*

In a *quitting game*, every player has only two actions, c for continue and q for quit. As soon as one or more of the players at any stage chooses q, the game stops and the players receive their payoffs, which are determined by the subset of players, say S, that choose simultaneously the action q. If nobody chooses q throughout all stages of play, then all players receive zero. The payoffs are defined as follows. For every nonempty subset $S \subset N$ of players, there is a payoff vector $v(S) \in \mathbb{R}^n$. The first stage on which S is the subset of players that choose q at this stage, every player $i \in N$ receives the payoff $v(S)_i$. A quitting game is a special limit-average-absorbing stochastic game. The example of Flesch et al. (1997) belongs to this class. We now state the result due to Solan and Vieille (2001).

Theorem 21. *Consider a quitting game satisfying the following assumptions: if player i alone quits, then i receives* 1*, and if player i quits with some other players, then i receives at most* 1*. Then the game has a subgame-perfect ϵ-equilibrium. Moreover, there is a cyclic ϵ-equilibrium strategy profile.*

Quitting games are special cases of "escape games" studied by Simon (2007). As shown by Simon (2012), a study of quitting games can be based on some methods of topological dynamics and homotopy theory. More comments on this issue can be found in Simon (2016).

Thuijsman and Raghavan (1997) studied n-person perfect information stochastic games and n-person ARAT stochastic games and showed the existence of pure equilibria in the limit-average payoff case. They also derived the existence of

ϵ-equilibria for 2-person switching control stochastic games with the same payoff criterion. A class of n-person stochastic games with the limit-average payoff criterion and additive transitions as in the ARAT case (see Sect. 5) was studied by Flesch et al. (2007). The payoff functions do not satisfy any separability in actions assumption. They established the existence of Nash equilibria that are history dependent. For 2-person absorbing games, they showed the existence of stationary ϵ-equilibria. In Flesch et al. (2008, 2009), the authors studied stochastic games with the limit-average payoffs where the state space X is the Cartesian product of some finite sets X_i, $i \in N$. For any state $x = (x_1, \ldots, x_n) \in X$ and any profile of actions $a = (a_1, \ldots, a_n)$, the transition probability is of the form $q(y|x, a) = q_1(y_1|x_1, a_1) \cdots q_n(y_n|x_n, a_n)$ where $y = (y_1, \ldots, y_n) \in X$. In both aperiodic and periodic cases, they established the existence of Nash equilibria for n-person games. In the two-person zero-sum case, there exists a stationary Markov perfect equilibrium.

A stochastic game is *recursive* if the payoffs at all nonabsorbing states are zero. The class of recursive stochastic games is important. The payoffs in any absorbing state can be interpreted as limit averages of stage payoffs as soon as the absorbing state is reached. If no absorbing state is reached, then the average payoff is zero. Moreover, as noted by Simon (2016), "by expanding the state space of any normal stochastic game so that there is a one-to-one relationship between the finite histories of play and the states, any state corresponds to a clopen (open and closed) subset of the infinite histories of play and every open subset of the infinite histories of play will correspond to some collection of states. A stochastic game where all non-zero payoffs are determined by membership in an open set of the infinite histories of play becomes in this way equivalent to a recursive game. Notice that if all absorbing payoffs are positive then the payoffs are lower semicontinuous, and if all absorbing payoffs are negative then the payoffs are upper semicontinuous (as functions on the infinite histories of play)." Flesch et al. (2010b) considered a class of n-person stochastic perfect information games assuming that in every state, the transitions are controlled by one player. The payoffs are equal to zero in every nonabsorbing state and are nonnegative in every absorbing state. They proposed a new iterative method to analyse these games under the expected limit-average payoff criterion and proved the existence of a subgame-perfect ϵ-equilibrium in pure strategies. They also showed the existence of the uniform equilibrium payoffs. Recursive n-person perfect information games, where each player controls one nonabsorbing state and the transitions are deterministic, were studied in Kuipers et al. (2016). Allowing also for negative payoffs in absorbing states (in contrast to Flesch et al. 2010b), the authors showed the existence of a subgame-perfect ϵ-equilibrium by a combinatorial method.

Correlated equilibria were introduced by Aumann (1974, 1987) for games in normal form. Correlation devices may be of different types; see Forges (2009). In Sect. 4 we considered a correlation device using public randomization. They are also called stationary, because at every stage a signal is generated according to the same probability distribution, independent of any data. There are also devices based on past signals that were sent to the players, but not on the past play. They are called "autonomous correlation devices" (see Forges 2009). An ϵ-equilibrium in an extended game that includes an autonomous correlation device is also called an extensive-form correlated ϵ-equilibrium in a multistage game. Solan (2001) characterized the set of extensive-form correlated ϵ-equilibria in stochastic games. He showed that every feasible and individually rational payoff in a stochastic game is an extensive-form correlated equilibrium payoff constructed with the help of an appropriately chosen device.

The following two results are due to Solan and Vieille (2002).

Theorem 22. *Every n-person stochastic game with finite state and action spaces has a uniform correlated equilibrium payoff using an autonomous correlation device.*

The construction of an equilibrium profile is based on the method of Mertens and Neyman (1981) applied to zero-sum games. The equilibrium path is sustained by the use of threat strategies. However, punishment occurs only if a player disobeys the recommendation of the correlation device. The second result is stronger in some sense but concerns positive recursive games, where the payoffs in absorbing states are nonnegative for all players.

Theorem 23. *Every positive recursive stochastic game with finite sets of states and actions has a uniform correlated equilibrium payoff and the correlation device can be taken to be stationary.*

The proof of the above result makes use of a variant of the method of Vieille (2000b).

In a recent paper, Mashiah-Yaakovi (2015) considered stochastic games with countable state spaces, finite sets of actions and Borel measurable bounded payoffs, defined on the space H_∞ of all plays. This class includes the G_δ-games of Blackwell (1969). The concept of uniform ϵ-equilibrium does not apply to this class of games, because the payoffs are not additive. She proved that these games have extensive-form correlated ϵ-equilibria.

Secchi and Sudderth (2002a) considered a special class of n-person stochastic "stay-in-a-set games" defined as follows. Let G_i be a fixed subset of X for each $i \in N$. Define $G^i_\infty := \{(x_1, a^1, x_2, a^2, \ldots)\}$, where $x_t \in G_i$ for every t. The payoff function for player $i \in N$ is the characteristic function of the set G_i^∞. They proved the existence of an ϵ-equilibrium (equilibrium) assuming that the state space is countable (finite) and the sets of actions are finite. Maitra and Sudderth (2003) generalized this result to the Borel state stay-in-a-set games with compact action sets using standard continuity assumption on the transition probability with respect to actions. Secchi and Sudderth (2002b) proved that every n-person stochastic game with countably many states, finite action sets and bounded upper semicontinuous payoff functions on H_∞ has an ϵ-equilibrium. All proofs in the aforementioned papers are partially based on the methods from gambling theory; see Dubins and Savage (2014).

Nonzero-sum infinite horizon games with perfect information are special cases of stochastic games. Flesch et al. (2010a) established the existence of subgame-perfect ϵ-equilibria in pure strategies in perfect information games with lower semicontinuous payoff functions on the space H_∞ of all plays. A similar result for games with upper semicontinuous payoffs was proved by Purves and Sudderth (2011). It is worth mentioning that the aforementioned results also hold for games with arbitrary nonempty action spaces and deterministic transitions. Solan and Vieille (2003) provided an example of a *two*-person game with perfect information that has no subgame-perfect ϵ-equilibrium in pure strategies, but does have a subgame-perfect ϵ-equilibrium in behavior strategies. Their game belongs to the class of deterministic stopping games. Recently, Flesch et al. (2014) showed that a subgame-perfect ϵ-equilibrium (in behavioral strategies) may not exist in perfect information games if the payoff functions are bounded and Borel measurable.

Additional general results on subgame-perfect equilibria in games of perfect information can be found in Alós-Ferrer and Ritzberger (2015, 2016). Two refinements of subgame-perfect ϵ-equilibrium concept were introduced and studied in continuous games of perfect information by Flesch and Predtetchinski (2015).

Two-person discounted stochastic games of perfect information with finite state and action spaces were treated in Küenle (1994). Making use of threat strategies, he constructed a history-dependent pure Nash equilibrium. However, it is worth to point out that pure stationary Nash equilibria need not exist in this class of games. A similar remark applies to irreducible stochastic games of perfect information with the limiting average payoff criterion. Counterexamples are described in Federgruen (1978) and Küenle (1994).

We close this section with a remark on "folk theorems" for stochastic games. It is worth mentioning that the techniques, based on threat strategies utilized very often in repeated games, cannot be immediately adapted to stochastic games, where the players use randomized (behavioral) strategies. Deviations are difficult to discover when the actions are selected at random. However, some folk theorems for various classes of stochastic games were proved in Dutta (1995), Fudenberg and Yamamoto (2011), Hörner et al. (2011, 2014), and Pęski and Wiseman (2015). Further comments can be found in Solan and Zillotto (2016).

Abreu et al. (1986, 1990) applied a method for analysing subgame-perfect equilibria in discounted repeated games that resembles the dynamic programming technique. The set of equilibrium payoffs is a set-valued fixed points of some naturally defined operator. A similar idea was used in stochastic games by Mertens and Parthasarathy (1991). The fixed point property for subgame-perfect equilibrium payoffs can be used to develop algorithms. Berg (2016) and Kitti (2016) considered some modifications of the aforementioned methods for discounted stochastic games with finite state spaces. They also demonstrated some techniques for computing (non-stationary) subgame-perfect equilibria in pure strategies provided that they exist. Sleet and Yeltekin (2016) applied the methods of Abreu et al. (1986, 1990) to some classes of dynamic games and provided a new approach for computing equilibrium value correspondences. Their idea is based on outer and inner approximations of the equilibrium value correspondence via step set-valued functions.

10 Nonzero-Sum Stochastic Games with Imperfect Monitoring

There are only a few papers on nonzero-sum stochastic games with imperfect monitoring (or incomplete information). Although in many models an equilibrium does not exist, some positive results were obtained for repeated games; see Forges (1992), Chap. IX in Mertens et al. (2015) and references cited therein. Altman et al. (2005, 2008) studied stochastic games, in which every player can only observe and control his "private state" and the state of the world is composed of the vector of private states. Moreover, the players do not observe the actions of their partners in the game. Such models of games are motivated by certain examples in wireless communications.

In the model of Altman et al. (2008), the state space $X = \prod_{i=1}^{n} X_i$, where X_i is a finite set of private states of player $i \in N$. The action space $A_i(x_i)$ of player $i \in N$ depends on $x_i \in X_i$ and is finite. It is assumed that player $i \in N$ has no information about the payoffs called costs. Hence, player i only knows the history of his private state process and the action chosen by himself in the past. Thus, a strategy π_i of player $i \in N$ is independent of realizations of state processes of other players and their actions. If $x = (x_1, \ldots, x_n) \in X$ is a state at some period of the game and $a = (a_1, \ldots, a_n)$ is the action profile selected independently by the players at that state, then the probability of going to state $y = (y_1, \ldots, y_n)$ is $q(y|x, a) = q_1(y_1|x_1, a_1) \cdots q_n(y_n|x_n, a_n)$, where $q_i(\cdot|x_i, a_i) \in \Pr(A_i(x_i))$. Thus the coordinate (or private) state processes are independent. It is assumed that every player i has a probability distribution ν_i of the initial state $x_i \in X_i$ and that the initial private states are independent. The initial distribution ν of the state $x \in X$ is determined by $\nu_1, \ldots, \nu_n$ in an obvious way and is known by the players. Further, it is supposed that player $i \in N$ is given some stage cost functions $c_i^j(x, a)$ ($j = 0, 1, \ldots, n_i$) depending on $x \in X$ and action profiles a available in that state. The cost function c_i^0 is to be minimized by player i in the long run, and c_i^j (for $j > 0$) are the costs that must satisfy some constraints described below.

Any strategy profile π together with the initial distribution ν and the transition probability q induces a unique probability measure on the space of all infinite plays. The expectation operator with respect to this measure is denoted by E_ν^π. The expected limit-average cost $C_i^j(\pi)$ is defined as follows:

$$C_i^j(\pi) := \limsup_{T\to\infty} \frac{1}{T} E_\nu^\pi \left(\sum_{t=1}^{T} c_i^j(x^t, a^t) \right).$$

Note that $x^t \in X$ and a^t is an action profile of all the players.

Let $b_i^j > 0$ ($j = 1, \ldots, n_i$) be bounds used to define constraints below. A strategy profile π is i-feasible if

$$C_i^j(\pi) \le b_i^j \quad \text{for each} \quad j = 1, \ldots, n_i.$$

Thus, π is feasible if it is i-feasible for every player $i \in N$.

A strategy profile π^* is called a *constrained Nash equilibrium,* if π^* is feasible and for every player $i \in N$ and his strategy π_i such that the profile (π_i, π^*_{-i}) is i-feasible, we have

$$C_i^0(\pi) \leq C_i^0(\pi_i, \pi^*_{-i}).$$

Note that a unilateral deviation of player i may increase his cost or it may violate his constraints. The aforementioned fact is illustrated in Altman et al. (2008) by an example in wireless communications.

Altman et al. (2008) made the following assumptions.

(I1) (*Ergodicity*) For every player $i \in N$ and any stationary strategy the state process on X_i is an irreducible Markov chain with one ergodic class and possibly some transient states.

(I2) (*Strong Slater condition*) There exists some $\eta > 0$ such that every player $i \in N$ has a strategy π_i^η with the property that for any strategy profile π_{-i} of other players

$$C_i^j(\pi_i^\eta, \pi_{-i}) \leq b_i^j - \eta \quad \text{for all} \quad j = 1, \ldots, n_i.$$

(I3) (*Information*) The players do not observe their costs.

Theorem 24. *Consider the game model that satisfies conditions* (I1)–(I3). *Then there exists a stationary constrained Nash equilibrium.*

Stochastic games with finite sets of states and actions and *imperfect public monitoring* were studied in Fudenberg and Yamamoto (2011) and Hörner et al. (2011). The players, in their models, observe states and receive only public signals on the chosen actions by the partners in the game. Fudenberg and Yamamoto (2011) and Hörner et al. (2011) established "folk theorems" for stochastic games under assumptions that relate to "irreducibility" conditions on the transition probability function. Moreover, Hörner et al. (2011, 2014) also studied algorithms for both computing the sets of all equilibrium payoffs in the normalized discounted games and for finding their limit as the discount factor tends to one. As shown in counterexamples in Flesch et al. (2003) an n-person stochastic game with non-observable actions of the players (and no public signals), observable payoffs and the expected limit-average payoff criterion does not possess ϵ-equilibrium. Cole and Kocherlakota (2001) studied discounted stochastic games with hidden states and actions. They provided an algorithm for finding a sequential equilibrium, where strategies depend on private information only through the privately observed state. Imperfect monitoring is also assumed in the model of the supermodular stochastic game studied by Balbus et al. (2013b), where the monotone convergence of Nash equilibrium payoffs in finite-stage games is proved.

11 Intergenerational Stochastic Games

This section develops a concept of equilibrium behavior and establishes its existence in various intergenerational games. Both paternalistic and non-paternalistic altruism cases are discussed. Consider an infinite sequence of generations labelled by $t \in \mathbb{N}$. There is a single good (called also a renewable resource) that can be used for consumption or productive investment. The set of all resource stocks S is an interval in $\mathbb{R}$. It is assumed that $0 \in S$. Every generation lives one period and derives utility from its own consumption and consumptions of some or all its descendants. Generation t observes the current stock $s_t \in S$ and chooses $a_t \in A(s_t) := [0, s_t]$ for consumption. The remaining part $y_t = s_t - a_t$ is left as an investment for its descendants. The next generation's inheritance or endowment is determined by a *weakly continuous* transition probability q from S to S (stochastic production function), which depends on $y_t \in A(s_t) \subset S$. Recall that the weak continuity of q means that $q(\cdot|y_m) \Rightarrow q(\cdot|y_0)$ if $y_m \to y_0$ in S (as $m \to \infty$). Usually, it is assumed that state 0 is absorbing, i.e., $q(\{0\}|0) = 1$. Let Φ be the set of all Borel functions $\phi : S \to S$ such that $\phi(s) \in A(s)$ for each $s \in S$. A *strategy* for generation t is a function $\phi_t \in \Phi$. If $\phi_t = \phi$ for all $t \in \mathbb{N}$ and some $\phi \in \Phi$, then we say that the generations employ a *stationary strategy*.

Suppose that all generations from $t + 1$ onward use a consumption strategy $c \in \Phi$. Then, in the *paternalistic model* generation t's utility when it consumes $a_t \in A(s_t)$ equals to $H(a_t, c)(s_t)$, where H is some real-valued function used for measurement of the satisfaction level of the generation. This implies that in models with paternalistic altruism each generation derives its utility from its own consumption and the *consumptions* of its successor or successors.

Such a game model reveals a time inconsistency. Strotz (1956) and Pollak (1968) were among the first, who noted this fact in the model of an economic agent whose preferences change over time. In related works, Phelps and Pollak (1968) and Peleg and Yaari (1973) observed that this situation is formally equivalent to one, in which decisions are made by a sequence of heterogeneous planners. They investigated the existence of consistent plans, what we shall call (stationary) Markov perfect equilibria. The solution concept is in fact a symmetric Nash equilibrium $(c^*, c^*, \ldots)$ in a game played by countably many short-lived players having the same utility functions. Therefore, we can say that a *stationary Markov perfect equilibrium* $(c^*, c^*, \ldots)$ corresponds to a strategy $c^* \in \Phi$ such that

$$H(c^*(s), c^*)(s) = \sup_{a \in A(s)} H(a, c^*)(s)$$

for every $s \in S$. We identify this equilibrium with c^*.

In other words, $c^* \in \Phi$ is a stationary Markov perfect equilibrium if

$$c^*(s) \in \arg\max_{a \in A(s)} H(a, c^*)(s) \quad \text{for each } s \in S.$$

There is now a substantial body of work on paternalistic models, see for instance, Alj and Haurie (1983), Harris and Laibson (2001), and Nowak (2010) and the results presented below in this section. At the beginning we consider three types of games, in which the existence of a stationary Markov perfect equilibrium was proved in a sequence of papers: Balbus et al. (2015a,b,c). Game (G1) describes a purely deterministic case, while games (G2) and (G3) deal with a stochastic production function. However, (G2) concerns a model with one descendant, whereas (G3) examines a model with infinitely many descendants. Let us mention that by an intergenerational game with k (k is finite or infinite) descendants (successors or followers), we mean a game in which each generation derives its utility from its own consumption and consumptions of its k descendants.

(G1) Let $S := [0, +\infty)$. Assume that $q(\cdot|y_t) = \delta_{p(y_t)}(\cdot)$, where $p : S \to S$ is a continuous and increasing production function such that $p(0) = 0$. We also assume that

$$H(a, c)(s) = \hat{u}(a, c(p(s - a)))$$

for some continuous and increasing in each variable function $\hat{u} : \mathbb{R}^2_+ \to \mathbb{R} \cup \{-\infty\}$. Moreover, we allow $\hat{u}$ to be unbounded from below. Hence, we assume that $\hat{u}(0, y) \geq -\infty$ for all $y \geq 0$ and $\hat{u}(x, 0) > -\infty$ for all $x > 0$. Furthermore, for any $y_1 > y_2$ in S and $h > 0$, we assume that the function $\Delta_h \hat{u}(x) := \hat{u}(x, y_1) - \hat{u}(x + h, y_2)$ has the *strict single crossing property* on $(0, +\infty)$, i.e., $\Delta_h \hat{u}(x) \geq 0$ implies that $\Delta_h \hat{u}(x') > 0$ for each $x' > x$ (see Milgrom and Shannon 1994).

(G2) Let $S := [0, +\infty)$. We study a model with a utility that reflects a generation's attitude toward risk. This fact is reflected by a positive risk coefficient r. In this setup, H takes the following form:

$$H(a, c)(s) = \begin{cases} u(a) + \beta \int_S v(c(s'))q(ds'|s - a), & \text{for} \quad r = 0 \\ u(a) - \frac{\beta}{r} \ln \int_S e^{-rv(c(s'))} q(ds'|s - a), & \text{for} \quad r > 0, \end{cases}$$

where $u : S \to \mathbb{R} \cup \{-\infty\}$ is increasing, strictly concave, continuous on $(0, +\infty)$ and $u(0) \geq -\infty$. In addition, the function $v : S \to \mathbb{R}$ is bounded, continuous and increasing. Further assumptions are as follows: for every $s \in S$, the set $Z_s := \{y \in S : \; q(\{s\}|y) > 0\}$ is countable and the transition law q is stochastically increasing. The latter fact means that, if $z \to Q(z|y)$ is the cumulative distribution function for $q(\cdot|y)$, then for all $y_1 < y_2$ and $z \in S$, we have $Q(z|y_1) \geq Q(z|y_2)$.

(G3) Let $S := [0, \bar{s}]$ for some $\bar{s} > 0$. In this case, we assume that the utility function of current generation t is as follows:

$$H(a, c)(s) = \tilde{u}(a) + E^c_s[w(a_{t+1}, a_{t+2}, \ldots)],$$

where $w : S^\infty \to \mathbb{R}$ is continuous and $\tilde{u} : S \mapsto \mathbb{R}$ is continuous, strictly concave and increasing. Here, E_s^c is an expectation operator with respect to the unique probability measure on the space of all feasible future histories (starting from the endowment s of generation t) of the consumption-investment process induced by the stationary strategy $c \in \Phi$ used by each generation τ ($\tau > t$) and the transition probability q. The function $\tilde{u}$ is also assumed to be continuous and strictly concave. Defining

$$\tilde{J}(c)(s) = E_s^c[w(a_k, a_{k+1}, a_{k+2}, \ldots)]$$

for every $k \in \mathbb{N}$, we obtain that

$$H(a, c)(s) = \tilde{u}(a) + \int_S \tilde{J}(c)(s')q(ds'|s-a).$$

In addition, $q(\cdot|y)$ is assumed to be nonatomic for $y > 0$.

Let I denote the set of nondecreasing lower semicontinuous functions $i : S \to \mathbb{R}$ such that $i(s) \in A(s)$ for each $s \in S$. Note that every $i \in I$ is continuous from the left and has at most a countable number of discontinuity points. Put

$$F := \{c \in \Phi : \ c(s) = s - i(s), \ i \in I, \ s \in S\}.$$

Clearly, any $c \in F$ is upper semicontinuous and continuous from the left. The idea of using the class F of strategies for analysing equilibria in deterministic bequest games comes from Bernheim and Ray (1983). Further, it was successfully applied to the study of other classes of dynamic games with simultaneous moves; see Sundaram (1989a) and Majumdar and Sundaram (1991).

Theorem 25. *Every intergenerational game* (G1), (G2) *and* (G3) *possesses a stationary Markov perfect equilibrium* $c^* \in F$.

The main idea of the proof is based upon the consideration of an operator L defined as follows: to each consumption strategy $c \in F$ used by descendant (or descendants) the function L assigns the maximal element c_0 from the set of best responses to c. It is shown that $c_0 \in F$. Moreover, F can be viewed as a convex subset of the vector space Y of real-valued continuous from the left functions $\eta : S \mapsto \mathbb{R}$ of bounded variation on every interval $S_n := [0, n]$, $n \in \mathbb{N}$, thus in particular on $[0, \bar{s}]$. We further equip Y with the topology of weak convergence. We assume that (η^m) *converges weakly* to some $\eta^0 \in Y$, if $\lim_{m\to\infty} \eta^m(s) = \eta^0(s)$ for every continuity point s of η^0. Then, due to Lemma 2 in Balbus et al. (2015c), F is compact and metrizable. Finally, the equilibrium point is obtained via the Schauder-Tychonoff fixed point theorem applied to the operator L.

Theorem 25 for game (G1) was proved in Balbus et al. (2015c). Related results for the purely deterministic case were considered by Bernheim and Ray (1983) and Leininger (1986). For instance, Leininger (1986) studied a class $\mathcal{U}$ of bounded from below utility functions for which every selector of the best response correspondence is nondecreasing. In particular, he noticed that this class is nonempty and it includes, for instance, the separable case, i.e., $u(x, y) = v(x) + bv(y)$, where v is strictly increasing and concave and $b > 0$. Bernheim and Ray (1983), on the other hand, showed that the functions u that are strictly concave in their first argument and satisfying the so-called increasing differences property (see Sect. 2) also belong to $\mathcal{U}$. Other functions u that meet conditions imposed by Bernheim and Ray (1983) and Leininger (1986) are of the form $u(x, y) = v_1(x)v_2(y)$, where v_1 is strictly concave and $v_2 \geq 0$ is continuous and increasing. The class $\mathcal{U}$ is not fully characterized. The class (G1) of games includes all above-mentioned examples and some new ones. Our result is also valid for a larger class of utilities that can be unbounded from below. Therefore, Theorem 25 is a generalization of Theorem 4.2 in Bernheim and Ray (1983) and Theorem 3 in Leininger (1986). The proofs given by Bernheim and Ray (1983) and Leininger (1986) do not work for unbounded utility functions. Indeed, Leininger (1986) uses a transformation of upper semicontinuous consumption strategies into the set of Lipschitz functions with constant 1. This clever "levelling" operation enables him to equip the space of continuous functions on the interval $[0, \bar{y}]$ with the topology of uniform convergence and to apply the Schauder fixed point theorem. His proof strongly makes use of the uniform continuity of u. This is the case, when the production function crosses the 45° line. If the production function does not cross the 45° line, a stationary equilibrium is then obtained as a limit of equilibria corresponding to the truncations of the production function. However, this part of the proof is descriptive and sketchy. Bernheim and Ray (1983), on the other hand, identify with the maximal best response consumption strategy, which is upper semicontinuous, a convex-valued upper hemicontinuous correspondence. Then, such a space of upper hemicontinuous correspondences is equipped with the Hausdorff topology. This fact implies the strategy space is compact, if endowments have an upper bound, i.e., when the production function p crosses the 45° line. If this is not satisfied, then a similar approximation technique as in Leininger (1986) is employed. Our proof does not follow the above-mentioned approximation methods. The weak topology introduced in the space Y implies that F is compact and allows to use an elementary but non-trivial analysis. For examples of deterministic bequest games with stationary Markov perfect equilibria given in closed form the reader is referred to Fudenberg and Tirole (1991) and Nowak (2006b, 2010).

Theorem 25 for game (G2) was proved by Balbus et al. (2015b), whereas for game (G3) by Balbus et al. (2015a). Within the stochastic framework, Theorem 25 is an attempt of saving the result reported by Bernheim and Ray (1989) on the existence of stationary Markov perfect equilibria in games with very general utility function and nonatomic shocks. If q is allowed to possess atoms, then a stationary Markov perfect equilibrium exists in the bequest games with one follower (see Theorems 1–2 in Balbus et al. 2015b). The latter result also embraces the purely deterministic case; see Example 1 in Balbus et al. (2015b), where the nature and role of assumptions are discussed. However, as shown in Example 3 in Balbus et al. (2015a), the existence of stationary Markov perfect equilibria in the class of F cannot be proved in intergenerational games where q has atoms and there are more than one descendant.

The result in Bernheim and Ray (1986) concerns "consistent plans" in models with finite time horizon. The problem is then simpler. The results of Bernheim and Ray (1986) were considerably extended by Harris (1985) in his paper on perfect equilibria in some classes of games of perfect information. It should be noted that there are other papers that contain certain results for bequest games with stochastic production function. Amir (1996b) studied games with one descendant for every generation and the transition probability such that the induced cumulative distribution function $Q(z|y)$ is convex in $y \in S$. This condition is rather restrictive. Nowak (2006a) considered similar games in which the transition probability is a convex combination of the Dirac measure at state $s = 0$ and some transition probability from S to S with coefficients depending on investments. Similar models were considered by Balbus et al. (2012, 2013a). The latter paper also studies some computational issues for stationary Markov perfect equilibria. One should note, however, that the transition probabilities in the aforementioned works are specific. However, the transition structure in Balbus et al. (2015a,b) is consistent with the transitions used in the theory of economic growth; see Bhattacharya and Majumdar (2007) and Stokey et al. (1989).

The interesting issue studied in the economics literature concerns the limiting behavior of the state process induced by a stationary Markov perfect equilibrium. Below we formulate a steady state result for a stationary Markov perfect equilibrium obtained for the game (G1). Under slightly more restrictive conditions it was shown by Bernheim and Ray (1987) that the equilibrium capital stock never exceeds the optimal planning stock in any period. Namely, it is assumed that

(B1) p is strictly concave, continuously differentiable and $\lim_{y\to 0+} p'(y) > 1$, $\lim_{y\to\infty} p'(y) < 1/\beta$, where $\beta \in (0, 1]$ is a discount factor;
(B2) $\hat{u}(a_t, a_{t+1}) = \hat{v}(a_t) + \beta\hat{v}(a_{t+1})$, where $\hat{v} : S \to \mathbb{R}$ is increasing, continuously differentiable, strictly concave and $\hat{v}(a) \to \infty$ as $a \to \infty$.

An *optimal consumption program* $\hat{a} := (\hat{a}_t)_{t\in\mathbb{N}}$ is the one which maximizes $\sum_{t=1}^{\infty} \beta^{t-1}\hat{v}(\hat{a}_t)$ subject to all feasibility constraints described in the model. The following result is stated as Theorems 3.2 and 3.3 in Bernheim and Ray (1987).

Theorem 26. *Assume* (B1)–(B2) *and consider game* (G1). *If c^* is a stationary Markov perfect equilibrium, then $i^*(s) = s - c^*(s) \le \hat{y}$, where $\hat{y} \in [0, \infty)$ is the limit of the sequence $(s_t - \hat{a}_t)_{t\in\mathbb{N}}$. If $\hat{y} > 0$, it solves $\beta p'(y) = 1$. If $\lim_{y\to 0+} p'(y) > 1/\beta$, then $\hat{y} > 0$.*

For further properties of stationary Markov perfect equilibria such as efficiency, and Pareto optimality, the reader is referred to Sect. 4 in Bernheim and Ray (1987).

For stochastic models it is of some interest to know whether a stationary Markov perfect equilibrium induces a Markov process having an invariant distribution. It turns out that the answer is positive if an additional stochastic monotonicity requirement is met:

(B3) If $y_1 < y_2$, then for any nondecreasing Borel measurable function $h : S \to \mathbb{R}$,

$$\int_S h(s)q(ds|y_1) \le \int_S h(s)q(ds|y_2).$$

By Theorem 25 for game (G3), there exists $c^* \in F$. Then $s \to i^*(s) = s - c^*(s)$ is nondecreasing on S. Put $q^*(B|s) := q(B|i^*(s))$ where B is a Borel subset of S and $s \in S$. From (B3), it follows that $s \to q^*(\cdot|s)$ is nondecreasing. Define the mapping $\Psi : \Pr(S) \to \Pr(S)$ by

$$\Psi\sigma(B) := \int_S q^*(B|s)\sigma(ds)$$

where $B \in \mathcal{B}(S)$. An *invariant distribution* for the Markov process induced by the transition probability q^* determined by i^* (and thus by c^*) is any fixed point of Ψ. Let $\Delta(q^*)$ be the set of invariant distributions for the process induced by q^*. In Sect. 4 in Balbus et al. (2015a), the following result was proved.

Theorem 27. *Assume* (B3) *and consider game* (G3). *Then, the set of invariant distributions $\Delta(q^*)$ is compact in the weak topology on* $\Pr(S)$.

For each $\sigma \in \Delta(q^*)$, $M(\sigma) := \int_S s\sigma(ds)$ is the mean of distribution σ. By Theorem 27, there exists σ^{**} with the highest mean over the set $\Delta(q^*)$.

One can ask for the uniqueness of invariant distribution. Theorem 4 in Balbus et al. (2015a) yields a positive answer to this question. However, this result concerns the model with multiplicative shocks, i.e., q is induced by the equation

$$s_{t+1} = f(y_t)\xi_t, \quad t \in \mathbb{N},$$

where $f : S \to S$ is a continuous increasing function such that $f(0) > 0$. In addition, there is a state $\hat{s} \in (0, \infty)$ such that $f(y) > y$ for $y \in (0, \hat{s})$ and $f(y) < y$ for $y \in (\hat{s}, \infty)$. Here $(\xi_t)_{t \in \mathbb{N}}$ is an i.i.d. sequence with the nonatomic distribution π. Assuming additionally the *monotone mixing condition,* we conclude from Theorem 4 in Balbus et al. (2015a) the uniqueness of the invariant distribution. Further discussion on these issues can be found in Stokey et al. (1989), Hopenhayn and Prescott (1992), Stachurski (2009), Balbus et al. (2015a) and the references cited therein.

In contrast to the paternalistic model one can also think of a *non-paternalistic* altruism. This notion is concerned with a model, in which each generation's utility is derived from its own consumption and the *utilities* of its all successors. The most general model with non-paternalistic altruism was formulated by Ray (1987). His work is of some importance, because it provides a proper definition of an equilibrium for the non-paternalistic case. According to Ray (1987), a stationary equilibrium consists of a pair of two functions: a saving policy (or strategy) and an indirect utility function. Such a pair constitutes an equilibrium if it is optimal for the current generation, provided its descendants use the same saving strategy and the same indirect utility function.

Assume that the generations from t onward use a consumption strategy $c \in \Phi$. Then, the expected utility of generation t, that inherits an endowment $s_t = s \in S := [0, \bar{s}]$, is of the form

$$W_t(c, v)(s) := (1 - \beta)\tilde{u}(c(s)) + \beta E_s^c[w(v(s_{t+1}), v(s_{t+2}), \ldots)]. \tag{6.8}$$

where $\tilde{u} : S \to K$ and $w : K^\infty \to K$ are continuous functions and $K := [0, \bar{k}]$ with some $\bar{k} \geq \bar{s}$. The function $v : S \to K$ is called an *indirect utility* and is assumed to be Borel measurable. Similarly, for any $c \in \Phi$ and $s = s_{t+1} \in S$, we can define

$$J(c, v)(s) := E_s^c[w(v(s_{t+2}), v(s_{t+3}), \ldots)],$$

which yields

$$W(c, v)(s) := W_t(c, v)(s) = (1 - \beta)\tilde{u}(c(s)) + \beta \int_S J(c, v)(s')q(ds'|s - c(s)).$$

Let us define

$$P(a, c, v)(s) := (1 - \beta)\tilde{u}(a) + \beta \int_S J(c, v)(s')q(ds'|s - a),$$

where $s \in S$, $a \in A(s)$ and $c \in \Phi$. If $s_t = s$, then $P(a, c, v)(s)$ is the utility for generation t choosing the consumption level $a \in A(s_t)$ in this state under the assumption that all future generations will employ a stationary strategy $c \in \Phi$ and the indirect utility is v.

A *stationary equilibrium* in the sense of Ray (1987) is a pair (c^*, v^*), with $c^* \in \Phi$, and $v^* : S \to K$ being a bounded Borel measurable function such that for every $s \in S$, we have that

$$v^*(s) = \sup_{a \in A(s)} P(a, c^*, v^*)(s) = P(c^*(s), c^*, v^*)(s) = W(c^*, v^*)(s). \tag{6.9}$$

Note that equality (6.9) says that there exist an indirect utility function v^* and a consumption strategy c^*, both depending on the current endowment, such that each generation finds it optimal to adopt this consumption strategy provided its descendants use the same strategy and exhibit the given indirect utility.

Let V be the set of all nondecreasing upper semicontinuous functions $v : S \to K$. Note that every $v \in V$ is continuous from the right and has at most a countable number of discontinuity points. By I we denote the subset of all functions $\varphi \in V$ such that $\varphi(s) \in A(s)$ for each $s \in S$. Let $F = \{c : c(s) = s - i(s),\ s \in S,\ i \in I\}$. We impose similar conditions to those imposed on model (G3). Namely, we shall assume that $\tilde{u}$ is strictly concave and increasing. Then, the following result holds.

Theorem 28. *In a non-paternalistic game as described above with nonatomic transitions, there exists a stationary equilibrium* $(c^*, v^*) \in F \times V$.

Theorem 28 was established as Theorem 1 in Balbus et al. (2016). Ray (1987) analysed games with non-paternalistic altruism and deterministic production functions. Unfortunately, his proof contains a mistake. The above result is strongly based on the assumption that the transitions are nonatomic and weakly continuous. The problem in the deterministic model of Ray (1987) remains open. However, Theorem 28 implies that an equilibrium exists if a "small nonatomic noise" is added to the deterministic transition function.

There is a great deal of work devoted to the so-called hyperbolic decision makers, in which the function w in (G3) has a specific form. Namely,

$$w(a_k, a_{k+1}, a_{k+2}, \ldots) = \alpha\beta \sum_{m=k}^{\infty} \beta^{m-k} \tilde{u}(a_m), \tag{6.10}$$

where $\alpha > 0$ and is interpreted as a *short-run* discount factor and $\beta < 1$ is known as a *long-run* discount coefficient. This model was studied by Harris and Laibson (2001) with the transition function defined via the difference equation

$$s_{t+1} = R(s_t - a_t) + \xi_t, \qquad R \ge 0 \quad \text{and} \quad t \in \mathbb{N}.$$

The random variables $(\xi_t)_{t\in\mathbb{N}}$ are nonnegative, Independent, and identically distributed with respect to a nonatomic probability measure. The function $\tilde{u}$ satisfies some restrictive condition concerning the risk aversion of the decisionmaker, but it may be unbounded from above. Working in the class of strategies with locally bounded variation, Harris and Laibson (2001) showed the existence of a stationary Markov perfect equilibrium in their model with concave utility function $\tilde{u}$. They also derived a strong hyperbolic Euler relation. The model considered by Harris and Laibson (2001) can also be viewed as a game between generations; see Balbus and Nowak (2008), Nowak (2010), and Jaśkiewicz and Nowak (2014a) where related versions are studied. However, its main interpretation in the economics literature says that it is a decision problem where the utility of an economic agent changes over time. Thus, the agent is represented by a *sequence of selves* and the problem is to find a time-consistent solution . This solution is actually a stationary Markov perfect equilibrium obtained by thinking about selves as players in an intergenerational game. For further details and references, the reader is referred to Harris and Laibson (2001) and Jaśkiewicz and Nowak (2014a). A decision model with time-inconsistent preferences involving selves who can stop the process at any stage was recently studied by Cingiz et al. (2016)

The model with the function w defined in (6.10) can be extended by adding to the transition probabilities an unknown parameter θ. Then, the natural solution for such a model is a *robust Markov perfect equilibrium*. Roughly speaking, this solution is based on the assumption that the generations involved in the game are risk-sensitive and accept a maxmin utility. More precisely, let Θ be a nonempty Borel subset of Euclidean space $\mathbb{R}^m$ ($m \geq 1$). Then, the endowment s_{t+1} for generation $t+1$ is determined by the transition q from $S \times \Theta$ to S that depends on the investment $y_t \in A(s_t)$ and a parameter $\theta_t \in \Theta$. This parameter is chosen according to a certain probability measure $\gamma_t \in \mathcal{P}$, where $\mathcal{P}$ denotes the action set of *nature* and it is assumed to be a Borel subset of $\Pr(\Theta)$.

Let Γ be the set of all sequences $(\gamma_t)_{t\in\mathbb{N}}$ of Borel measurable mappings $\gamma_t : D \to \mathcal{P}$, where $D = \{(s,a) : s \in S, a \in A(s)\}$. For any $t \in \mathbb{N}$ and $\gamma = (\gamma_t)_{t\in\mathbb{N}} \in \Gamma$, we set $\gamma^t := (\gamma_\tau)_{\tau\geq t}$. Clearly, $\gamma^t \in \Gamma$. A *Markov strategy* for *nature* is a sequence $\gamma = (\gamma_t)_{t\in\mathbb{N}} \in \Gamma$. Note that γ^t can be called a Markov strategy used by nature from period t onward.

For any $t \in \mathbb{N}$, define H^t as the set of all sequences

$$h^t = (a_t, \theta_t, s_{t+1}, a_{t+1}, \theta_{t+1}, \ldots), \quad \text{where } (s_k, a_k) \in D,\ \theta_k \in \Theta \text{ and } k \geq t.$$

H^t is the set of all feasible future histories of the process from period t onward. Endow H^t with the product σ-algebra. Assume in addition that $\tilde{u} \leq 0$, the generations employ a stationary strategy $c \in \Phi$ and nature chooses some $\gamma \in \Gamma$. Then the choice of nature is a probability measure depending on $(s_t, c(s_t))$. Let $E_{s_t}^{c,\gamma^t}$ denote as usual the expectation operator corresponding to the unique probability measure on H^t induced by a stationary strategy $c \in \Phi$ used by each generation τ ($\tau \geq t$), a Markov strategy of nature $\gamma^t \in \Gamma$ and the transition probability q. Assume that all generations from t onward use $c \in \Phi$ and nature applies a strategy $\gamma^t \in \Gamma$. Then, the generation t's expected utility is of the following form:

$$\hat{W}(c)(s_t) := \inf_{\gamma^t\in\Gamma} E_{s_t}^{c,\gamma^t} \left(\tilde{u}(c(s_t)) + \alpha\beta \sum_{m=t+1}^{\infty} \beta^{m-t-1} \tilde{u}(c(s_\tau)) \right).$$

This definition of utility in an intergenerational game provides an intuitive notion of ambiguity aversion, which can be regarded as the generations' diffidence for any lack of precise definition of uncertainty, something that provides room for the malevolent influence of nature. Letting

$$\hat{J}(c)(s_j) = \inf_{\gamma^j\in\Gamma} E_{s_j}^{c,\gamma^j} \left(\sum_{m=j}^{\infty} \beta^{m-j} \tilde{u}(c(s_\tau)) \right)$$

we one can show that

$$\hat{W}(c)(s_t) = \tilde{u}(c(s)) + \inf_{\xi \in \mathcal{P}} \alpha\beta \int_S \hat{J}(c)(s_{t+1})q(ds_{t+1}|s_t - c(s_t), \xi).$$

For any $s \in S$, $a \in A(s)$, and $c \in \Phi$, we set

$$\hat{P}(a, c)(s) = \tilde{u}(a) + \inf_{\xi \in \mathcal{P}} \alpha\beta \int_S \hat{J}(c)(s')q(ds'|s - a, \xi).$$

If $s = s_t$, then $\hat{P}(a, c)(s)$ is the utility for generation t choosing $a \in A(s_t)$ in this state when all future generations employ a stationary strategy $c \in \Phi$.

A *robust Markov perfect equilibrium* is a function $c^* \in \Phi$ such that for every $s \in S$ we have

$$\sup_{a \in A(s)} \hat{P}(a, c^*)(s) = \hat{P}(c^*(s), c^*)(s) = \hat{W}(c^*)(s).$$

The existence of a robust Markov perfect equilibrium in the aforementioned model was proved by Balbus et al. (2014) under the assumption that the transition probability is a convex combination of probability measures $\mu_1, \ldots, \mu_l$ on S with coefficients depending on investments $y = s - a$. A robust Markov perfect equilibrium was obtained in the class of functions F under the condition that all measures $\mu_1, \ldots, \mu_l$ are nonatomic. If $\mu_1, \ldots, \mu_l$ have atoms, then some stochastic dominance conditions are imposed, but the equilibrium was obtained in the class of Lipschitz continuous functions with constant 1. A different approach was presented in the work of Jaśkiewicz and Nowak (2014b), where the set of endowments S and the set of consumptions are Borel and the parameter set Θ is finite. Assuming again that the transition probability is a finite convex combination of probability measures $\mu_1, \ldots, \mu_l$ on S depending on the parameter θ with coefficients depending on the inheritance s and consumption level a, they have established a twofold result. First, they proved the existence of a robust Markov perfect equilibrium in the class of randomized strategies. Second, assuming that $\mu_1, \ldots, \mu_l$ are nonatomic, and making use of the purification theorem of Dvoretzky-Wald-Wolfowitz, they replaced a randomized equilibrium by a pure one.

The models of intergenerational games with general spaces of consumptions and endowments were also examined by Jaśkiewicz and Nowak (2014a). A novel feature in this approach is the fact that generation t can employ the *entropic risk measure* to calculate its utilities. More precisely, if Z is a random variable with the distribution π, then its entropic risk measure is $\mathcal{E}(Z) = \frac{1}{r} \ln \int_\Omega e^{rZ(\omega)}\pi(d\omega)$, where $r < 0$ is a risk coefficient. If r is sufficiently close to zero, then making use of the Taylor expansion one can see that

$$\mathcal{E}(Z) \approx EZ + \frac{r}{2}\mathrm{Var}(Z).$$

This means that a generation which uses the entropic risk measure to calculate its utility is risk averse and takes into account not only the expected value of a random future successors' utilities derived from consumptions but also their variance. Assuming that each generation cares only about its m descendants and assuming that the transition probability is a convex combination of finitely many nonatomic measures on the endowment space with coefficients that may depend on s and a, Jaśkiewicz and Nowak (2014a) proved the existence of stationary Markov perfect equilibrium in pure strategies. The same result was shown for games with infinitely many descendants in the case of hyperbolic preferences. In both cases the proof consists of two parts. First, a randomized stationary Markov perfect equilibrium was shown to exist. Second, making use of the specific structure of the transition probability and applying the Dvoretzky-Wald-Wolfowitz theorem a desired pure stationary Markov perfect equilibrium was obtained.

A related game to the above mentioned models is the one with quasi-geometric discounting from the dynamic consumer theory; see Chatterjee and Eyigungor (2016). Particularly, the authors showed that in natural cases such a game does not possess a Markov perfect equilibrium in the class of continuous strategies. However, a continuous Markov perfect equilibrium exists, if the model was reformulated involving lotteries. These two models were then numerically compared. It is known that the numerical analysis of equilibrium in models with hyperbolic (quasi-geometric) discounting shows difficulties in achieving convergence even in a simple, deterministic optimal growth problem that has a smooth closed-form solution. Maliar and Maliar (2016) defined some restrictions on the equilibrium strategies under which the numerical methods studied deliver a unique smooth solution for many deterministic and stochastic models.

Finally, we wish to point out that Markov Perfect Equilibria for stochastic bequest games with transition probabilities and utilities depending on time were shown to exist in Balbus et al. (2017, 2018).

12 Stopping Games

Stopping games were introduced by Dynkin (1969) as a generalization of optimal stopping problems. They were used in several models in economics and operations research, for example, in equipment replacement, job search, and consumer purchase behavior; see Heller (2012).

Dynkin (1969) dealt with the following problem. Two players observe a bivariate sequence of adapted random variables $(X(k), Y(k))_{k\in\mathbb{N}_0}$, where $\mathbb{N}_0 = \mathbb{N} \cup \{0\}$. Player 1 chooses a stopping time τ_1 such that $\{\tau_1 = k\} \subset \{X(k) \geq 0\}$, whereas player 2 selects τ_2 such that $\{\tau_2 = k\} \subset \{X(k) < 0\}$. If $\tau_1 \wedge \tau_2$ is finite, then player 2 pays $Y(\tau)$ to player 1 and the game terminates. Hence, the objective of player 1 (respectively 2) is to maximize (minimize) $R(\tau_1, \tau_2) = E[Y(\tau_1 \wedge \tau_2)]$. Dynkin (1969) characterized ϵ-optimal stopping times and proved that the game has a value provided that $\sup_{k\in\mathbb{N}_0} |Y(k)|$ is integrable. This model was later extended by Kiefer (1971) and Neveu (1975). In particular, Neveu (1975) showed the existence of a game value in a slightly modified model. Namely, he dealt with the following expected payoff function:

$$R(\tau_1, \tau_2) = E[X(\tau_1)1[\tau_1 < \tau_2] + Y(\tau_2)1[\tau_2 \leq \tau_1]],$$

where $(X(k))_{k\in\mathbb{N}_0}$ and $(Y(k))_{k\in\mathbb{N}_0}$ are $\mathbb{R}$-valued adapted stochastic processes such that $\sup_{k\in\mathbb{N}_0}(X^+(k) + Y^-(k))$ are integrable and $X(k) \leq Y(k)$ for all $k \in \mathbb{N}_0$. The game considered by Neveu (1975) was generalized by Yasuda (1985), who dropped the latter assumption on monotonicity. In his model, the expected payoff function takes the following form:

$$R(\tau_1, \tau_2) = E[X(\tau_1)1[\tau_1 < \tau_2] + Y(\tau_2)1[\tau_2 < \tau_1] + Z(\tau_1)1[\tau_1 = \tau_2]],$$

where as usual $(X(k))_{k\in\mathbb{N}_0}$, $(Y(k))_{k\in\mathbb{N}_0}$ and $(Z(k))_{k\in\mathbb{N}_0}$ are adapted integrable random variables. Yasuda (1985) considered randomized strategies instead of pure ones. According to Yasuda (1985) a strategy for a player is an adapted random sequence $p = (p_k)_{k\in\mathbb{N}_0}$ (or $q = (q_k)_{k\in\mathbb{N}_0}$) such that $0 \leq p_k, q_k \leq 1$ with probability one. Here, p_k (or q_k) stands for the probability that the player stops the game at time k conditional on the event that the game was not stopped before. In computing the payoff induced by a pair of strategies (p, q), one assumes that the randomizations performed by the players in various stages are mutually independent and independent of the payoff processes. Thus, a strategy that corresponds to a stopping time σ is $p_k = 0$ on the event $[\sigma > k]$ and $p_k = 1$ on the event $[\sigma \leq k]$. Yasuda (1985) proved the existence of the value in the set of randomized strategies in finite and discounted infinite time horizon problems.

Before formulating the next result, we define the stopping stages for players 1 and 2 by $\theta_1 := \inf\{k \in \mathbb{N}_0 : \ P(k) \leq p_k\}$, and $\theta_2 := \inf\{k \in \mathbb{N}_0 : \ Q(k) \leq q_k\}$, where $(P(k), Q(k))_{k\in\mathbb{N}_0}$ is a double sequence of i.i.d. random variables uniformly distributed over $[0, 1]$ satisfying certain independence assumptions imposed in Rosenberg et al. (2001). Set $\theta = \theta_1 \wedge \theta_2$. Clearly, θ is the stage at which the game stops. Let us define

$$R(p, q) = E[X(\theta_1)1[\theta_1 < \theta_2] + Y(\theta_2)1[\theta_2 < \theta_1] + Z(\theta_1)1[\tau_1 = \tau_2 < +\infty]]$$

and its β-discounted evaluation

$$R_\beta(p, q) = (1 - \beta)E\big[\beta^{\theta+1}(X(\theta_1)1[\theta_1 < \theta_2] + Y(\theta_2)1[\theta_2 < \theta_1] + Z(\theta_1)1[\tau_1 = \tau_2 < +\infty])\big].$$

The following result was proved by Rosenberg et al. (2001).

Theorem 29. *Assume that* $E[\sup_{k\in\mathbb{N}_0}(|X(k)| + |Y(k)| + |Z(k)|)] < +\infty$. *Then the stopping games with the payoffs* $R(p, q)$ *and* $R_\beta(p, q)$ *have values, say* v *and* v_β, *respectively. Moreover,* $\lim_{\beta\to 1} v_\beta = v$.

Let us now turn to nonzero-sum Dynkin games. They were considered in several papers; see, for instance, Ferenstein (2007), Krasnosielska-Kobos (2016), Morimoto (1986), Nowak and Szajowski (1999), Ohtsubo (1987), Ohtsubo (1991), Solan and Vieille (2001), and Szajowski (1994). Obviously, the list of references is by no means exhaustive. We start with presenting a result for two-player nonzero-sum stopping games. Assume that the aforementioned sequences $(X(k))_{k\in\mathbb{N}_0}$, $(Y(k))_{k\in\mathbb{N}_0}$ and $(Z(k))_{k\in\mathbb{N}_0}$ are bounded in $\mathbb{R}^2$ and let ρ be a uniform bound on the payoffs. The payoff of the game is $R(p,q)$ except that $R(p,q) \in \mathbb{R}^2$. Shmaya and Solan (2004) proved the following result.

Theorem 30. *For each $\epsilon > 0$, the stopping game has an ϵ-equilibrium $(p^*_\epsilon, q^*_\epsilon)$.*

Theorem 30 does not hold, if the payoffs are not uniformly bounded. Its proof is based upon a stochastic version of the Ramsey theorem that was also proved by Shmaya and Solan (2004). It states that for every colouring of a complete infinite graph by finitely many colours, there is a complete infinite monochromatic subgraph. Shmaya and Solan (2004) applied a variation of this result to reduce the problem of the existence of an ϵ-equilibrium in a general stopping game to that of studying properties of ϵ-equilibria in a simple class of stochastic games with finite state space. A similar result for deterministic 2-player nonzero-sum stopping games was reported by Shmaya et al. (2003).

All the aforementioned works deal with the two-player case and/or assume some special structure of the payoffs. Recently, Hamadène and Hassani (2014) studied n-person nonzero-sum Dynkin games. Such a game is terminated at $\tau := \tau_1 \wedge \ldots \wedge \tau_n$, where τ_i is a stopping time chosen by player i. Then, the corresponding payoff for player i is given by

$$R_i(\tau_1, \ldots, \tau_n) = W^{i,I}_\tau,$$

where I_s denotes the set of players who make the decision to stop, that is, $I_s = \{m \in \{1,\ldots,n\} : \tau = \tau_m\}$ and W^{i,I_s} is the payoff stochastic process of player i. The main assumption says that the payoff is less when the player belongs to the group involved in the decision to stop than when he is not. Hamadène and Hassani (2014) showed that the game has a Nash equilibrium in pure strategies. The proof is based on the approximation scheme whose limit provides a Nash equilibrium.

Krasnosielska-Kobos and Ferenstein (2013) is another paper that is concerned with multi-person stopping games. More precisely, they consider a game in which players sequentially observe the offers $X(1), X(2), \ldots$ at jump times $T_1, T_2, \ldots$ of a Poisson process. It is assumed that the random variables $X(1), X(2), \ldots$ form an i.i.d. sequence. Each accepted offer results in a reward $R(k) = X(k)r(T_k)$, where r is a non-increasing discount function. If more than one player accepts the offer, then the player with the highest priority gets the reward. By making use of the solution to the multiple optimal stopping time problem with above reward structure, Krasnosielska-Kobos and Ferenstein (2013) constructed a Nash equilibrium which is Pareto efficient.

Mashiah-Yaakovi (2014), on the other hand, studied subgame-perfect equilibria in stopping games. It is assumed that at every stage one of the players is chosen according to a stochastic process, and that player decides whether to continue the interaction or to stop it. The terminal payoff vector is obtained by another stochastic process. Mashiah-Yaakovi (2014) defines a weaker concept of subgame-perfect equilibrium, namely, a δ-approximate subgame-perfect ϵ-equilibrium. A strategy profile is a δ-approximate subgame-perfect ϵ-equilibrium if it induces an ϵ-equilibrium in every subgame, except perhaps a set of subgames that occur with probability at most δ. A 0-approximate subgame-perfect ϵ-equilibrium is actually a subgame-perfect ϵ-equilibrium. The concept of approximate subgame-perfect equilibrium relates to the concept of "trembling-hand perfect equilibrium" introduced by Selten (1975). A stopping game in which, at every stage, one player who decides to stop or continue the game is chosen according to a (periodic in some sense) stochastic process is also studied in Mashiah-Yaakovi (2009). This assumption extends the random priority in stopping games considered, for example, in Szajowski (1994, 1995). Once the chosen player decides to stop, the players receive terminal payoffs that are determined by a second stochastic process. Periodic subgame-perfect ϵ-equilibria in pure strategies are studied under some quite general conditions. Some bases for stopping n-person games with fixed priorities were provided by Enns and Ferenstein (1987).

Finally, it is worth pointing out that there are three notions of random stopping times. The above mentioned randomized strategies used by Yasuda (1985) and Rosenberg et al. (2001) are also called behavioral stopping times. A randomized stopping time, on the other hand, is a nonnegative adapted real-valued process $\rho = (\rho_k)_{k\in\mathbb{N}\cup\{\infty\}}$ that satisfies $\sum_{k\in\mathbb{N}\cup\{\infty\}} \rho_k = 1$. The third concept allows to define mixed stopping times ν. Roughly speaking, they are product measurable functions in which the first coordinate r is chosen according to the uniform distribution over the interval $[0, 1]$ at the outset. Then, the stopping time is $\nu(r, \cdot)$. For more details, the reader is referred to Rosenberg et al. (2001). As communicated to us by Eilon Solan, the classes of mixed and randomized stopping times are equivalent by a proper generalization of Kuhn's theorem.

Acknowledgements We thank Tamer Başar and Georges Zaccour for inviting us to write this chapter and their help. We also thank Elżbieta Ferenstein, János Flesch, Eilon Solan, Yeneng Sun, Krzysztof Szajowski and two reviewers for their comments on an earlier version of this survey.

References

Abreu D, Pearce D, Stacchetti E (1986) Optimal cartel equilibria with imperfect monitoring. J Econ Theory 39:251–269

Abreu D, Pearce D, Stacchetti E (1990) Toward a theory of discounted repeated games with imperfect monitoring. Econometrica 58:1041–1063

Adlakha S, Johari R (2013) Mean field equilibrium in dynamic games with strategic complementarities. Oper Res 61:971–989

Aliprantis C, Border K (2006) Infinite dimensional analysis: a Hitchhiker's guide. Springer, New York

Alj A, Haurie A (1983) Dynamic equilibria in multigenerational stochastic games. IEEE Trans Autom Control 28:193–203

Alós-Ferrer C, Ritzberger K (2015) Characterizing existence of equilibrium for large extensive form games: a necessity result. Econ Theory 63:407–430

Alós-Ferrer C, Ritzberger K (2016) Equilibrium existence for large perfect information games. J Math Econ 62:5–18

Altman E (1996) Non-zero-sum stochastic games in admission, service and routing control in queueing systems. Queueing Syst Theory Appl 23:259–279

Altman E, Avrachenkov K, Bonneau N, Debbah M, El-Azouzi R, Sadoc Menasche D (2008) Constrained cost-coupled stochastic games with independent state processes. Oper Res Lett 36:160–164

Altman E, Avrachenkov K, Marquez R, Miller G (2005) Zero-sum constrained stochastic games with independent state processes. Math Methods Oper Res 62:375–386

Altman E, Hordijk A, Spieksma FM (1997) Contraction conditions for average and α-discount optimality in countable state Markov games with unbounded rewards. Math Oper Res 22: 588–618

Amir R (1996a) Continuous stochastic games of capital accumulation with convex transitions. Games Econ Behavior 15:132–148

Amir R (1996b) Strategic intergenerational bequests with stochastic convex production. Econ Theory 8:367–376

Amir R (2003) Stochastic games in economics: the lattice-theoretic approach. In: Neyman A, Sorin S (eds) Stochastic games and applications. Kluwer, Dordrecht, pp 443–453

Artstein Z (1989) Parametrized integration of multifunctions with applications to control and optimization. SIAM J Control Optim 27:1369–1380

Aumann RJ (1974) Subjectivity and correlation in randomized strategies. J Math Econ 1:67–96

Aumann RJ (1987) Correlated equilibrium as an expression of Bayesian rationality. Econometrica 55:1–18

Balbus Ł, Jaśkiewicz A, Nowak AS (2014) Robust Markov perfect equilibria in a dynamic choice model with quasi-hyperbolic discounting. In: Haunschmied J et al (eds) Dynamic games in economics, dynamic modeling and econometrics in economics and finance 16. Springer, Berlin/Heidelberg, pp 1–22

Balbus Ł, Jaśkiewicz A, Nowak AS (2015a) Existence of stationary Markov perfect equilibria in stochastic altruistic growth economies. J Optim Theory Appl 165:295–315

Balbus Ł, Jaśkiewicz A, Nowak AS (2015b) Stochastic bequest games. Games Econ Behavior 90:247–256

Balbus Ł, Jaśkiewicz A, Nowak AS (2015c) Bequest games with unbounded utility functions. J Math Anal Appl 427:515–524

Balbus Ł, Jaśkiewicz A, Nowak AS (2016) Non-paternalistic intergenerational altruism revisited. J Math Econ 63:27–33

Balbus Ł, Nowak AS (2004) Construction of Nash equilibria in symmetric stochastic games of capital accumulation. Math Methods Oper Res 60:267–277

Balbus Ł, Nowak AS (2008) Existence of perfect equilibria in a class of multigenerational stochastic games of capital accumulation. Automatica 44:1471–1479

Balbus Ł, Reffett K, Woźny Ł (2012) Stationary Markovian equilibria in altruistic stochastic OLG models with limited commitment. J Math Econ 48:115–132

Balbus Ł, Reffett K, Woźny Ł (2013a) A constructive geometrical approach to the uniqueness of Markov stationary equilibrium in stochastic games of intergenerational altruism. J Econ Dyn Control 37:1019–1039

Balbus Ł, Reffett K, Woźny Ł (2013b) Markov stationary equilibria in stochastic supermodular games with imperfect private and public information. Dyn Games Appl 3:187–206

Balbus Ł, Reffett K, Woźny Ł (2014a) Constructive study of Markov equilibria in stochastic games with complementarities. J Econ Theory 150:815–840
Balbus Ł, Jaśkiewicz A, Nowak AS, Woźny L (2017) A note on Markov perfect equilibria in a class of non-stationary bequest games. J Math Anal Appl 456:394–401
Balbus Ł, Jaśkiewicz A, Nowak AS (2018, in press) Markov perfect equilibria in a dynamic decision model with quasi-hyperbolic discounting. Ann Oper Res
Barelli P, Duggan J (2014) A note on semi-Markov perfect equilibria in discounted stochastic games. J Econ Theory 15:596–604
Başar T, Olsder GJ (1995) Dynamic noncooperative game theory. Academic, New York
Berg K (2016) Elementary subpaths in discounted stochastic games. Dyn Games Appl 6(3): 304–323
Bernheim D, Ray D (1983) Altruistic growth economies I. Existence of bequest equilibria. Technical report no 419, Institute for Mathematical Studies in the Social Sciences, Stanford University
Bernheim D, Ray D (1986) On the existence of Markov-consistent plans under production uncertainty. Rev Econ Stud 53:877–882
Bernheim D, Ray D (1987) Economic growth with intergenerational altruism. Rev Econ Stud 54:227–242
Bernheim D, Ray D (1989) Markov perfect equilibria in altruistic growth economies with production uncertainty. J Econ Theory 47:195–202
Bhattacharya R, Majumdar M (2007) Random dynamical systems: theory and applications. Cambridge University Press, Cambridge
Billingsley P (1968) Convergence of probability measures. Wiley, New York
Blackwell D (1965) Discounted dynamic programming. Ann Math Stat 36:226–235
Blackwell D (1969) Infinite G_δ-games with imperfect information. Zastosowania Matematyki (Appl Math) 10:99–101
Borkar VS, Ghosh MK (1993) Denumerable state stochastic games with limiting average payoff. J Optim Theory Appl 76:539–560
Browder FE (1960) On continuity of fixed points under deformations of continuous mappings. Summa Brasiliensis Math 4:183–191
Carlson D, Haurie A (1996) A turnpike theory for infinite-horizon open-loop competitive processes. SIAM J Control Optim 34:1405–1419
Castaing C, Valadier M (1977) Convex analysis and measurable multifunctions. Lecture notes in mathematics, vol 580. Springer, New York
Chatterjee S, Eyigungor B (2016) Continuous Markov equilibria with quasi-geometric discounting. J Econ Theory 163:467–494
Chiarella C, Kemp MC, Van Long N, Okuguchi K (1984) On the economics of international fisheries. Int Econ Rev 25:85–92
Cingiz K, Flesch J, Herings JJP, Predtetchinski A (2016) Doing it now, later, or never. Games Econ Behavior 97:174–185
Cole HL, Kocherlakota N (2001) Dynamic games with hidden actions and hidden states. J Econ Theory 98:114–126
Cottle RW, Pang JS, Stone RE (1992) The linear complementarity problem. Academic, New York
Curtat LO (1996) Markov equilibria of stochastic games with complementarities. Games Econ Behavior 17:177–199
Doraszelski U, Escobar JF (2010) A theory of regular Markov perfect equilibria in dynamic stochastic games: genericity, stability, and purification. Theor Econ 5:369–402
Doraszelski U, Pakes A (2007) A framework for applied dynamic analysis in IO. In: Amstrong M, Porter RH (eds) Handbook of industrial organization, vol 3. North Holland, Amsterdam/ London, pp 1887–1966
Doraszelski U, Satterthwaite M (2010) Computable Markov-perfect industry dynamics. Rand J Econ 41:215–243
Dubins LE, Savage LJ (2014) Inequalities for stochastic processes. Dover, New York

Duffie D, Geanakoplos J, Mas-Colell A, McLennan A (1994) Stationary Markov equilibria. Econometrica 62:745–781

Duggan J (2012) Noisy stochastic games. Econometrica 80:2017–2046

Dutta PK (1995) A folk theorem for stochastic games. J Econ Theory 66:1–32

Dutta PK, Sundaram RK (1992) Markovian equilibrium in class of stochastic games: existence theorems for discounted and undiscounted models. Econ Theory 2:197–214

Dutta PK, Sundaram RK (1993) The tragedy of the commons? Econ Theory 3:413–426

Dutta PK, Sundaram RK (1998) The equilibrium existence problem in general Markovian games. In: Majumdar M (ed) Organizations with incomplete information. Cambridge University Press, Cambridge, pp 159–207

Dynkin EB (1969) The game variant of a problem on optimal stopping. Sov Math Dokl 10:270–274

Dynkin EB, Evstigneev IV (1977) Regular conditional expectations of correspondences. Theory Probab Appl 21:325–338

Eaves B (1972) Homotopies for computation of fixed points. Math Program 3:1–22

Eaves B (1984) A course in triangulations for solving equations with deformations. Springer, Berlin

Elliott RJ, Kalton NJ, Markus L (1973) Saddle-points for linear differential games. SIAM J Control Optim 11:100–112

Enns EG, Ferenstein EZ (1987) On a multi-person time-sequential game with priorities. Seq Anal 6:239–256

Ericson R, Pakes A (1995) Markov-perfect industry dynamics: a framework for empirical work. Rev Econ Stud 62:53–82

Escobar JF (2013) Equilibrium analysis of dynamic models of imperfect competition. Int J Ind Organ 31:92–101

Federgruen A (1978) On N-person stochastic games with denumerable state space. Adv Appl Probab 10:452–471

Ferenstein E (2007) Randomized stopping games and Markov market games. Math Methods Oper Res 66:531–544

Filar JA, Schultz T, Thuijsman F, Vrieze OJ (1991) Nonlinear programming and stationary equilibria in stochastic games. Math Program 50:227–237

Filar JA, Vrieze K (1997) Competitive Markov decision processes. Springer, New York

Fink AM (1964) Equilibrium in a stochastic n-person game. J Sci Hiroshima Univ Ser A-I Math 28:89–93

Flesch J, Schoenmakers G, Vrieze K (2008) Stochastic games on a product state space. Math Oper Res 33:403–420

Flesch J, Schoenmakers G, Vrieze K (2009) Stochastic games on a product state space: the periodic case. Int J Game Theory 38:263–289

Flesch J, Thuijsman F, Vrieze OJ (1997) Cyclic Markov equilibrium in stochastic games. Int J Garne Theory 26:303–314

Flesch J, Thuijsman F, Vrieze OJ (2003) Stochastic games with non-observable actions. Math Methods Oper Res 58:459–475

Flesch J, Kuipers J, Mashiah-Yaakovi A, Schoenmakers G, Solan E, Vrieze K (2010a) Perfect-information games with lower-semicontinuous payoffs. Math Oper Res 35:742–755

Flesch J, Kuipers J, Schoenmakers G, Vrieze K (2010b) Subgame-perfection in positive recursive games with perfect information. Math Oper Res 35:193–207

Flesch J, Kuipers J, Mashiah-Yaakovi A, Schoenmakers G, Shmaya E, Solan E, Vrieze K (2014) Non-existence of subgame-perfect ϵ-equilibrium in perfect-information games with infinite horizon. Int J Game Theory 43:945–951

Flesch J, Thuijsman F, Vrieze OJ (2007) Stochastic games with additive transitions. Eur J Oper Res 179:483–497

Flesch J, Predtetchinski A (2015) On refinements of subgame perfect ϵ-equilibrium. Int J Game Theory 45:523–542

Forges E (1986) An approach to communication equilibria. Econometrica 54:1375–1385

Forges F (1992) Repeated games of incomplete information: nonzero-sum. In: Aumann RJ, Hart S (eds) Handbook of game theory, vol 1. North Holland, Amsterdam, pp 155–177

Forges F (2009) Correlated equilibria and communication in games. In: Meyers RA (ed) Encyclopedia of complexity and systems science. Springer, New York, pp 1587–1596

Fudenberg D, Levine D (1983) Subgame-perfect equilibria of finite and infinite horizon games. J Econ Theory 31:251–268

Fudenberg D, Tirole J (1991) Game theory. MIT Press, Cambridge

Fudenberg D, Yamamoto Y (2011) The folk theorem for irreducible stochastic games with imperfect public monitoring. J Econ Theory 146:1664–1683

Gale D (1967) On optimal development in a multisector economy. Rev Econ Stud 34:1–19

Glicksberg IL (1952) A further generalization of the Kakutani fixed point theorem with application to Nash equilibrium points. Proc Am Math Soc 3:170–174

Govindan S, Wilson R (2003) A global Newton method to compute Nash equilibria. J Econ Theory 110:65–86

Govindan S, Wilson R (2009) Global Newton method for stochastic games. J Econ Theory 144:414–421

Haller H, Lagunoff R (2000) Genericity and Markovian behavior in stochastic games. Econometrica 68:1231–1248

Hamadène S, Hassani M (2014) The multi-player nonzero-sum Dynkin game in discrete time. Math Methods Oper Res 79:179–194

Harris C (1985) Existence and characterization of perfect equilibrium in games of perfect information. Econometrica 53:613–628

Harris C, Laibson D (2001) Dynamic choices of hyperbolic consumers. Econometrica 69:935–957

Harris C, Reny PJ, Robson A (1995) The existence of subgame-perfect equilibrium in continuous games with almost perfect information: a case for public randomization. Econometrica 63: 507–544

Harsanyi JC (1973a) Oddness of the number of equilibrium points: a new proof. Int J Game Theory 2:235–250

Harsanyi JC (1973b) Games with randomly disturbed payoffs: a new rationale for mixed-strategy equilibrium points. Int J Game Theory 2:1–23

Haurie A, Krawczyk JB, Zaccour G (2012) Games and dynamic games. World Scientific, Singapore

He W, Sun Y (2017) Stationary Markov perfect equilibria in discounted stochastic games. J Econ Theory 169:35–61

Heller Y (2012) Sequential correlated equilibria in stopping games. Oper Res 60:209–224

Herings JJP, Peeters RJAP (2004) Stationary equilibria in stochastic games: structure, selection, and computation. J Econ Theory 118:32–60

Herings JJP, Peeters RJAP (2010) Homotopy methods to compute equilibria in game theory. Econ Theory 42:119–156

Himmelberg CJ (1975) Measurable relations. Fundam Math 87:53–72

Himmelberg CJ, Parthasarathy T, Raghavan TES, Van Vleck FS (1976) Existence of p-equilibrium and optimal stationary strategies in stochastic games. Proc Am Math Soc 60:245–251

Hopenhayn H, Prescott E (1992) Stochastic monotonicity and stationary distributions for dynamic economies. Econometrica 60:1387–1406

Hörner J, Sugaya T, Takahashi S, Vieille N (2011) Recursive methods in discounted stochastic games: an algorithm for $\delta \to 1$ and a folk theorem. Econometrica 79:1277–1318

Hörner J, Takahashi S, Vieille N (2014) On the limit perfect public equilibrium payoff set in repeated and stochastic game. Games Econ Behavior 85:70–83

Horst U (2005) Stationary equilibria in discounted stochastic games with weakly interacting players. Games Econ Behavior 51:83–108

Jaśkiewicz A, Nowak AS (2006) Approximation of noncooperative semi-Markov games. J Optim Theory Appl 131:115–134

Jaśkiewicz A, Nowak AS (2014a) Stationary Markov perfect equilibria in risk sensitive stochastic overlapping generations models. J Econ Theory 151:411–447

Jaśkiewicz A, Nowak AS (2014b) Robust Markov perfect equilibria. J Math Anal Appl 419: 1322–1332
Jaśkiewicz A, Nowak AS (2015a) On pure stationary almost Markov Nash equilibria in nonzero-sum ARAT stochastic games. Math Methods Oper Res 81:169–179
Jaśkiewicz A, Nowak AS (2015b) Stochastic games of resource extraction. Automatica 54: 310–316
Jaśkiewicz A, Nowak AS (2016) Stationary almost Markov perfect equilibria in discounted stochastic games. Math Oper Res 41:430–441
Jaśkiewicz A, Nowak AS (2018a) Zero-sum stochastic games. In: Basar T, Zaccour G (eds) Handbook of dynamic game theory, Birkhäuser, Basel
Jaśkiewicz A, Nowak AS (2018b, in press) On symmetric stochastic games of resource extraction with weakly continuous transitions. TOP
Kakutani S (1941) A generalization of Brouwer's fixed point theorem. Duke Math J 8:457–459
Kiefer YI (1971) Optimal stopped games. Theory Probab Appl 16:185–189
Kitti M (2016) Subgame-perfect equilibria in discounted stochastic games. J Math Anal Appl 435:253–266
Klein E, Thompson AC (1984) Theory of correspondences. Wiley, New York
Kohlberg E, Mertens JF (1986) On the strategic stability of equilibria. Econometrica 54:1003–1037
Krasnosielska-Kobos A (2016) Construction of Nash equilibrium based on multiple stopping problem in multi-person game. Math Methods Oper Res (2016) 83:53–70
Krasnosielska-Kobos A, Ferenstein E (2013) Construction of Nash equilibrium in a game version of Elfving's multiple stopping problem. Dyn Games Appl 3:220–235
Krishnamurthy N, Parthasarathy T, Babu S (2012) Existence of stationary equilibrium for mixtures of discounted stochastic games. Curr Sci 103:1003–1013
Krishnamurthy N, Parthasarathy T, Ravindran G (2012) Solving subclasses of multi-player stochastic games via linear complementarity problem formulations – a survey and some new results. Optim Eng 13:435–457
Kuipers J, Flesch J, Schoenmakers G, Vrieze K (2016) Subgame-perfection in recursive perfect information games, where each player controls one state. Int J Game Theory 45:205–237
Kuratowski K, Ryll-Nardzewski C (1965) A general theorem on selectors. Bull Polish Acad Sci (Ser Math) 13:397–403
Küenle HU (1994) On Nash equilibrium solutions in nonzero-sum stochastic games with complete information. Int J Game Theory 23:303–324
Küenle HU (1999) Equilibrium strategies in stochastic games with additive cost and transition structure and Borel state and action spaces. Int Game Theory Rev 1:131–147
Leininger W (1986) The existence of perfect equilibria in model of growth with altruism between generations. Rev Econ Stud 53:349–368
Lemke CE (1965) Bimatrix equilibrium points and mathematical programming. Manag Sci 11:681–689
Lemke CE, Howson JT Jr (1964) Equilibrium points of bimatrix games. SIAM J Appl Math 12:413–423
Levhari D, Mirman L (1980) The great fish war: an example using a dynamic Coumot-Nash solution. Bell J Econ 11:322–334
Levy YJ (2013) Discounted stochastic games with no stationary Nash equilibrium: two examples. Econometrica 81:1973–2007
Levy YJ, McLennan A (2015) Corrigendum to: discounted stochastic games with no stationary Nash equilibrium: two examples. Econometrica 83:1237–1252
Maitra A, Sudderth W (2003) Borel stay-in-a-set games. Int J Game Theory 32:97–108
Maitra A, Sudderth WD (2007) Subgame-perfect equilibria for stochastic games. Math Oper Res 32:711–722
Majumdar MK, Sundaram R (1991) Symmetric stochastic games of resource extraction. The existence of non-randomized stationary equilibrium. In: Raghavan et al (eds) Stochastic games and related topics. Kluwer, Dordrecht, pp 175–190

Maliar, L, Maliar, S (2016) Ruling out multiplicity of smooth equilibria in dynamic games: a hyperbolic discounting example. Dyn Games Appl 6:243–261
Mas-Colell A (1974) A note on a theorem of F. Browder. Math Program 6:229–233
Mashiah-Yaakovi A (2009) Periodic stopping games. Int J Game Theory 38:169–181
Mashiah-Yaakovi A (2014) Subgame-perfect equilibria in stopping games. Int J Game Theory 43:89–135
Mashiah-Yaakovi A (2015) Correlated equilibria in stochastic games with Borel measurable payoffs. Dyn Games Appl 5:120–135
Maskin E, Tirole J (2001) Markov perfect equilibrium: I. Observable actions. J Econ Theory 100:191–219
Mertens JF (2002) Stochastic games. In: Aumann RJ, Hart S (eds) Handbook of game theory with economic applications, vol 3. North Holland, Amsterdam/London, pp 1809–1832
Mertens JF (2003) A measurable measurable choice theorem. In: Neyman A, Sorin S (eds) Stochastic games and applications. Kluwer, Dordrecht, pp 107–130
Mertens JF, Neyman A (1981) Stochastic games. Int J Game Theory 10:53–56
Mertens JF, Parthasarathy T (1991) Nonzero-sum stochastic games. In: Raghavan et al (eds) Stochastic games and related topics. Kluwer, Dordrecht, pp 145–148
Mertens JF, Parthasarathy T (2003) Equilibria for discounted stochastic games. In: Neyman A, Sorin S (eds) Stochastic games and applications. Kluwer, Dordrecht, pp 131–172
Mertens JF, Sorin S, Zamir S (2015) Repeated games. Cambridge University Press, Cambridge
Milgrom P, Roberts J (1990) Rationalizability, learning and equilibrium in games with strategic complementarities. Econometrica 58:1255–1277
Milgrom P, Shannon C (1994) Monotone comparative statics. Econometrica 62:157–180
Mohan SR, Neogy SK, Parthasarathy T (1997) Linear complementarity and discounted polystochastic game when one player controls transitions. In: Ferris MC, Pang JS (eds) Proceedings of the international conference on complementarity problems. SIAM, Philadelphia, pp 284–294
Mohan SR, Neogy SK, Parthasarathy T (2001) Pivoting algorithms for some classes of stochastic games: a survey. Int Game Theory Rev 3:253–281
Monderer D, Shapley LS (1996) Potential games. Games Econ Behavior 14:124–143
Montrucchio L (1987) Lipschitz continuous policy functions for strongly concave optimization problems. J Math Econ 16:259–273
Morimoto H (1986) Nonzero-sum discrete parameter stochastic games with stopping times. Probab Theory Relat Fields 72:155–160
Nash JF (1950) Equilibrium points in n-person games. Proc Nat Acad Sci USA 36:48–49
Neveu J (1975) Discrete-parameter martingales. North-Holland, Amsterdam
Neyman A (2017) Continuous-time stochastic games. Games Econ Behavior 104:92–130
Neyman A, Sorin S (eds) (2003) Stochastic games and applications. Kluwer, Dordrecht
Nowak AS (1985) Existence of equilibrium stationary strategies in discounted noncooperative stochastic games with uncountable state space. J Optim Theory Appl 45:591–602
Nowak AS (1987) Nonrandomized strategy equilibria in noncooperative stochastic games with additive transition and reward structure. J Optim Theory Appl 52:429–441
Nowak AS (2003a) N-person stochastic games: extensions of the finite state space case and correlation. In: Neyman A, Sorin S (eds) Stochastic games and applications. Kluwer, Dordrecht, pp 93–106
Nowak AS (2003b) On a new class of nonzero-sum discounted stochastic games having stationary Nash equilibrium point. Int J Game Theory 32:121–132
Nowak AS (2006a) On perfect equilibria in stochastic models of growth with intergenerational altruism. Econ Theory 28:73–83
Nowak AS (2006b) A multigenerational dynamic game of resource extraction. Math Social Sci 51:327–336
Nowak AS (2006c) A note on equilibrium in the great fish war game. Econ Bull 17(2):1–10
Nowak AS (2007) On stochastic games in economics. Math Methods Oper Res 66:513–530
Nowak AS (2008) Equilibrium in a dynamic game of capital accumulation with the overtaking criterion. Econ Lett 99:233–237

Nowak AS (2010) On a noncooperative stochastic game played by internally cooperating generations. J Optim Theory Appl 144:88–106
Nowak AS, Altman E (2002) ε-Equilibria for stochastic games with uncountable state space and unbounded costs. SIAM J Control Optim 40:1821–1839
Nowak AS, Jaśkiewicz A (2005) Nonzero-sum semi-Markov games with the expected average payoffs. Math Methods Oper Res 62:23–40
Nowak AS, Raghavan TES (1992) Existence of stationary correlated equilibria with symmetric information for discounted stochastic games. Math Oper Res 17:519–526
Nowak AS, Raghavan TES (1993) A finite step algorithm via a bimatrix game to a single controller non-zero sum stochastic game. Math Program Ser A 59:249–259
Nowak AS, Szajowski K (1999) Nonzero-sum stochastic games. In: Ann Int Soc Dyn Games, vol 4. Birkhäuser, Boston, pp 297–343
Nowak AS, Więcek P (2007) On Nikaido-Isoda type theorems for discounted stochastic games. J Math Anal Appl 332:1109–1118
Ohtsubo Y (1987) A nonzero-sum extension of Dynkin's stopping problem. Math Oper Res 12:277–296
Ohtsubo Y (1991) On a discrete-time nonzero-sum Dynkin problem with monotonicity. J Appl Probab 28:466–472
Parthasarathy T (1973) Discounted, positive, and noncooperative stochastic games. Int J Game Theory 2:25–37
Parthasarathy T, Sinha S (1989) Existence of stationary equilibrium strategies in nonzero-sum discounted stochastic games with uncountable state space and state-independent transitions. Int J Game Theory 18:189–194
Peleg B, Yaari M (1973) On the existence of consistent course of action when tastes are changing. Rev Econ Stud 40:391–401
Pęski M, Wiseman T (2015) A folk theorem for stochastic games with infrequent state changes. Theoret Econ 10:131–173
Phelps E, Pollak R (1968) On second best national savings and game equilibrium growth. Rev Econ Stud 35:195–199
Pollak R (1968) Consistent planning. Rev Econ Stud 35:201–208
Potters JAM, Raghavan TES, Tijs SH (2009) Pure equilibrium strategies for stochastic games via potential functions. In: Advances in dynamic games and their applications. Annals of the international society of dynamic games, vol 10. Birkhäuser, Boston, pp 433–444
Purves RA, Sudderth WD (2011) Perfect information games with upper semicontinuous payoffs. Math Oper Res 36:468–473
Puterman ML (1994) Markov decision processes: discrete stochastic dynamic programming. Wiley, Hoboken
Raghavan TES, Syed Z (2002) Computing stationary Nash equilibria of undiscounted single-controller stochastic games. Math Oper Res 27:384–400
Raghavan TES, Tijs SH, Vrieze OJ (1985) On stochastic games with additive reward and transition structure. J Optim Theory Appl 47:451–464
Ramsey FP (1928) A mathematical theory of savings. Econ J 38:543–559
Ray D (1987) Nonpaternalistic intergenerational altruism. J Econ Theory 40:112–132
Reny PJ, Robson A (2002) Existence of subgame-perfect equilibrium with public randomization: a short proof. Econ Bull 3(24):1–8
Rieder U (1979) Equilibrium plans for non-zero sum Markov games. In: Moeschlin O, Pallaschke D (eds) Game theory and related topics. North-Holland, Amsterdam, pp 91–102
Rogers PD (1969) Non-zero-sum stochastic games. Ph.D. dissertation, report 69–8, Univ of California
Rosen JB (1965) Existence and uniqueness of equilibrium points for concave n-person games. Econometrica 33:520–534
Rosenberg D, Solan E, Vieille N (2001) Stopping games with randomized strategies. Probab Theory Relat Fields 119:433–451

Rubinstein A (1979) Equilibrium in supergames with the overtaking criterion. J Econ Theory 21:1–9
Secchi P, Sudderth WD (2002a) Stay-in-a-set games. Int J Game Theory 30:479–490
Secchi P, Sudderth WD (2002b) N-person stochastic games with upper semi-continuous payoffs. Int J Game Theory 30:491–502
Secchi P, Sudderth.WD (2005) A simple two-person stochastic game with money. In: Annals of the international society of dynamic games, vol 7. Birkhäuser, Boston, pp 39–66
Selten R (1975) Re-examination of the perfectness concept for equilibrium points in extensive games. Int J Game Theory 4:25–55
Shapley LS (1953) Stochastic games. Proc Nat Acad Sci USA 39:1095–1100
Shubik M, Whitt W (1973) Fiat money in an economy with one non-durable good and no credit: a non-cooperative sequential game. In: Blaquière A (ed) Topics in differential games. North-Holland, Amsterdam, pp 401–448
Shmaya E, Solan E (2004) Two player non-zero sum stopping games in discrete time. Ann Probab 32:2733–2764
Shmaya E, Solan E, Vieille N (2003) An applications of Ramsey theorem to stopping games. Games Econ Behavior 42:300–306
Simon RS (2007) The structure of nonzero-sum stochastic games. Adv Appl Math 38:1–26
Simon R (2012) A topological approach to quitting games. Math Oper Res 37:180–195
Simon RS (2016) The challenge of nonzero-sum stochastic games. Int J Game Theory 45:191–204
Sleet C, Yeltekin S (2016) On the computation of value correspondences for dynamic games. Dyn Games Appl 6:174–186
Smale S (1976) A convergent process of price adjustment and global Newton methods. J Math Econ 3:107–120
Sobel MJ (1871) Non-cooperative stochastic games. Ann Math Stat 42:1930–1935
Solan E (1998) Discounted stochastic games. Math Oper Res 23:1010–1021
Solan E (1999) Three-person absorbing games. Math Oper Res 24:669–698
Solan E (2001) Characterization of correlated equilibria in stochastic games. Int J Game Theory 30:259–277
Solan E (2017, in press) Acceptable strategy profiles in stochastic games. Games Econ Behavior
Solan E, Vieille N (2001) Quitting games. Math Oper Res 26:265–285
Solan E, Vieille N (2002) Correlated equilibrium in stochastic games. Games Econ Behavior 38:362–399
Solan E, Vieille N (2003) Deterministic multi-player Dynkin games. J Math Econ 39:911–929
Solan E, Vieille N (2010) Computing uniformly optimal strategies in two-player stochastic games. Econ Theory 42:237–253
Solan E, Ziliotto B (2016) Stochastic games with signals. In: Annals of the international society of dynamic games, vol 14. Birkhäuser, Boston, pp 77–94
Sorin S (1986) Asymptotic properties of a nonzero-sum stochastic games. Int J Game Theory 15:101–107
Spence M (1976) Product selection, fixed costs, and monopolistic competition. Rev Econ Stud 43:217–235
Stachurski J (2009) Economic dynamics: theory and computation. MIT Press, Cambridge, MA
Stokey NL, Lucas RE, Prescott E (1989) Recursive methods in economic dynamics. Harvard University Press, Cambridge
Strotz RH (1956) Myopia and inconsistency in dynamic utility maximization. Rev Econ Stud 23:165–180
Sundaram RK (1989a) Perfect equilibrium in a class of symmetric dynamic games. J Econ Theory 47:153–177
Sundaram RK (1989b) Perfect equilibrium in a class of symmetric dynamic games. Corrigendum. J Econ Theory 49:385–187
Szajowski K (1994) Markov stopping game with random priority. Z Oper Res 39:69–84
Szajowski K (1995) Optimal stopping of a discrete Markov process by two decision makers. SIAM J Control Optim 33:1392–1410

Takahashi M (1964) Equilibrium points of stochastic non-cooperative n-person games. J Sci Hiroshima Univ Ser A-I Math 28:95–99
Thuijsman F, Raghavan TES (1997) Perfect information stochastic games and related classes. Int J Game Theory 26:403–408
Topkis D (1978) Minimizing a submodular function on a lattice. Oper Res 26:305–321
Topkis D (1998) Supermodularity and complementarity. Princeton University Press, Princeton
Valadier M (1994) Young measures, weak and strong convergence and the Visintin-Balder theorem. Set-Valued Anal 2:357–367
Van Long N (2011) Dynamic games in the economics of natural resources: a survey. Dyn Games Appl 1:115–148
Vieille N (2000a) Two-player stochastic games I: a reduction. Isr J Math 119:55–92
Vieille N (2000b) Two-player stochastic games II: the case of recursive games. Israel J Math 119:93–126
Vieille N (2002) Stochastic games: Recent results. In: Aumann RJ, Hart S (eds) Handbook of game theory with economic applications, vol 3. North Holland, Amsterdam/London, pp 1833–1850
Vives X (1990) Nash equilibrium with strategic complementarities. J Math Econ 19:305–321
von Weizsäcker CC (1965) Existence of optimal programs of accumulation for an infinite horizon. Rev Econ Stud 32:85–104
Vrieze OJ, Thuijsman F (1989) On equilibria in repeated games with absorbing states. Int J Game Theory 18:293–310
Whitt W (1980) Representation and approximation of noncooperative sequential games. SIAM J Control Optim 18:33–48
Więcek P (2009) Pure equilibria in a simple dynamic model of strategic market game. Math Methods Oper Res 69:59–79
Więcek P (2012) N-person dynamic strategic market games. Appl Math Optim 65:147–173
Yasuda M (1985) On a randomised strategy in Neveu's stopping problem. Stoch Process Appl 21:159–166

Mean Field Games

7

Peter E. Caines, Minyi Huang, and Roland P. Malhamé

Contents

P. E. Caines (✉)
Department of Electrical and Computer Engineering, McGill University, Montreal, QC, Canada
e-mail: peterc@cim.mcgill.ca

M. Huang
School of Mathematics and Statistics, Carleton University, Ottawa, ON, Canada
e-mail: mhuang@math.carleton.ca

R. P. Malhamé
Department of Electrical Engineering, École Polytechnique de Montréal, Montreal, QC, Canada
e-mail: roland.malhame@polymtl.ca

T. Başar, G. Zaccour (eds.), *Handbook of Dynamic Game Theory*,
https://doi.org/10.1007/978-3-319-44374-4_7

Abstract

Mean field game (MFG) theory studies the existence of Nash equilibria, together with the individual strategies which generate them, in games involving a large number of asymptotically negligible agents modeled by controlled stochastic dynamical systems. This is achieved by exploiting the relationship between the finite and corresponding infinite limit population problems. The solution to the infinite population problem is given by (i) the Hamilton-Jacobi-Bellman (HJB) equation of optimal control for a generic agent and (ii) the Fokker-Planck-Kolmogorov (FPK) equation for that agent, where these equations are linked by the probability distribution of the state of the generic agent, otherwise known as the system's mean field. Moreover, (i) and (ii) have an equivalent expression in terms of the stochastic maximum principle together with a McKean-Vlasov stochastic differential equation, and yet a third characterization is in terms of the so-called master equation. The chapter first describes problem areas which motivate the development of MFG theory and then presents the theory's basic mathematical formalization. The main results of MFG theory are then presented, namely the existence and uniqueness of infinite population Nash equilibiria, their approximating finite population ε-Nash equilibria, and the associated best response strategies. This is followed by a presentation of the three main mathematical methodologies for the derivation of the principal results of the theory. Next, the particular topics of major-minor agent MFG theory and the common noise problem are briefly described and then the final section concisely presents three application areas of MFG theory.

Keywords

Mean field games · MFG · MFG equilibria · Mean field game equations · MFG PDEs · MFG analytic methods · MFG probabilistic methods · Major minor agent and common noise games · Non-linear Markov processes · MFG applications

1 Introduction

1.1 The Fundamental Idea of Mean Field Game Theory

Mean field game (MFG) theory studies the existence of Nash equilibria, together with the individual strategies which generate them, in games involving a large number of asymptotically negligible agents modeled by controlled stochastic dynamical systems. This is achieved by exploiting the relationship between the finite and corresponding infinite limit population problems. The solution to the infinite population problem is given by (i) the Hamilton-Jacobi-Bellman (HJB) equation of optimal control for a generic agent and (ii) the Fokker-Planck-Kolmogorov (FPK) equation for that agent, where these equations are linked by the distribution of the state of the generic agent, otherwise known as the system's mean field. Moreover, (i) and (ii) have an equivalent expression in terms of the stochastic maximum principle together with a McKean-Vlasov stochastic differential equation, and yet a third characterization is in terms of the so-called master equation. An important feature of MFG solutions is that they have fixed-point properties regarding the individual responses to and the formation of the mean field which conceptually correspond to equilibrium solutions of the associated games.

1.2 Background

Large population dynamical multi-agent noncooperative and cooperative phenomena occur in a wide range of designed and natural settings such as communication, environmental, epidemiological, transportation, and energy systems, and they underlie much economic and financial behavior. Here, large is taken to mean numerically large with respect to some implied normal range or infinite (as a discrete or uncountable set). Analysis of such systems with even a moderate number of agents is regarded as being extremely difficult using the finite population game theoretic methods which were developed over several decades for multi-agent control systems (see, e.g., Basar and Ho 1974; Ho 1980; Basar and Olsder 1999; and Bensoussan and Frehse 1984). In contrast to the dynamical system formulation of multi-agent games, the continuum population game theoretic models of economics (Aumann and Shapley 1974; Neyman 2002) are static, as, in general, are the large population models employed in network games (Altman et al. 2002) and classical transportation analysis (Correa and Stier-Moses 2010; Haurie 1985; Wardrop 1952). However, dynamical (also termed sequential) stochastic games were analyzed in the continuum limit in the work of Jovanovic and Rosenthal (1988) and Bergin and Bernhardt (1992), where a form of the mean field equations can be recognized in a discrete-time dynamic programming equation linked with an evolution equation for the population state distribution.

Subsequently, what is now called MFG theory originated in the equations for dynamical games with (i) large finite populations of asymptotically negligible agents together with (ii) their infinite limits, in the work of (Huang et al. 2003, 2007), Huang et al. (2006) (where the framework was called the Nash certainty equivalence principle; see Caines (2014)) and independently in that of Lasry and Lions (2006a,b, 2007), where the now standard terminology of mean field games (MFGs) was introduced. The closely related notion of oblivious equilibria for large population dynamic games was also independently introduced by Weintraub et al. (2005, 2008) within the framework of discrete-time Markov decision processes (MDP).

1.3 Scope

The theory and methodology of MFG has rapidly developed since its inception and is still advancing. Consequently, the objective of this article is only to present the fundamental conceptual framework of MFG in the continuous time setting and the main techniques that are currently available. Moreover, the important topic of numerical methods will not be included, but it is addressed elsewhere in this volume by other contributors.

2 Problem Areas and Motivating Examples

Topics which motivate MFG theory or form potential areas of applications include the following:

2.1 Engineering

In the domain of power grid network control, an MFG methodology is being applied to create decentralized schemes for power network peak load reduction and compensation of fluctuations originating in renewable sources (see Sect. 7). Vast numbers of individual electric water-heating devices are planned to be coordinated in a decentralized way using an MFG architecture which would limit the required flows of information, such that individual controls give rise to a desired mean consumption.

For cell phone communication networks where coded signals can overlap in the frequency spectrum (called CDMA networks), a degradation of individual reception can occur when multiple users emit in the same frequency band. Compensation for this by users increasing their individual signal powers will shorten battery life and is collectively self-defeating. However, in the resulting dynamic game, a Nash equilibrium is generated when each cellular user controls its transmitted power as specified by MFG theory (see Sect. 7). Other applications include decentralized charging control of large populations of plug-in electric vehicles (Ma et al. 2013).

2.2 Economics and Finance

Human capital growth has been considered in an MFG setting by Guéant et al. (2011) and Lucas and Moll (2014) where the individuals invest resources (such as time and money) for the improvement of personal skills to better position themselves in the labor market when competing with each other.

Chan and Sircar (2015) considered the mean field generalization of Bertrand and Cournot games in the production of exhaustible resources where the price acts as a medium for the producers to interact. Furthermore, an MFG formulation has been used by Carmona et al. (2015) to address systemic risk as characterized by a large number of banks having reached a default threshold by a given time, where interbank loaning and lending is regarded as an instrument of control.

2.3 Social Phenomena

Closely related to the application of MFG theory to economics and finance is its potential application to a whole range of problems in social dynamics. As a short list of current examples, we mention:

2.3.1 Opinion Dynamics

The evolution of the density of the opinions of a mass of agents under hypotheses on the dynamics and stubbornness of the agents is analyzed in an MFG framework in Bauso et al. (2016).

2.3.2 Vaccination Games

When the cost to each individual is represented as a function of (a) the risk of side effects, (b) the benefits of being vaccinated, and (c) the proportion of the population which is vaccinated, as in Bauch and Earn (2004), it is evident that an MFG formulation is relevant, and this has been pursued in the work of Laguzet and Turinici (2015).

2.3.3 Congestion Studies

MFG methodology has been employed in the study of crowds and congested flows in Dogbé (2010) and Lachapelle and Wolfram (2011), where numerical methods reveal the possibility of lane formation.

3 Mathematical Framework

3.1 Agent Dynamics

In MFG theory individual agents are modeled by controlled stochastic systems which may be coupled by their dynamics, their cost functions, and their observation functions.

The principal classes of dynamical models which are used in MFG theory are sketched below; in all of them, the individual agent controls its own state process (invariably denoted here by x_i or x_θ) and is subject to individual and possibly common stochastic disturbances.

Concerning terminology, throughout this article, the term *strategy* of an agent means the functional mapping from an agent's information set to its control actions over time, in other words, the *control law* of that agent.

3.1.1 Diffusion Models

In the diffusion-based models of large population games, the state evolution of a collection of N agents $\mathscr{A}_i, 1 \leq i \leq N < \infty$, is specified by a set of N controlled stochastic differential equations (SDEs) which in the important linear case take the form:

$$dx_i = (A_i x_i + B_i u_i)dt + C_i dw_i, \quad 1 \le i \le N, \tag{7.1}$$

on a finite or infinite time interval, where for the ith agent $\mathscr{A}_i$, $x_i \in \mathbb{R}^n$ is the state, $u_i \in \mathbb{R}^m$ the control input, and $w_i \in \mathbb{R}^r$ a standard Wiener process and where $\{w_i, 1 \le i \le N\}$ are independent processes. For simplicity, all collections of system initial conditions are taken to be independent and have finite second moment.

A simplified form of the general case is given by the following set of controlled SDEs which for each agent $\mathscr{A}_i$ includes state coupling with all other agents:

$$dx_i(t) = \frac{1}{N}\sum_{j=1}^{N} f(t, x_i(t), u_i(t), x_j(t))dt + \sigma dw_i(t) \tag{7.2}$$

$$= \int_{R^n} f(t, x_i(t), u_i(t), z)\left\{\frac{1}{N}\sum_{j=1}^{N}\delta_{x_j}(dz)\right\} dt + \sigma dw_i(t)$$

$$=: f\left[t, x_i(t), u_i(t), \left\{\frac{1}{N}\sum_{j=1}^{N}\delta_{x_j}\right\}\right] dt + \sigma dw_i(t) \tag{7.3}$$

$$= f[t, x_i(t), u_i(t), \mu_t^N]dt + \sigma dw_i(t), \tag{7.4}$$

where the function $f[\cdot,\cdot,\cdot,\cdot]$, with the empirical measure of the population states $\mu_t^N := \frac{1}{N}\sum_{j=1}^{N}\delta_{x_j}$ at the instant t as its fourth argument, is defined via

$$f[t, x(t), u(t), \nu_t] := \int_{R^n} f(t, x(t), u(t), z)\nu_t(dz), \tag{7.5}$$

for any measure flow ν_t, as in Cardaliaguet (2012) and Kolokoltsov et al. (2012). For simplicity, we do not consider diffusion coefficients depending on the system state or control.

Equation (7.2) is defined on a finite or infinite time interval, where, here, for the sake of simplicity, only the uniform (i.e., nonparameterized) generic agent case is presented. The dynamics of a generic agent in the infinite population limit of this system is then described by the following controlled McKean-Vlasov equation

$$dx_t = f[x_t, u_t, \mu_t]dt + \sigma dw_t, \quad 1 \le i \le N, \ \ 0 \le t \le T,$$

where $f[x, u, \mu_t] = \int_{\mathbb{R}} f(x, u, y)\mu_t(dy)$, $\mu_t(\cdot)$ denotes the distribution of the state of the generic agent at $t \in [0, T]$ and the initial condition measure μ_0 is specified. (The dynamics used in Lasry and Lions (2006a,b, 2007) and Cardaliaguet (2012) are of the form $dx_i(t) = u_i(t)dt + dw_i(t)$, where u_i, x_i, w_i are scalar-valued processes.)

It is reasonable to speculate that results described below for the case of system dynamics driven by a Wiener process would hold in the general case of a Wiener process plus a point process and ultimately to the general case of Lévy processes; indeed, in an operator framework, this generalization is carried out in the work of Kolokoltsov et al. (see below).

3.1.2 Nonlinear Markov Processes

The mean field game dynamic modeling framework has been significantly generalized by Kolokoltsov et al. (2012) via the introduction of controlled nonlinear Markov processes where, in this framework, instead of diffusion SDEs, the evolution of a typical agent is described by an integrodifferential generator of Lévy-Khintchine type, where, as in the diffusion models described in the rest of this paper, the coefficients of the dynamical system of each agent, and its associated costs, are permitted to depend upon the empirical measure of the population of agents. As a consequence, by virtue of the Markov property, game theoretic best response problems in this framework can still be solved within the HJB formalism, and moreover the sensitivity analysis of the controls and dynamics with respect to perturbations in the population measure flow is facilitated.

3.1.3 Markov Chains and Other Discrete-Time Processes

The dynamical evolution of the state x_i of the ith agent $\mathscr{A}_i$ is formulated as a discrete-time Markov decision process (MDP). The so-called anonymous sequential games (Bergin and Bernhardt 1992; Jovanovic and Rosenthal 1988) deal with a continuum of agents, where a generic agent's cost function depends on its own state and action, and the joint state-action distribution of the agent population.

In the context of industry dynamics, Weintraub et al. (2005, 2008) adopted a large finite population, where the dynamics may be described by a Markov transition kernel model $P_{t+1} := P(x_i(t+1)|x_i(t), x_{-i}(t), u_i(t))$, where x_{-i} denotes the states of other players; also see Adlakha et al. (2015).

3.1.4 Finite State Models

Within continuous time modeling, Gomes et al. (2013) formulated a mean field game of switching among finite states and determined the equilibrium by a coupled system of ordinary differential equations. Finite state mean field games have applications in social-economic settings and networks (Gomes et al. 2014; Kolokoltsov and Malafeyev 2017; Kolokoltsov and Bensoussan 2016).

3.2 Agent Cost Functions

Throughout this article we shall only refer to cost functions which are the additive (or integral) composition over a finite or infinite time interval of instantaneous (running) costs; in MFG theory these will depend upon the individual state of an agent along with its control and possibly a function of the states of all other agents in the system. As usual in stochastic decision problems, the cost function for any agent will be defined by the expectation of the integrated running costs over all possible sample paths of the system. An important class of such functions is the so-called ergodic cost functions which are defined as the time average of integral cost functions.

3.2.1 Individual Agent Performance Functions in Noncooperative Games

The principal types of games considered in MFG theory are, first, noncooperative games, where each agent seeks to minimize its own loss represented by its cost function. In the most basic finite population linear-quadratic diffusion case, the agent $\mathscr{A}_i$, $1 \leq i \leq N$, possesses a cost function of the form:

$$J_i^N(u_i, u_{-i}) = E\int_0^T \{\|x_i(t) - Hm_N(t)\|_Q^2 + \|u_i(t)\|_R^2\}dt, \tag{7.6}$$

where $\|\cdot\|_M^2$ denotes the squared (semi-)norm arising from the positive semi-definite matrix M, where we assume the cost-coupling term to be of the form $m_N(t) := \overline{x}_N(t) + \eta$, $\eta \in \mathbb{R}^n$, where u_{-i} denotes all agents' control laws except for that of the ith agent, $\overline{x}_N$ denotes the population average state $(1/N)\sum_{i=1}^N x_i$, and where, here and below, the expectation is taken over an underlying sample space which carries all initial conditions and Wiener processes.

For the nonlinear case introduced in Sect. 3.1.1, a corresponding finite population mean field cost function is

$$J_i^N(u_i, u_{-i}) := E\int_0^T \Big((1/N)\sum_{j=1}^N L(x_i(t), u_i(t), x_j(t))\Big)dt, \quad 1 \leq i \leq N, \tag{7.7}$$

where $L(\cdot)$ is the pairwise cost rate function. Setting the infinite population cost rate $L[x, u, \mu_t] = \int_{\mathbb{R}} L(x, u, y)\mu_t(dy)$, hence the corresponding infinite population expected cost for a generic agent $\mathscr{A}_i$ is given by

$$J_i(u_i, \mu) \;:=\; E\int_0^T L[x(t), u_i(t), \mu_t]dt, \tag{7.8}$$

which is the general expression appearing in Huang et al. (2006) and Nourian and Caines (2013) and which includes those of Lasry and Lions (2006a,b, 2007), Cardaliaguet (2012). $e^{-\rho t}$ discounted costs are employed for infinite time horizon cost functions (Huang et al. 2003, 2007), while the long-run average cost is used for ergodic MFG problems (Bardi 2012; Lasry and Lions 2006a,b, 2007; Li and Zhang 2008).

3.2.2 Risk-Sensitive Performance Functions

This article will solely focus on additive type costs although other forms can be adopted for the individual agents. One important such form is a risk-sensitive cost function:

$$J_i^N(u_i, u_{-i}) = E \exp\left[\int_0^T (1/N) \sum_{j=1}^N L(x_i(t), u_i(t), x_j(t)) dt\right],$$

which allows the use of dynamic programming to compute the best response. For related analyses in the linear-exponential-quadratic-Gaussian (LEQG) case, see, e.g., Tembine et al. (2014).

3.2.3 Performance Functions in Major-Minor Agent Systems

We start with the most basic finite population linear-quadratic case with a major agent $\mathscr{A}_0$ having state x_0 and N minor agents $\mathscr{A}_i$, $1 \le i \le N$, with states x_i. The SDEs of $\mathscr{A}_0$ and $\mathscr{A}_i$ are given by

$$\begin{aligned} dx_0 &= (A_0 x_0 + B_0 u_0 + F_0 m_N) dt + D_0 dw_0, \\ dx_i &= (A x_i + B u_i + F m_N + G x_0) dt + D dw_i, \quad 1 \le i \le N, \end{aligned}$$

where $m_N = \frac{1}{N} \sum_{i=1}^N x_i$ and the initial states are $x_0(0)$ and $x_i(0)$. The major agent has a cost function of the form:

$$J_0^N(u_0, u_{-0}) = E \int_0^T \{\|x_0(t) - H_0 m_N(t)\|_{Q_0}^2 + \|u_i(t)\|_{R_0}^2\} dt,$$

and the minor agent $\mathscr{A}_i$ possesses a cost function of the form:

$$J_i^N(u_i, u_{-i}) = E \int_0^T \{\|x_i(t) - H_1 m_N(t) - H_2 x_0(t)\|_Q^2 + \|u_i(t)\|_R^2\} dt.$$

Correspondingly, in the nonlinear case with a major agent, the N nonlinear equations in (7.2) are generalized to include the state of a major agent described by an additional SDE, giving a system described by $N + 1$ equations. The cost functions are given by

$$J_0^N(u_0, u_{-0}) := E \int_0^T (1/N) \sum_{j=1}^N L_0(x_0(t), u_0(t), x_j(t)) dt,$$

and

$$J_i^N(u_i, u_{-i}) := E \int_0^T (1/N) \sum_{j=1}^N L(x_i(t), u_i(t), x_0(t), x_j(t)) dt.$$

Consequently, the infinite population mean field cost functions for the major and minor agents respectively are given by

$$J_0(u_0,\mu) \;:=\; E\int_0^T L_0[x_0(t),u_0(t),\mu_t]dt,$$

and

$$J_i(u_i,\mu) \;:=\; E\int_0^T L[x_i(t),u_i(t),x_0(t),\mu_t]dt,$$

where $L_0[x_0(t), u_0(t), \mu_t]$ and $L[x_i(t), u_i(t), x_0(t), \mu_t]$ correspond to their finite population versions as in the basic minor agent only case.

3.3 Information Patterns

We now introduce the following definitions and characterizations of information patterns in dynamic game theory which shall be used in the rest of this article.

The States of a Set of Agents: A state in dynamic games is taken to be either (i) an individual (agent) state as defined in the Sect. 3.1, in which case it will constitute a component of the global system state, namely, the union of the individual states, or (ii) the global state, which is necessarily sufficient to describe the dynamical evolution of all the agents once the system inputs are specified. We emphasize that in this setting (see, for instance, (7.2)), only knowledge of the entire system state (i.e., the union of all the individual states) plus all the system inputs would in general permit such an extrapolation.

Moreover, in the infinite population case, the (global) system state may refer to the statistical or probability distribution of the population of individual states, i.e., the *mean field*.

Variety of Information Patterns: Information on dynamical states: For any given agent, this may constitute (i) the initial state, (ii) the partial past history, or (iii) the purely current state values of either (i) that individual agent or (ii) a partial set of all the agents or (iii) the entire set of the agents.

Open-Loop and Closed-Loop Control Laws: The common definition of an open-loop control law for an agent is that it is solely a function of the information set consisting of time and the initial state of that agent or of the whole system (i.e., the global initial state). A closed-loop (i.e., feedback) control law is one which is a function of time and the current state of that agent or the global state of the system subject to the given information pattern constraints, where a particular case of importance is that in which an agent's strategy at any instant depends only upon its current state.

A significant modification of the assertion above must be made in the classical mean field game situation with no common noise or correlating major agent; indeed, in that case all agents in the population will be employing an infinite population-based Nash equilibrium strategy. As a result, the probability distribution of the generic agent, which can be identified with the global state as defined earlier, becomes deterministically predictable for all future times, provided it is known at the initial time. Consequently, for such MFGs, an adequate characterization of sufficient information is the initial global state. Furthermore, in the MFG framework among others, an agent lacking complete observations on its own current state may employ recursive filtering theory to estimate its own state.

Statistical Information on the Population: For any individual agent, information is typically available on that agent's own dynamics (i.e., the structure and the parameters of its own controlled dynamic system); but it is a distinct assumption that no such information is available to it concerning other individual agents. Furthermore, this information may be available in terms of a distribution (probabilistic or otherwise) over the population of agents and not associated with any identifiable individual agent. This is particularly the case in the MFG framework.

Who Knows What About Whom: A vital aspect of information patterns in game theory is that knowledge concerning (i) other agents' control actions, or, more generally, concerning (ii) their strategies (i.e., their control laws), may or may not be available to any given agent. This is a fundamental issue since the specification of an agent's information pattern in terms of knowledge of other agent's states, system dynamics, cost functions, and parameters leads to different possible methods to solve for different types of equilibrium strategies and even for their existence.

In the MFG setting, if the common assumption is adopted that all agents will compute their best response in reaction to the best responses of all other agents (through the system dynamics), it is then optimal for each agent to solve for its strategy through the solution of the MFG equations. The result is that each agent will know the control strategy of every other agent, but not its control action since the individual state of any other agent is not available to a given agent. Note however that the state distribution of any (i.e., other generic) agent is known by any given generic agent since this is the system's mean field which is generated by the MFG equations.

3.4 Solution Concepts: Equilibria and Optima

In contrast to the situation in classical stochastic control, in the game theoretic context, the notion of an optimal level of performance and associated optimal control for the entire system is in general not meaningful. The fundamental solution concept is that of an equilibrium and here we principally consider the notion of a Nash equilibrium.

3.4.1 Equilibria in Noncooperative Games: Nash Equilibria

For a set of agents $\mathscr{A}_i, 1 \leq i \leq N < \infty$, let $\mathscr{U}^N := \mathscr{U}_1 \times \ldots \times \mathscr{U}_N$ denote the joint admissible strategy space, where each space $\mathscr{U}_i$ consists of a set of strategies (i.e., control laws) u_i which are functions of information specified for $\mathscr{A}_i$ via the underlying information pattern.

The joint strategy $u \triangleq (u_1, \ldots, u_N)$ (sometimes written as $\{u_i, 1 \leq i \leq N\}$) lying in $\mathscr{U}_1 \times \ldots \times \mathscr{U}_N$ will constitute an input for a specific system in one of the classes specified in Sects. 3.1 and 3.2.

The joint strategy (or control law) $u^{\circ,N} \triangleq \{u_i^\circ, 1 \leq i \leq N\} \in \mathscr{U}^N$ is said to generate an *ε-Nash equilibrium*, $\varepsilon \geq 0$, if for each i,

$$J_i^N(u_i^\circ, u_{-i}^\circ) - \varepsilon \leq \inf_{u_i \in \mathscr{U}_i} J_i^N(u_i, u_{-i}^\circ) \leq J_i^N(u_i^\circ, u_{-i}^\circ). \tag{7.9}$$

In case $\varepsilon = 0$, the equilibrium is called a *Nash equilibrium*.

This celebrated concept has the evident interpretation that when all agents except agent $\mathscr{A}_i$ employ a set of control laws $\{u_j^\circ, j \neq i, 1 \leq j \leq N\}$, any deviation by $\mathscr{A}_i$ from u_i° can yield a cost reduction of at most ε.

In the MFG framework, in its basic noncooperative formulation, the objective of each agent is to find strategies (i.e., control laws) which are compatible with respect to the information pattern and other dynamical constraints and which minimize its individual performance function. Consequently the resulting problem is necessarily game theoretic and the central results of the topic concern the existence of Nash Equilibria and their properties.

For a system of N players, under the hypothesis of closed-loop state information (see Sect. 3.3), we shall define the *set of value functions* $\{V_i(t,x), 1 \leq i \leq N,\}$ *in a Nash equilibrium*, as the set of costs of N agent $\mathscr{A}_i$, $1 \leq i \leq N$, with respect to the time and *global state* pair (t, x). The set of value functions and its existence may be characterized by a set of coupled HJB equations.

Under closed-loop information, the Nash equilibrium, if it exists, is *sub-game perfect* in the sense that by restricting to any remaining period of the original game, the set of strategies is still a Nash equilibrium for the resulting sub-game. In this case, the strategy of each agent is determined as a function of time and the current states of the agents and is usually called a *Markov strategy*.

3.4.2 Pareto Optima

A set of strategies yields a *Pareto optimum* if a change of strategies which strictly decreases the cost incurred by one agent strictly increases the cost incurred by at least one other agent.

3.4.3 Social Optima and Welfare Optimization

Within the framework of this article, a *social cost* or *(negative) welfare function* is defined as the sum of the individual cost functions of a set of agents (Huang et al. 2012). As a result a cooperative game may be defined which consists of the agents minimizing the social cost as a cooperative optimal control problem, where the individual strategies will depend upon the information pattern. We observe that a social optimum is necessarily a Pareto optimum with respect to the vector of individual costs since otherwise at least one summand of the social cost function may be strictly reduced without any other agent's cost increasing. The so-called *person-by-person optimality* is the property that at the social optimum, the strategy change of any single agent can yield no improvement of the social cost and so provides a useful necessary condition for social optimality. The exact solution of this problem in general requires a centralized information pattern.

3.4.4 Team Optima

Team problems are distinguished from cooperative game problems by the fact that only one cost function is defined a priori for the entire set of agents while they have access to different sets of information. A necessary condition for a solution to be team optimal is that the person-by-person optimality condition is satisfied (Ho 1980). Team problems do not in general reduce to single agent optimum problems due to the variety of information patterns that are possible for the set of agents.

3.4.5 Mean Field Type Control Optimality

Mean field type control deals with optimal control problems where the mean field of the state process either is involved in the cost functional in a nonlinear manner, such as being associated with the variance of the state, or appears in the system dynamics, or both, and is a function of the single agent's control. Unlike standard stochastic optimal control problems, mean field type control problems do not possess an iterated expectation structure due to the mean field term (i.e., there is time inconsistency), which excludes the direct (i.e., without state extension) application of dynamic programming. In this case, the stochastic maximum principle is an effective tool for characterizing the optimal control; see Andersson and Djehiche (2011) and the monograph of Bensoussan et al. (2013). Carmona et al. (2013) considered a closely related problem termed the control of McKean-Vlasov dynamics. Mean field games dealing with mean field type dynamics and costs and addressing time consistency are considered by Djehiche and Huang (2016).

4 Analytic Methods: Existence and Uniqueness of Equilibria

The objective of each agent in the classes of games under consideration is to find strategies which are admissible with respect to the given dynamic and information constraints and which achieve one of corresponding types of equilibria or optima described in the previous section. In this section we present some of the main analytic methods for establishing the existence, uniqueness, and the nature of the related control laws and their equilibria.

The fundamental feature of MFG theory is the relation between the game theoretic behavior (assumed here to be noncooperative) of finite populations of agents and the infinite population behavior characterized by a small number of equations in the form of PDEs or SDEs.

4.1 Linear-Quadratic Systems

The basic mean field problem in the linear-quadratic case has an explicit solution characterizing a Nash equilibrium (see Huang et al. 2003, 2007). Consider the scalar infinite time horizon discounted case, with nonuniform parameterized agents $\mathscr{A}_\theta$ (representing a generic agent $\mathscr{A}_i$ taking parameter θ) with parameter distribution $F(\theta), \theta \in \Theta$ and system parameters identified as $A_\theta = a_\theta, B_\theta = b_\theta, Q := 1$, $R = r > 0$, $H = \gamma$; the extension to the vector case and more general parameter dependence on θ is straightforward. The so-called Nash certainty equivalence (NCE) scheme generating the equilibrium solution takes the form:

$$\rho s_\theta = \frac{ds_\theta}{dt} + a_\theta s_\theta - \frac{b_\theta^2}{r}\Pi_\theta s_\theta - x^*, \tag{7.10}$$

$$\frac{d\overline{x}_\theta}{dt} = (a_\theta - \frac{b_\theta^2}{r}\Pi_\theta)\overline{x}_\theta - \frac{b_\theta^2}{r}s_\theta, \qquad 0 \le t < \infty, \tag{7.11}$$

$$\overline{x}(t) = \int_\Theta \overline{x}_\theta(t)dF(\theta), \tag{7.12}$$

$$x^*(t) = \gamma(\overline{x}(t) + \eta), \tag{7.13}$$

$$\rho\Pi_\theta = 2a_\theta\Pi_\theta - \frac{b_\theta^2}{r}\Pi_\theta^2 + 1, \quad \Pi_\theta > 0, \quad \text{Riccati Equation} \tag{7.14}$$

where the control law of the generic parameterized agent $\mathscr{A}_\theta$ has been substituted into the system equation (7.1) and is given by $u_\theta^0(t) = -\frac{b_\theta}{r}(\Pi_\theta x_\theta(t) + s_\theta(t))$, $0 \le t < \infty$. u_θ^0 is the optimal tracking feedback law with respect to $x^*(t)$ which is an affine function of the mean field term $\overline{x}(t)$, the average with respect to the parameter distribution F of the $\theta \in \Theta$ parameterized state means $x_\theta(t)$ of the agents. Subject to the conditions for the NCE scheme to have a solution, each agent is necessarily in a Nash equilibrium with respect to all full information causal (i.e., non-anticipative) feedback laws with respect to the remainder of agents when these are employing the law u_θ^0 associated with their own parameter.

It is an important feature of the best response control law u_θ^0 that its form depends only on the parametric distribution F of the entire set of agents, and at any instant it is a feedback function of only the state of the agent $\mathscr{A}_\theta$ itself and the deterministic mean field-dependent offset s_θ, and is thus decentralized.

4.2 Nonlinear Systems

For the general nonlinear case, the MFG equations on $[0, T]$ are given by the linked equations for (i) the value function V for each agent in the continuum, (ii) the FPK equation for the SDE for that agent, and (iii) the specification of the best response feedback law depending on the mean field measure μ_t and the agent's state $x(t)$. In the uniform agent scalar case, these take the following form:

The mean field game HJB-FPK equations are as follows:

$$[\text{HJB}] \quad -\frac{\partial V(t,x)}{\partial t} = \inf_{u\in U}\left\{ f[x,u,\mu_t]\frac{\partial V(t,x)}{\partial x} + L[x,u,\mu_t]\right\} + \frac{\sigma^2}{2}\frac{\partial^2 V(t,x)}{\partial x^2} \tag{7.15}$$

$$V(T,x) = 0,$$

$$[\text{FPK}] \quad \frac{\partial p_\mu(t,x)}{\partial t} = -\frac{\partial\{f[x,u^\circ(t,x),\mu_t]p_\mu(t,x)\}}{\partial x} + \frac{\sigma^2}{2}\frac{\partial^2 p_\mu(t,x)}{\partial x^2} \tag{7.16}$$

$$(t,x) \in [0,T] \times \mathbb{R}$$

$$p_\mu(0,x) = p_{\mu_0}(x),$$

$$[\text{BR}] \quad u^\circ(t,x) = \varphi(t,x|\mu_t), \tag{7.17}$$

where $p_\mu(t,\cdot)$ is the density of the measure μ_t, which is assumed to exist, and the function $\varphi(t,x|\mu_t)$ is the infimizer in the HJB equation. The (t,x,μ_t)-dependent feedback control gives an optimal control (also known as a best response (BR) strategy) for the generic individual agent with respect to the infinite population-dependent performance function (7.8) (where the infinite population is represented by the generic agent measure μ).

By the very definition of the solution to FPK equations, the solution μ above will be the state distribution in the process distribution solution pair (x,μ) in

$$[\text{SDE}] \quad dx_t = f[x_t,u_t^\circ,\mu_t]dt + \sigma dw_t, \quad 1\le i\le N, \quad 0\le t\le T. \tag{7.18}$$

This equivalence of the controlled sample path pair (x,μ) solution to the SDE and the corresponding FPK PDE is very important from the point of view of the existence, uniqueness, and game theoretic interpretation of the solution to the system's equation.

A solution to the mean field game equations above may be regarded as an equilibrium solution for an infinite population game in the sense that each BR feedback control (generated by the HJB equation) enters an FPK equation – and hence the corresponding SDE – and so generates a pair (x,μ), where each generic agent in the infinite population with state distribution μ solves the same optimization problem and hence regenerates μ.

In this subsection we briefly review the main methods which are currently available to establish the existence and uniqueness of solutions to various sets of MFG equations. In certain cases the methods are based upon iterative techniques which converge subject to various well-defined conditions. The key feature of the methods is that they yield individual state and mean field-dependent feedback control laws generating ε-Nash equilibria together with an upper bound on the approximation error.

The general nonlinear MFG problem is approached by different routes in the basic sets of papers Huang et al. (2007, 2006), Nourian and Caines (2013), Carmona and Delarue (2013) on one hand, and Lasry and Lions (2006a,b, 2007), Cardaliaguet (2012), Cardaliaguet et al. (2015), Fischer (2014), Carmona and Delarue (2014) on the other. Roughly speaking, the first set uses an infinite to finite population approach (to be called the top-down approach) where the infinite population game equations are first analyzed by fixed-point methods and then ε-Nash equilibrium results are obtained for finite populations by an approximation analysis, while the latter set analyzes the Nash equilibria of the finite population games, with each agent using only individual state feedback, and then proceeds to the infinite population limit (to be called the bottom-up approach).

4.3 PDE Methods and the Master Equation

In Lasry and Lions (2006a,b, 2007), it is proposed to obtain the MFG equation system by a finite N agent to infinite agent (or bottom-up) technique of solving a sequence of games with an increasing number of agents. Each solution would then give a Nash equilibrium for the corresponding finite population game. In this framework there are then two fundamental problems to be tackled: first, the proof of the convergence, in an appropriate sense, of the finite population Nash equilibrium solutions to limits which satisfy the infinite population MFG equations and, second, the demonstration of the existence and uniqueness of solutions to the MFG equations.

In the expository notes of Cardaliaguet (2012), the analytic properties of solutions to the infinite population HJB-FPK PDEs of MFG theory are established for finite time horizon using PDE methods including Schauder fixed-point theory and the theory of viscosity solutions. The relation to finite population games is then derived, that is to say an ε-Nash equilibrium result is established, predicated upon the assumption of strictly individual state feedback for the agents in the sequence of finite games. We observe that the analyses in both cases above will be strongly dependent upon the hypotheses concerning the functional form of the controlled dynamics of the individual agents and their cost functions, each of which may possibly depend upon the mean field measure.

4.3.1 Basic PDE Formulation

In the exposition of the basic analytic MFG theory (Cardaliaguet 2012), agents have the simple dynamics:

$$dx_t^i = u_t^i dt + \sqrt{2} dw_t^i \tag{7.19}$$

and the cost function of agent i is given in the form:

$$J_i^N(u_i, u_{-i}) = E\int_0^T \left[\frac{1}{2}(u_t^i)^2 + F(x_t^i, \mu_t^{N,-i})\right] dt + EG(x_t^i, \mu_t^{N,-i}),$$

where $\mu_t^{N,-i}$ is the empirical distribution of the states of all other agents. This leads to MFG equations in the simple form:

$$-\partial_t V - \Delta V + \frac{1}{2}|DV|^2 = F(x,m), \quad (x,t) \in \mathbb{R}^d \times (0,T) \tag{7.20}$$

$$\partial_t m - \Delta m - \text{div}(mDV) = 0, \tag{7.21}$$

$$m(0) = m_0, \quad V(x,T) = G(x, m(T)), \quad x \in \mathbb{R}^d, \tag{7.22}$$

where $V(t,x)$ and $m(t,x)$ are the value function and the density of the state distribution, respectively.

The first step is to consider the HJB equation with some fixed measure μ; it is shown by use of the Hopf-Cole transform that a unique Lipschitz continuous solution v to the new HJB equation exists for which a certain number of derivatives are Hölder continuous in space and time and for which the gradient Dv is bounded over R^n.

The second step is to show that the FPK equation with DV appearing in the divergence term has a unique solution function which is as smooth as V. Moreover, as a time-dependent measure, m is Hölder continuous with exponent $\frac{1}{2}$ with respect to the Kantorovich-Rubinstein (KR) metric.

Third, the resulting mapping of μ to V and thence to m, denoted Ψ, is such that Ψ is a continuous map from the (KR) bounded and complete space of measures with finite second moment (hence a compact space) into the same. It follows from Schauder fixed-point theorem that Ψ has a fixed point, which consequently constitutes a solution to the MFG equations with the properties listed above.

The fourth and final step is to show that the Lasry-Lions monotonicity condition (a form of strict passivity condition) on F

$$\int_{\mathbb{R}^d} F(x,m_1) - F(x,m_2) d(m_1 - m_2) > 0, \qquad \forall m_1 \neq m_2,$$

combined with a similar condition for G allowing for equality implies the uniqueness of the solution to the MFG equations.

Within the PDE setting, in-depth regularity investigation of the HJB-FPK equation under different growth and convexity conditions on the Hamiltonian have been developed by Gomes and Saude (2014).

4.3.2 General Theory: The Master Equation Method

The master equation formulation was initially introduced by P-L Lions and has been investigated by various researchers (Bensoussan et al. 2013; Cardaliaguet et al. 2015; Carmona and Delarue 2014).

The program of working from the finite population game equations and their solution to the infinite population MFG equations and their solution has been carried out in Cardaliaguet et al. (2015) for a class of systems with simple dynamics but which, in an extension of the standard MFG theory, include a noise process common to all agents in addition to their individual system noise processes. The basic idea is to reinterpret the value function of a typical player in a game of N players as a function $U(t, x_i, m)$ of time, its own state, and the empirical distribution of the states of all other players.

The analysis using the master equation begins with a set of equations which may be interpreted as the dynamic programming equations for the population. Furthermore, the information set permitted for this optimization is the full N agent system state. The derivation of the master equation is carried out (on an appropriately dimensioned torus) by arguing that the value function of a representative agent i, from a population of N, is a function of time, its own state, and a measure formed by $N - 1$ particle states and by taking the limit when $N \to \infty$. In the end, the state space for the master equation is the joint space of a generic agent state and a probability distribution.

The main result of the extensive analysis in Cardaliaguet et al. (2015) is the convergence of the set of Nash value functions $V_i^N(t_0, x), 1 \leq i \leq N$, of the set of agents for the population of size N, with initial condition $x = (x_1, \ldots, x_N)$ at time t_0, to the set of corresponding infinite population value functions U (given as solutions to the master equation), evaluated at the corresponding initial state x_i and the empirical measure m_x^N. This convergence is in the average sense

$$\frac{1}{N}\Sigma|V_i^N(t_0, x) - U(t_0, x_i, m_x^N)| \to 0, \qquad \text{as} \qquad N \to \infty.$$

In Bensoussan et al. (2015), following the derivation of the master equation for mean field type control, the authors apply the standard approach of introducing the system of HJB-FPK equations as an equilibrium solution, and then the Master equation is obtained by decoupling the HJB equation from the Fokker-Planck-Kolmogorov equation. Carmona and Delarue (2014) take a different route by deriving the master equation from a common optimality principle of dynamic programming with constraints. Gangbo and Swiech (2015) analyze the existence and smoothness of the solution for a first-order master equation which corresponds to a mean field game without involving Wiener processes.

4.4 The Hybrid Approach: PDEs and SDEs, from Infinite to Finite Populations

The infinite to finite route is top-down: one does not solve the game of N agents directly. The solution procedure involves four steps. First, one passes directly to the infinite population situation and formulates the dynamical equation and cost function for a single agent interacting with an infinite population possessing a fixed state distribution μ. Second, the stochastic optimization problem for that generic agent is then solved via dynamic programming using the HJB equation and the resulting measure for the optimally controlled agent is generated via the agent's SDE or equivalently FPK equation. Third, one solves the resulting fixed-point problem by the use of various methods (e.g., by employing the Banach, Schauder, or Kakutani fixed-point theorems). Finally, fourth, it is shown that the infinite population Nash equilibrium control laws are ε-Nash equilibrium for finite populations. This formulation was introduced in the sequence of papers Huang et al. (2003, 2006, 2007) and used in Nourian and Caines (2013), Sen and Caines (2016); it corresponds to the "limit first" method employed by Carmona, Delarue, and Lachapelle (2013) for mean field games.

Specifically, subject to Lipschitz and differentiability conditions on the dynamical and cost functions, and adopting a contraction argument methodology, one establishes the existence of a solution to the HJB-FPK equations via the Banach fixed-point theorem; the best response control laws obtained from these MFG equations are necessarily Nash equilibria within all causal feedback laws for the infinite population problem. Since the limiting distribution is equal to the original measure μ, a fixed point is obtained; in other words a consistency condition is satisfied. By construction this must be (i) a self-sustaining population distribution when all agents in the infinite population apply the corresponding feedback law, and (ii) by its construction via the HJB equation, it must be a Nash equilibrium for the infinite population. The resulting equilibrium distribution of a generic agent is called the mean field of the system.

The infinite population solution is then related to the finite population behavior by an ε-Nash equilibrium theorem which states that the cost of any agent can be reduced by at most ε when it changes from the infinite population feedback law to another while all other agents stick to their infinite population-based control strategies. Specifically, it is then shown (Huang et al. 2006) that the set of strategies $\{u_i^\circ(t) = \varphi_i(t, x_i(t)|\mu_t), 1 \leq i \leq N\}$ yields an ε-Nash equilibrium for all ε, i.e., for all $\varepsilon > 0$, there exists $N(\varepsilon)$ such that for all $N \geq N(\varepsilon)$

$$J_i^N(u_i^\circ, u_{-i}^\circ) - \varepsilon \leq \inf_{u_i \in \mathcal{U}_i} J_i^N(u_i, u_{-i}^\circ) \leq J_i^N(u_i^\circ, u_{-i}^\circ). \quad (7.23)$$

4.5 The Probabilistic Approach

4.5.1 Maximum Principle Solutions Within the Probabilistic Formulation

A different solution framework for the mean field game with nonlinear diffusion dynamics is to take a stochastic maximum principle approach (Carmona and Delarue 2013) for determining the best response of a representative agent. The procedure is carried out in the following steps: (i) A measure flow μ_t is introduced to specify the empirical state distribution associated with an infinite population. (ii) An optimal control problem is solved for that agent by introducing an adjoint process, which then determines the closed-loop system. (iii) The measure flow μ_t is then required to be equal to the law of the closed-loop state processes. This procedure yields a McKean-Vlasov forward-backward stochastic differential equation.

Necessary and sufficient conditions are available to establish the validity of the stochastic maximum (SM) principle approach to MFG theory. In particular, convexity conditions on the dynamics and the cost function, with respect to the state and controls, may be taken as sufficient conditions for the main results characterizing an MFG equilibrium through the solution of the forward-backward stochastic differential equations (FBSDEs), where the forward equation is that of the optimally controlled state dynamics and the backward equation is that of the adjoint process generating the optimal control, where these are linked by the mean field measure process. Furthermore the Lasry-Lions monotonicity condition on the cost function with respect to the mean field forms the principal hypothesis yielding the uniqueness of the solutions.

4.5.2 Weak Solutions Within the Probabilistic Formulation

Under a weak formulation of mean field games (Carmona and Lacker 2015), the stochastic differential equation in the associated optimal control problem is interpreted according to a weak solution. This route is closely related to the weak formulation of stochastic optimal control problems, also known as the martingale approach.

The solution of the mean field game starts by fixing a mean field, as a measure to describe the effect of an infinite number of agents, and a nominal measure for the probability space. Girsanov's transformation is then used to define a new probability measure under which one determines a diffusion process with a controlled drift and a diffusion term. Subsequently, the optimal control problem is solved under this new measure. Finally the consistency condition is introduced such that the distribution of the closed-loop state process agrees with the mean field. Hence the existence and uniqueness of solutions to the MFG equations under the specified conditions are obtained for weak solutions.

The proof of existence under the weak formulation relies on techniques in set-valued analysis and a generalized version of Kakutani's theorem.

4.6 MFG Equilibrium Theory Within the Nonlinear Markov Framework

The mean field game dynamic modeling framework is significantly generalized by Kolokoltsov et al. (2012) via the introduction of controlled nonlinear Markov processes where, instead of diffusion SDEs, the evolution of a typical agent is described by an integrodifferential generator of Levy-Khintchine type; as a consequence, by virtue of the Markov property, optimal control problems in this framework can still be solved within the HJB formalism.

Similar to the diffusion models described in the rest of this paper, the coefficients of the dynamical system of each agent, and its associated costs, are permitted to depend upon the empirical measure of the population of agents.

In the formal analysis, again similar to the procedures in the rest of the paper (except for Cardaliaguet et al. 2015), this measure flow is initially fixed at the infinite population limit measure and the agents then optimize their behavior via the corresponding HJB equations.

Finally, invoking the standard consistency requirement of mean field games, the MFG equations are obtained when the probability law of the resulting closed-loop configuration state is set equal to the infinite population distribution (i.e., the limit of the empirical distributions).

Concerning the methods used by Kolokoltsov et al. (2012), we observe that there are two methodologies which are employed to ensure the existence of solutions to the kinetic equations (corresponding to the FPK equations in this generalized setup) and the HJB equations: First, (i) the continuity of the mapping from the population measure to the measure generated by the kinetic (i.e., generalized FPK) equations is proven, and (ii) the compactness of the space of measures is established; then (i) and (ii) yield the existence (but not necessarily uniqueness) of a solution measure corresponding to any fixed control law via the Schauder fixed-point theory. Second, an estimate of the sensitivity of the best response mapping (i.e., control law) with respect to an a priori fixed measure flow is proven by an application of the Duhamel principle to the HJB equation. This analysis provides the ingredients which are then used for an existence theory for the solutions to the joint FPK-HJB equations of the MFG. Within this framework an ε-Nash equilibrium theory is then established in a straightforward manner.

5 Major and Minor Agents

The basic structure of mean field games can be remarkably enriched by introducing one or more major agents to interact with a large number of minor agents. A major agent has significant influence, while a minor agent has negligibly small influence on others. Such a differentiation of the strength of agents is well motivated by many practical decision situations, such as a sector consisting of a dominant corporation and many much smaller firms, the financial market with institutional traders and a huge number of small traders. The traditional game theoretic literature has studied such models of mixed populations and coined the name mixed games, but this is only in the context of static cooperative games (Haimanko 2000; Hart 1973; Milnor and Shapley 1978).

Huang (2010) introduced a large population LQG game model with mean field couplings which involves a large number of minor agents and also a major agent. A distinctive feature of the mixed agent MFG problem is that even asymptotically (as the population size N approaches infinity), the noise process of the major agent causes random fluctuation of the mean field behavior of the minor agents. This is in contrast to the situation in the standard MFG models with only minor agents. A state-space augmentation approach for the approximation of the mean field behavior of the minor agents is taken in order to Markovianize the problem and hence to obtain ε-Nash equilibrium strategies. The solution of the mean field game reduces to two local optimal control problems, one for the major agent and the other for a representative minor agent.

Nourian and Caines (2013) extend the LQG model for major and minor (MM) agents (Huang 2010) to the case of a nonlinear MFG systems. The solution to the mean field game problem is decomposed into two nonstandard nonlinear stochastic optimal control problems (SOCPs) with random coefficient processes which yield forward adapted stochastic best response control processes determined from the solution of (backward in time) stochastic Hamilton-Jacobi-Bellman (SHJB) equations. A core idea of the solution is the specification of the conditional distribution of the minor agent's state given the sample path information of the major agent's noise process. The study of mean field games with major-minor agents and nonlinear diffusion dynamics has also been developed in Carmona and Zhu (2016) and Bensoussan et al. (2013) which rely on the machinery of FBSDEs.

An extension of the model in Huang (2010) to the systems of agents with Markov jump parameters in their dynamics and random parameters in their cost functions is studied in Wang and Zhang (2012) for a discrete-time setting.

In MFG problems with purely minor agents, the mean field is deterministic, and this obviates the need for observations on other agents' states so as to determine the mean field. However, a new situation arises for systems with a major agent whose state is partially observed; in this case, best response controls generating equilibria exist which depend upon estimates of the major agent's state (Sen and Caines 2016; Caines and Kizilkale 2017).

6 The Common Noise Problem

An extension of the basic MFG system model occurs when what is called common noise is present in the global system, that is to say there is a common Wiener process whose increments appear on the right-hand side of the SDEs of every agent in the system (Ahuja 2016; Bensoussan et al. 2015; Cardaliaguet et al. 2015; Carmona and Delarue 2014). Clearly this implies that asymptotically in the population size, the individual agents cannot be independent even when each is using local state plus mean field control (which would give rise to independence in the standard case). The study of this case is well motivated by applications such as economics, finance, and, for instance, the presence of common climatic conditions in renewable resource power systems.

There are at least two approaches to this problem. First it may be treated explicitly in an extension of the master equation formulation of the MFG equations as indicated above in Sect. 4 and, second, common noise may be taken to be the state process of a passive (that is to say uncontrolled) major agent whose state process enters each agent's dynamical equation (as in Sect. 3.2.3).

This second approach is significantly more general than the former since (i) the state process of a major agent will typically have nontrivial dynamics, and (ii) the state of the major agent typically enters the cost function of each agent, which is not the case in the simplest common noise problem. An important difference in these treatments is that in the common noise framework, the control of each agent will be a function of (i.e., it is measurable with respect to) its own Wiener process and the common noise, while in the second formulation each minor agent can have complete, partial, or no observations on the state of the major agent which in this case is the common noise.

7 Applications of MFG Theory

As indicated in the Introduction, a key feature of MFG theory is the vast scope of its potential applications of which the following is a sample: Smart grid applications: (i) Dispersed residential energy storage coordinated as a virtual battery for smoothing intermittent renewable sources (Kizilkale and Malhamé 2016); (ii) The recharging control of large populations of plug-in electric vehicles for minimizing system electricity peaks (Ma et al. 2013). Communication systems: (i) Power control in cellular networks to maintain information throughput subject to interference (Aziz and Caines 2017); (ii) Optimization of frequency spectrum utilization in cognitive wireless networks; (iii) Decentralized control for energy conservation in ad hoc environmental sensor networks. Collective dynamics: (i) Crowd dynamics with xenophobia developing between two groups (Lachapelle and Wolfram 2011) and collective choice models (Salhab et al. 2015); (ii) Synchronization of coupled oscillators (Yin et al. 2012). Public health models: Mean field game-based anticipation of individual vaccination strategies (Laguzet and Turinici 2015).

For lack of space and in what follows, we further detail only three examples, respectively, drawn among smart grid applications, communication system applications, and economic applications as follows.

7.1 Residential Power Storage Control for Integration of Renewables

The general objective in this work is to coordinate the loads of potentially millions of dispersed residential energy devices capable of storage, such as electric space heaters, air conditioners, or electric water heaters; these will act as a virtual battery whose storage potential is directed at mitigating the potentially destabilizing effect of the high power system penetration of renewable intermittent energy sources (e.g., solar and wind). A macroscopic level model produces tracking targets for the mean temperature of the controlled loads, and then an application of MFG theory generates microscopic device level decentralized control laws (Kizilkale and Malhamé 2016).

A scalar linear diffusion model with state x_i is used to characterize individual heated space dynamics and includes user activity-generated noise and a heating source. A quadratic cost function is associated with each device which is designed so that (i) pressure is exerted so that devices drift toward z which is set either to their maximum acceptable comfort temperature H if extra energy storage is desired or to their minimum acceptable comfort temperature $L < H$ if load deferral is desired and (ii) average control effort and temperature excursions away from initial temperature are penalized. This leads to the cost function:

$$E \int_0^\infty e^{-\delta t}[q_t(x_i - z)^2 + q_{x_0}(x_i - x_i(0))^2 + ru_i^2]dt, \tag{7.24}$$

where $q_t = \lambda |\int_0^t (\bar{x}(\tau) - y)d\tau|$, $\bar{x}$ is the population average state with $L < \bar{x}(0) < H$, and $L < y < H$ is the population target. In the design the parameter $\lambda > 0$ is adjusted to a suitable level so as to generate a stable population behavior.

7.2 Communication Networks

The so-called CDMA communication networks are such that cell phone signals can interfere by overlapping in the frequency spectrum causing a degradation of individual signal to noise ratios and hence the quality of service. In a basic version of the standard model there are two state variables for each agent: the transmitted power $p \in R^+$ and channel attenuation $\beta \in R$. Conventional power control algorithms in mobile devices use gradient-type algorithms with bounded step size for the transmitted power which may be represented by the so-called rate adjustment model: $dp^i = u_p^i dt + \sigma_p^i dW_p^i$, $u_p^i \leq |u_{max}|, 1 \leq i \leq N$, where N represents the number of the users in the network, and W_p^i, $1 \leq i \leq N$, independent standard Wiener processes. Further, a standard model for time-varying channel attenuation is the lognormal model, where the channel gain for the ith agent with respect to the base station is given by $e^{\beta^i(t)}$ at the instant t, $0 \leq t \leq T$ and the received power at the base station from the $i$the agent is given by the product $e^{\beta^i} \times p^i$. The channel state, $\beta^i(t)$, evolves according to the power attenuation dynamics: $d\beta^i = -a^i(\beta^i + b^i)dt + \sigma_\beta^i \mathrm{d}W_\beta^i, t \geq 0, 1 \leq i \leq N$. For the generic agent $\mathscr{A}_i$ in the infinite user's case, the cost function $L_i(\beta_i, p_i)$ is given by

$$\begin{aligned}
&\lim_{N\to\infty} E\left[\int_0^T \left\{ -\frac{e^{\beta^i} p^i}{\frac{1}{N}\sum_{j=1}^n p^j e^{\beta^j} + \eta} + p^i \right\} dt\right] \\
&= \int_0^T \left\{ -\frac{e^{\beta^i} p^i}{\int_{\Omega_\beta \times \Omega_p} e^\beta p \mu_t(\beta, p) d\beta dp + \eta} + p^i \right\} dt,
\end{aligned}$$

where μ_t denotes the system mean field. As a result, the power control problem may be formulated as a dynamic game between the cellular users whereby each agent's cost function $L_i(\beta, p)$ involves both its individual transmitted power and its signal to noise ratio. The application of MFG theory yields a Nash equilibrium together with the control laws generated by the system's MFG equations (Aziz and Caines 2017). Due to the low dimension of the system state in this formulation, and indeed in that with mobile agents in a planar zone, the MFG PDEs can be solved efficiently.

7.3 Stochastic Growth Models

The model described below is a large population version of the so-called capital accumulation games (Amir 1996). Consider N agents (as economic entities). The capital stock of agent i is x_t^i and modeled by

$$dx_t^i = \left[A(x_t^i)^\alpha - \delta x_t^i\right]dt - C_t^i dt - \sigma x_t^i dw_t^i, \quad t \geq 0, \tag{7.25}$$

where $A > 0$, $0 < \alpha < 1$, $x_0^i > 0$, $\{w_t^i, 1 \leq i \leq N\}$ are i.i.d. standard Wiener processes. The function $F(x) := Ax^\alpha$ is the Cobb-Douglas production function with capital x and a constant labor size; $(\delta dt + \sigma dw_t^i)$ is the stochastic capital depreciation rate; and C_t is the consumption rate.

The utility functional of agent i takes the form:

$$J_i(C^1, \ldots, C^N) = E\left[\int_0^T e^{-\rho t} U(C_t^i, C_t^{(N,\gamma)})dt + e^{-\rho T} S(X_T)\right], \tag{7.26}$$

where $C_t^{(N,\gamma)} = \frac{1}{N}\sum_{i=1}^N (C_t^i)^\gamma$ is the population average utility from consumption. The motivation of taking the utility function $U(C_t^i, C_t^{(N,\gamma)})$ is based on relative performance. We take $\gamma \in (0,1)$ and the utility function (Huang and Nguyen 2016):

$$U(C_t^i, C_t^{(N,\gamma)}) = \frac{1}{\gamma}(C_t^i)^{\gamma(1-\lambda)}\left(\frac{(C_t^i)^\gamma}{C_t^{(N,\gamma)}}\right)^\lambda, \quad \lambda \in [0,1]. \tag{7.27}$$

So $U(C_t^i, C_t^{(N,\gamma)})$ may be viewed as a weighted geometric mean of the own utility $U_0 = (C_t^i)^\gamma/\gamma$ and the relative utility $U_1 = (C_t^i)^\gamma/(\gamma C_t^{(N,\gamma)})$. For a given θ, $U(c,\theta)$ is a hyperbolic absolute risk aversion (HARA) utility since $U(c,\theta) = \frac{c^\gamma}{\gamma\theta^\lambda}$, where $1-\gamma$ is usually called the relative risk aversion coefficient. We further take $S(x) = \frac{\eta x^\gamma}{\gamma}$, where $\eta > 0$ is a constant.

Concerning growth theory in economics, human capital growth has been considered in an MFG setting by Lucas and Moll (2014) and Guéant et al. (2011) where the individuals invest resources (such as time and money) for the improvement of personal skills to better position themselves in the labor market when competing with each other (Guéant et al. 2011).

References

Adlakha S, Johari R, Weintraub GY (2015) Equilibria of dynamic games with many players: existence, approximation, and market structure. J Econ Theory 156:269–316

Ahuja S (2016) Wellposedness of mean gield games with common noise under a weak monotonicity condition. SIAM J Control Optim 54(1):30–48

Altman E, Başar T, Srikant R (2002) Nash equilibria for combined flow control and routing in networks: asymptotic behavior for a large number of users. IEEE Trans Autom Control 47(6):917–930

Amir R (1996) Continuous stochastic games of capital accumulation with convex transitions. Games Econ Behav 15:111–131

Andersson D, Djehiche B (2011) A maximum principle for SDEs of mean-field type. Appl Math Optim 63(3):341–356

Aumann RJ, Shapley LS (1974) Values of non-atomic games. Princeton University Press, Princeton

Aziz M, Caines PE (2017) A mean field game computational methodology for decentralized cellular network optimization. IEEE Trans Control Syst Technol 25(2):563–576

Bardi M (2012) Explicit solutions of some linear-quadratic mean field games. Netw Heterog Media 7(2):243–261

Başar T, Ho YC (1974) Informational properties of the Nash solutions of two stochastic nonzero-sum games. J Econ Theory 7:370–387

Başar T, Olsder GJ (1999) Dynamic noncooperative game theory. SIAM, Philadelphia

Bauch CT, Earn DJD (2004) Vaccination and the theory of games. Proc Natl Acad Sci U.S.A. 101:13391–13394

Bauso D, Pesenti R, Tolotti M (2016) Opinion dynamics and stubbornness via multi-population mean-field games. J Optim Theory Appl 170(1):266–293

Bensoussan A, Frehse J (1984) Nonlinear elliptic systems in stochastic game theory. J Reine Angew Math 350:23–67

Bensoussan A, Frehse J, Yam P (2013) Mean field games and mean field type control theory. Springer, New York

Bensoussan A, Frehse J, and Yam SCP (2015) The master equation in mean field theory. J Math Pures Appl 103:1441–1474

Bergin J, Bernhardt D (1992) Anonymous sequential games with aggregate uncertainty. J Math Econ 21:543–562

Caines PE (2014) Mean field games. In: Samad T, Baillieul J (eds) Encyclopedia of Systems and Control. Springer, Berlin

Caines PE, Kizilkale AC (2017, in press) ε-Nash equilibria for partially observed LQG mean field games with a major player. IEEE Trans Autom Control

Cardaliaguet P (2012) Notes on mean field games. University of Paris, Dauphine

Cardaliaguet P, Delarue F, Lasry J-M, Lions P-L (2015, preprint) The master equation and the convergence problem in mean field games

Carmona R, Delarue F (2013) Probabilistic analysis of mean-field games. SIAM J Control Optim 51(4):2705–2734

Carmona R, Delarue F (2014) The master equation for large population equilibriums. In: Crisan D., Hambly B., Zariphopoulou T (eds) Stochastic analysis and applications. Springer Proceedings in Mathematics & Statistics, vol 100. Springer, Cham

Carmona R, Delarue F, Lachapelle A (2013) Control of McKean-Vlasov dynamics versus mean field games. Math Fin Econ 7(2):131–166

Carmona R, Fouque J-P, Sun L-H (2015) Mean field games and systemic risk. Commun Math Sci 13(4):911–933

Carmona R, Lacker D (2015) A probabilistic weak formulation of mean field games and applications. Ann Appl Probab 25:1189–1231

Carmona R, Zhu X (2016) A probabilistic approach to mean field games with major and minor players. Ann Appl Probab 26(3):1535–1580

Chan P, Sircar R (2015) Bertrand and Cournot mean field games. Appl Math Optim 71:533–569

Correa JR, Stier-Moses NE (2010) Wardrop equilibria. In: Cochran JJ (ed) Wiley Encyclopedia of Operations Research and Management Science. John Wiley & Sons, Inc, Hoboken

Djehiche B, Huang M (2016) A characterization of sub-game perfect equilibria for SDEs of mean field type. Dyn Games Appl 6(1):55–81

Dogbé C (2010) Modeling crowd dynamics by the mean-field limit approach. Math Comput Model 52(9–10):1506–1520

Fischer M (2014, preprint) On the connection between symmetric N-player games and mean field games. arXiv:1405.1345v1
Gangbo W, Swiech A (2015) Existence of a solution to an equation arising from mean field games. J Differ Equ 259(11):6573–6643
Gomes DA, Mohr J, Souza RR (2013) Continuous time finite state mean field games. Appl Math Optim 68(1):99–143
Gomes DA, Saude J (2014) Mean field games models – a brief survey. Dyn Games Appl 4(2): 110–154
Gomes D, Velho RM, Wolfram M-T (2014) Socio-economic applications of finite state mean field games. Phil Trans R Soc A 372:20130405. http://dx.doi.org/10.1098/rsta.2013.0405
Guéant O, Lasry J-M, Lions P-L (2011) Mean field games and applications. In: Paris-Princeton lectures on mathematical finance. Springer, Heidelberg, pp 205–266
Haimanko O (2000) Nonsymmetric values of nonatomic and mixed games. Math Oper Res 25: 591–605
Hart S (1973) Values of mixed games. Int J Game Theory 2(1):69–85
Haurie A, Marcotte P (1985) On the relationship between Nash-Cournot and Wardrop equilibria. Networks 15(3):295–308
Ho YC (1980) Team decision theory and information structures. Proc IEEE 68(6):15–22
Huang M (2010) Large-population LQG games involving a major player: the Nash certainty equivalence principle. SIAM J Control Optim 48:3318–3353
Huang M, Caines PE, Malhamé RP (2003) Individual and mass behaviour in large population stochastic wireless power control problems: centralized and Nash equilibrium solutions. In: Proceedings of the 42nd IEEE CDC, Maui, pp 98–103
Huang M, Caines PE, Malhamé RP (2007) Large-population cost-coupled LQG problems with non-uniform agents: individual-mass behavior and decentralized ε-Nash equilibria. IEEE Trans Autom Control 52:1560–1571
Huang M, Caines PE, Malhamé RP (2012) Social optima in mean field LQG control: centralized and decentralized strategies. IEEE Trans Autom Control 57(7):1736–1751
Huang M, Malhamé RP, Caines PE (2006) Large population stochastic dynamic games: closed-loop McKean-Vlasov systems and the Nash certainty equivalence principle. Commun Inf Syst 6(3):221–251
Huang M, Nguyen SL (2016) Mean field games for stochastic growth with relative utility. Appl Math Optim 74:643–668
Jovanovic B, Rosenthal RW (1988) Anonymous sequential games. J Math Econ 17(1):77–87
Kizilkale AC, Malhamé RP (2016) Collective target tracking mean field control for Markovian jump-driven models of electric water heating loads. In: Vamvoudakis K, Sarangapani J (eds) Control of complex systems: theory and applications. Butterworth-Heinemann/Elsevier, Oxford, pp 559–589
Kolokoltsov VN, Bensoussan A (2016) Mean-field-game model for botnet defense in cyber-security. Appl Math Optim 74(3):669–692
Kolokoltsov VN, Li J, Yang W (2012, preprint) Mean field games and nonlinear Markov processes. Arxiv.org/abs/1112.3744v2
Kolokoltsov VN, Malafeyev OA (2017) Mean-field-game model of corruption. Dyn Games Appl 7(1):34–47. https://doi.org/10.1007/s13235-015-0175-x
Lachapelle A, Wolfram M-T (2011) On a mean field game approach modeling congestion and aversion in pedestrian crowds. Transp Res B 45(10):1572–1589
Laguzet L, Turinici G (2015) Individual vaccination as Nash equilibrium in a SIR model with application to the 2009–2010 influenza A (H1N1) epidemic in France. Bull Math Biol 77(10):1955–1984
Lasry J-M, Lions P-L (2006a) Jeux à champ moyen. I – Le cas stationnaire. C Rendus Math 343(9):619–625
Lasry J-M, Lions P-L (2006b) Jeux à champ moyen. II Horizon fini et controle optimal. C Rendus Math 343(10):679–684
Lasry J-M, Lions P-L (2007). Mean field games. Japan J Math 2:229–260

Li T, Zhang J-F (2008) Asymptotically optimal decentralized control for large population stochastic multiagent systems. IEEE Trans Automat Control 53:1643–1660
Lucas Jr. RE, Moll B (2014) Knowledge growth and the allocation of time. J Political Econ 122(1):1–51
Ma Z, Callaway DS, Hiskens IA (2013) Decentralized charging control of large populations of plug-in electric vehicles. IEEE Trans Control Syst Technol 21(1):67–78
Milnor JW, Shapley LS (1978) Values of large games II: oceanic games. Math Oper Res 3:290–307
Neyman A (2002) Values of games with infinitely many players. In: Aumann RJ, Hart S (eds) Handbook of game theory, vol 3. Elsevier, Amsterdam, pp 2121–2167
Nourian M, Caines PE (2013) ε-Nash mean field game theory for nonlinear stochastic dynamical systems with major and minor agents. SIAM J Control Optim 51:3302–3331
Salhab R, Malhamé RP, Le Ny L (2015, preprint) A dynamic game model of collective choice in multi-agent systems. ArXiv:1506.09210
Sen N, Caines PE (2016) Mean field game theory with a partially observed major agent. SIAM J Control Optim 54:3174–3224
Tembine H, Zhu Q, Başar T (2014) Risk-sensitive mean-field games. IEEE Trans Autom Control 59:835–850
Wang BC, Zhang J-F (2012) Distributed control of multi-agent systems with random parameters and a major agent. Automatica 48(9):2093–2106
Wardrop JG (1952) Some theoretical aspects of road traffic research. Proc Inst Civ Eng Part II, 1:325–378
Weintraub GY, Benkard C, Van Roy B (2005) Oblivious equilibrium: a mean field approximation for large-scale dynamic games. Advances in neural information processing systems, MIT Press, Cambridge
Weintraub GY, Benkard CL, Van Roy B (2008) Markov perfect industry dynamics with many firms. Econometrica 76(6):1375–1411
Yin H, Mehta PG, Meyn SP, Shanbhag UV (2012) Synchronization of coupled oscillators is a game. IEEE Trans Autom Control 57:920–935

Zero-Sum Differential Games

8

Pierre Cardaliaguet and Catherine Rainer

Contents

P. Cardaliaguet (✉)
Université Paris-Dauphine, Paris, France
e-mail: cardaliaguet@ceremade.dauphine.fr

C. Rainer
Département, Laboratoire de Mathématiques LMBA, Université de Brest, Brest Cedex, France
e-mail: Catherine.Rainer@univ-brest.fr

T. Başar, G. Zaccour (eds.), *Handbook of Dynamic Game Theory*,
https://doi.org/10.1007/978-3-319-44374-4_4

Abstract

The chapter is devoted to two-player, zero-sum differential games, with a special emphasis on the existence of a value and its characterization in terms of a partial differential equation, the Hamilton-Jacobi-Isaacs equation. We discuss different classes of games: in finite horizon, in infinite horizon, and pursuit-evasion games. We also analyze differential games in which the players do not have a full information on the structure of the game or cannot completely observe the state. We complete the chapter by a discussion on differential games depending on a singular parameter: for instance, we provide conditions under which the differential game has a long-time average.

Keywords

Differential games · Zero-sum games · Viscosity solutions · Hamilton-Jacobi equations · Bolza problem · Pursuit-evasion games · Search games · Incomplete information · Long-time average · Homogenization

1 Introduction

Differential game theory investigates conflict problems in systems which are driven by differential equations. The topic lies at the intersection of game theory (several players are involved) and of controlled systems (the state is driven by differential equations and is controlled by the players). This chapter is devoted to the analysis of two-player, zero-sum differential games.

The typical example of such games is the pursuit-evasion game, in which a pursuer tries to steer in minimal time an evader to a given target. This kind of problem appears, for instance, in aerospace problems. Another example is optimal control with disturbances: the second player is seen as a perturbation against which the controller has to optimize the behavior of the system; one then looks at a worst-case design (sometimes also called Knightian uncertainty, in the economics literature).

We discuss here several aspects of zero-sum differential games. We first present several examples and explain, through the so-called verification theorem, the main formal ideas on these games. In Sects. 3 (deterministic dynamics), 4 (stochastic ones), and 5 (pursuit-evasion games), we explain in a rigorous way when the game has a value and characterize this value in terms of a partial differential equation, the Hamilton-Jacobi-Isaacs (HJI) equation. In the games discussed in Sects. 3, 4, and 5, the players have a perfect knowledge of the game and a perfect observation of the action of their opponent. In Sect. 6 we describe zero-sum differential games with an information underpinning. We complete the chapter by the analysis of differential games with a singular parameter, for instance, defined on a large time interval or depending on a small parameter (Sect. 7).

This article is intended as a brief introduction to two-player, zero-sum differential games. It is impossible to give a fair account of this large topic within a few pages: therefore, we have selected some aspects of the domain, the choice reflecting only the taste of the authors. It is also impossible to give even a glimpse of the huge literature on the subject, because it intersects and motivates researches on partial differential equations, stochastic processes, and game theory. The interested reader will find in the Reference part a list of monographs, surveys, and a selection of pioneering works for further reading. We also quote directly in the text several authors, either for their general contributions to the topic or for a specific result (in which case we add the publication date, so that the reader can easily find the reference).

2 Isaacs' Approach

We start the analysis of zero-sum differential games by briefly describing some typical examples (or classes of examples). Then we explain in a simple framework the basic principle for solving a game in practice. This part is a (modest) tribute to Isaacs, who introduced most ideas on the subject. Let us stress the point that Isaacs' theory goes much beyond the few points developed here. In particular, whereas we will be interested mostly in cases where the value function is smooth, Isaacs investigated in detail their possible singularities.

2.1 Examples

We present here several classical differential games, from the most emblematic ones (the pursuit-evasion games) to the classical ones (the Bolza problem or the infinite horizon problem).

2.1.1 Pursuit-Evasion Games

Pursuit-evasion differential games are a class of differential games in which one of the players (the pursuer) tries to steer the state to a given target in minimal time, while his opponent (the evader) is doing his best to postpone the capture. This class of games is very natural, and it is the first one to have been studied historically: Isaacs introduced in his famous monograph (Isaacs 1965) a number of examples, which he contributed to solve (at least partially). It is interesting to quote some of them, if only to provide a hint of what a differential game is.

1. Lion and Man

   A lion and a man are in a circular arena. They observe each other and have equal maximum speed. The lion tries to catch the man as quickly as possible, while the man wants to escape the lion for as long as he can, forever if possible. Capture holds when the position of both players coincides. This game was originally posed in 1925 by Rado.

   A possible strategy for the lion to capture the man is the following: the lion first moves to the center of the arena and then remains on the radius that passes through the man's position. Since the man and the lion have the same maximum speed, the lion can remain on the radius and simultaneously move toward the man. One can show that the lion can get arbitrarily close to the man; Besicovitch ('52) also proved that the man can postpone forever the capture.

2. Homicidal Chauffeur Game

   This is one of the most well-known examples of a differential game. Introduced by Isaacs in a 1951 report for the RAND Corporation, the game consists of a car striving as quickly as possible to run over a pedestrian. The catch is that the car has a bounded acceleration and a large maximum speed, while the pedestrian has no inertia but relatively small maximum speed. A nice survey by Patsko and Turova ('09) is dedicated to the problem.

3. Princess and Monster

   A third very well-known game is princess and monster. It was also introduced by Isaacs, who described it as follows:

   > The monster searches for the princess, the time required being the payoff. They are both in a totally dark room (of any shape), but they are each cognizant of its boundary. Capture means that the distance between the princess and the monster is within the capture radius, which is assumed to be small in comparison with the dimension of the room. The monster, supposed highly intelligent, moves at a known speed. We permit the princess full freedom of locomotion.

   This game is a typical example of "search games" and is discussed in Sect. 6.1.

Let us now try to formalize a little the pursuit-evasion games. In these games, the state (position of the players and, in the case of the homicidal chauffeur game, the angular velocity of the chauffeur) is a point X in some Euclidean space $\mathbb{R}^d$. It evolves in time and we denote by X_t the position at time t. The state X_t moves according to a differential equation

$$\dot{X}_t = f(X_t, u_t, v_t)$$

where $\dot{X}_t$ stands for the derivative of the map $t \to X_t$ and f is the evolution law depending on the controls u_t and v_t of the players. The first player chooses u_t and the second one v_t and both players select their respective controls at each instant of time according to the position of the state and the control played until then by their opponent. Both controls are in general restricted to belong to given sets: $u_t \in U$ and $v_t \in V$ for all t. The capture occurs as the state of the system reaches a given set of positions $C \subseteq \mathbb{R}^d$. If we denote by θ_C the capture time, then

$$\theta_C = \inf\{t \geq 0,\ X_t \in C\}.$$

The pursuer (i.e., player 1) wants to minimize θ_C, and the evader (player 2) wants to maximize it.

For instance, in the lion and man game, $X = (x, y) \in \mathbb{R}^2 \times \mathbb{R}^2$, where x is the position of the lion (the pursuer) and y the position of the man. The dynamics here is particularly simple since players are supposed to choose their velocities at each time:

$$\dot{x}_t = u_t, \qquad \dot{y}_t = v_t$$

where u_t and v_t are bounded (in norm) by the same maximal speed (say 1, to fix the ideas). So here f is just $f(X, u, v) = (u, v)$ and $U = V$ is the unit ball of $\mathbb{R}^2$. Capture occurs as the position of the two players coincide, i.e., as the state reaches the target $C := \{X = (x, y),\ x = y\}$. One also has to take into account the fact that the lion and the man have to remain in the arena, which leads to constraints on the state.

2.1.2 The Bolza and the Infinite Horizon Problems

There are two other well-known classes of differential games: the Bolza problem, where the game is played during a fixed duration, and the infinite horizon problem, where the game is played forever, but payoff is discounted. In both cases, the state is a point X in $\mathbb{R}^d$ (for some $d \in \mathbb{N}\backslash\{0\}$) which evolves in time according to an ordinary differential equation of the form

$$\dot{X}_t = f(X_t, u_t, v_t).$$

As in pursuit-evasion games, the first player chooses u_t in a given set U and the second one v_t in a fixed set V. The criterion, however, is different.

- In the Bolza problem, the horizon is fixed, given by a terminal time $T > 0$. The cost of the first player (it is a payoff for the second one) is of the form

$$\int_0^T \ell(X_t, u_t, v_t)\, dt + g(X_T)$$

  where ℓ is called the running cost and g the terminal cost.
- In the infinite horizon problem, the cost takes the form

$$\int_0^\infty e^{-rt} \ell(X_t, u_t, v_t)\, dt$$

  where $r > 0$ is a discount rate and where the running cost ℓ is supposed to be bounded (to ensure the convergence of the integral).

It often happens that the state is driven by a stochastic differential equation instead of an ordinary one:

$$dX_t = f(X_t, u_t, v_t)dt + \sigma(X_t, u_t, v_t)dB_t,$$

where B_t is a Brownian motion and σ is a matrix. The costs take the same form, but in expectation. For instance, for the Bolza problem, it becomes

$$\mathbb{E}\left[\int_0^T \ell(X_t, u_t, v_t)\, dt + g(X_T)\right].$$

2.2 A Verification Theorem for Pursuit-Evasion Games

Let us now start the analysis of two-player zero-sum differential games by presenting some ideas due to Isaacs and his followers. Although most of these ideas rely on an a priori regularity assumption on the value function which does not hold in general, they shed an invaluable light on the problem by revealing – in an idealized and simplified framework – many phenomena that will be encountered in a more technical setting throughout the other sections. Another interesting aspect of these techniques is that they allow to solve explicitly several games, in contrast with the subsequent theories which are mainly concerned with theoretical issues and the numerical approximation of the games.

We restrict ourselves to present the very basic aspects of this theory, a deeper analysis being beyond the scope of this treatment. The interested reader can find significant developments in the monographs by Isaacs (1965), Blaquière et al. (1969), Başar and Olsder (1999), Lewin (1994), and Melikyan (1998).

2.2.1 Definition of the Value Functions

We consider the pursuit-evasion game with dynamics given by

$$\dot{X}_t = f(X_t, u_t, v_t),$$

where u_t belongs for any t to some control set U and is chosen at each time by the first player and v_t belongs to some control set V and is chosen by the second player. The state of the system X_t lies in $\mathbb{R}^d$, and we will always assume that $f : \mathbb{R}^d \times U \times V \to \mathbb{R}^d$ is smooth and bounded, so that the solution of the above differential equation will be defined on $[0, +\infty)$. We denote by $C \subseteq \mathbb{R}^d$ the target and we recall that the first player (the pursuer) aims at minimizing the capture time.

We now explain how the players choose their controls at each time.

Definition 1 (Feedback strategies). A feedback strategy for the first player (resp. for the second player) is a map $\overline{u} : \mathbb{R}^+ \times \mathbb{R}^d \to U$ (resp. $\overline{v} : \mathbb{R}^+ \times \mathbb{R}^d \to V$).

This means that each player chooses at each time t his action as a function of the time and of the current position of the system. As will be extensively explained later on, other definitions of strategies are possible (and in fact more appropriate to prove the existence of the value).

The main issue here is that, given two arbitrary strategies $\overline{u} : \mathbb{R}^+ \times \mathbb{R}^d \to U$ and $\overline{v} : \mathbb{R}^+ \times \mathbb{R}^d \to V$ and an initial position $x_0 \in \mathbb{R}^d$, the system

$$\begin{cases} \dot{X}_t = f(X_t, \overline{u}(t, X_t), \overline{v}(t, X_t)) \\ x(0) = x_0 \end{cases} \tag{8.1}$$

does not necessarily have a solution. Hence, we have to restrict the choice of strategies for the players. We suppose that the players choose their strategies in "sufficiently large" sets $\overline{U}$ and $\overline{V}$ of feedback strategies, such that, for any $\overline{u} \in \overline{U}$ and $\overline{v} \in \overline{V}$, there exists a unique absolutely continuous solution to (8.1). This solution is denoted by $X_t^{x_0,\overline{u},\overline{v}}$.

Let us explain what it means to reach the target: for given a trajectory $X : [0, +\infty) \to \mathbb{R}^d$, let

$$\theta_C(X) := \inf\{t \geq 0 \mid X_t \in C\}.$$

We set $\theta_C(X) := +\infty$ if $X_t \notin C$ for all $t \geq 0$.

Definition 2. For a fixed admissible pair $(\overline{U}, \overline{V})$ of sets of strategies, the upper and the lower value functions are respectively given by

$$\mathbf{V}^+(x_0) := \inf_{\overline{u}\in\overline{U}} \sup_{\overline{v}\in\overline{V}} \theta_C(X^{x_0,\overline{u},\overline{v}}) \qquad \text{and} \qquad \mathbf{V}^-(x_0) := \sup_{\overline{v}\in\overline{V}} \inf_{\overline{u}\in\overline{U}} \theta_C(X^{x_0,\overline{u},\overline{v}}).$$

We say that the game has a value if $\mathbf{V}^+ = \mathbf{V}^-$. In this case we say that the map $\mathbf{V} := \mathbf{V}^+ = \mathbf{V}^-$ is the value of the game. We say that a strategy $\overline{u}^* \in \overline{U}$ (resp. $\overline{v}^* \in \overline{V}$) is optimal for the first player (resp. the second player) if

$$\mathbf{V}^-(x_0) := \sup_{\overline{v}\in\overline{V}} \theta_C(X^{x_0,\overline{u}^*,\overline{v}}) \qquad (\text{resp.}\mathbf{V}^+(x_0) := \inf_{\overline{u}\in\overline{U}} \theta_C(X^{x_0,\overline{u},\overline{v}^*})).$$

Let us note that upper and lower value functions a priori depend on the sets of strategies $(\overline{U}, \overline{V})$. Actually, with this definition of value function, the fact that the game has a value is an open (but not very interesting) issue. The reason is that the set of strategies is too restrictive and has to be extended in a suitable way: this issue will be addressed in the next section. However, we will now see that the approach is nevertheless useful to understand some crucial ideas on the problem.

2.2.2 The Verification Theorem

The following verification theorem – due to Isaacs – allows us to show that a given function is indeed the value function of the game: it will be enough to check that this (supposedly smooth) function is the solution of a partial differential equation (PDE) called the Hamilton-Jacobi-Isaacs equation (HJI).

We associate with the dynamics of the game a *Hamiltonian* H. For this we suppose that the so-called Isaacs' condition holds, which means that an infinitesimal game has a value:

$$H(x, p) := \inf_{u\in U}\sup_{v\in V}\langle p, f(x,u,v)\rangle = \sup_{v\in V}\inf_{u\in U}\langle p, f(x,u,v)\rangle \quad \forall (x,p)\in \mathbb{R}^d\times\mathbb{R}^d. \tag{8.2}$$

Let us note that Isaacs' condition holds, for instance, if the dynamics is separate, i.e., if $f(x,u,v) = f_1(x,u) + f_2(x,v)$.

For any $(x,p) \in \mathbb{R}^d \times \mathbb{R}^d$, we denote by $\tilde{u}(x,p)$ and $\tilde{v}(x,p)$ the (supposedly unique) optimal elements in the definition of H:

$$\max_{v\in V}\langle p, f\,(x, \tilde{u}(x,p), v)\rangle = \min_{u\in U}\langle p, f(x,u,\tilde{v}(x,p))\rangle = H(x,p).$$

Theorem 1 (Verification Theorem). *Let us assume that the target is closed and that Isaacs'condition (8.2) holds. Suppose that there is a nonnegative map* $\mathbf{V}:\mathbb{R}^d\to\mathbb{R}$, *continuous on* $\mathbb{R}^d$ *and of class* $\mathcal{C}^1$ *over* $\mathbb{R}^d\backslash C$, *with* $\mathbf{V}(x)=0$ *on* C *which satisfies the Hamilton-Jacobi-Isaacs (HJI) equation:*

$$H(x, D\mathbf{V}(x)) + 1 = 0 \qquad \forall x\in\mathbb{R}^d\backslash C. \tag{8.3}$$

Let us furthermore assume that the maps $x\to\overline{u}^*(x) := \tilde{u}(x, D\mathbf{V}(x))$ *and* $x\to\overline{v}^*(x) := \tilde{v}(x, D\mathbf{V}(x))$ *belong to* $\overline{U}$ *and* $\overline{V}$, *respectively.*

Then the game has a value and $\mathbf{V}$ *is the value of the game:* $\mathbf{V}=\mathbf{V}^-=\mathbf{V}^+$. *Moreover, the strategies* $\overline{u}^*(x)$ *and* $\overline{v}^*(x)$ *are optimal, in the sense that*

$$\mathbf{V}(x) = \theta_C(X^{x_0,\overline{u}^*,\overline{v}^*}) = \inf_{\overline{u}\in\overline{U}}\theta_C(X^{x_0,\overline{u},\overline{v}^*}) = \sup_{\overline{v}\in\overline{V}}\theta_C(X^{x_0,\overline{u}^*,\overline{v}})$$

for all $x\in\mathbb{R}^d\backslash C$.

This result is quite striking since it reduces the resolution of the game to that of solving a PDE and furthermore provides at the same time the optimal feedbacks of the players. Unfortunately it is of limited interest because the value functions $\mathbf{V}^-$ and $\mathbf{V}^+$ are very seldom smooth enough for this result to apply. For this reason Isaacs' theory is mainly concerned with the singularities of the value functions, i.e., the set of points where they fail to be continuous or differentiable.

The converse also holds true: if the game has a value and if this value has the regularity described in the theorem, then it satisfies the HJI equation (8.3). We will explain in the next section that this statement holds also in a much more general setting.

Sketch of proof: Let us fix $x_0\in\mathbb{R}^d$. We first claim that

$$\inf_{\overline{u}}\theta_C(X^{x_0,\overline{u},\overline{v}^*}) \ge V(x_0).$$

For this, let $\overline{u}\in\overline{U}$ and let us set $X_t = X^{x_0,\overline{u},\overline{v}^*}$ and $\tau := \theta_C(X^{x_0,\overline{u},\overline{v}^*})$. Then, for any $t\in[0,\tau)$, we have

$$\begin{aligned}\tfrac{d}{dt}\mathbf{V}(X_t) &= \langle D\mathbf{V}(X_t), f(X_t, \overline{u}(t, X_t), \overline{v}^*(X_t))\rangle \\ &\geq \inf_{u\in U}\langle D\mathbf{V}(X_t), f(X_t, u, \overline{v}^*(X_t))\rangle \\ &= H(X_t, D\mathbf{V}(X_t)) \ = \ -1.\end{aligned}$$

Integrating the above inequality between 0 and $t \leq \tau$, we get

$$\mathbf{V}(X_t) - \mathbf{V}(x_0) \geq -t.$$

For $t = \tau$, we have $\mathbf{V}(X_\tau) = 0$ since X_τ belongs to C and $\mathbf{V} = 0$ on C. So by continuity of $\mathbf{V}$ on $\mathbb{R}^d$,

$$\mathbf{V}(x_0) \leq \tau = \theta_C(X^{x_0,\overline{u},\overline{v}^*}).$$

One can show in a similar way that $\mathbf{V}(x_0) \geq \theta_C(X^{x_0,\overline{u}^*,\overline{v}})$ for any $\overline{v} \in \overline{V}$. Hence,

$$\theta_C(X^{x_0,\overline{u}^*,\overline{v}}) \leq \mathbf{V}(x_0) \leq \theta_C(X^{x_0,\overline{u},\overline{v}^*}),$$

which shows at the same time that the game has a value, that this value is $\mathbf{V}$, and that the strategies $\overline{u}^*$ and $\overline{v}^*$ are optimal. □

2.2.3 The Hamiltonian System

We now give some hints about the explicit computation of the value function $\mathbf{V}$. The key idea is that it is possible to compute $\mathbf{V}$ along the characteristics associated with the HJI equation (8.3). This system of ordinary differential equations is a Hamiltonian system.

Throughout the rest of this section, we assume that the value function $\mathbf{V}$ and the data are smooth enough to perform the computation.

Theorem 2. *Let $x_0 \in \mathbb{R}^d \backslash C$ be an initial position and $\overline{u}^*$ and $\overline{v}^*$ be the optimal strategies given in Theorem (1) and let us set $X_t = X_t^{x_0,\overline{u}^*,\overline{v}^*}$. Then the pair (X, P), where $P_t := D\mathbf{V}(X_t)$, is a solution of the Hamiltonian system*

$$\begin{cases} \dot{X}_t = \frac{\partial H}{\partial p}(X_t, P_t) \\ \dot{P}_t = -\frac{\partial H}{\partial x}(X_t, P_t) \end{cases} \tag{8.4}$$

on $[0, \mathbf{V}(x_0))$.

The variable P is often called the adjoint variable of X. In control theory (i.e., when f only depends on u or on v), the existence of such an adjoint is an optimality condition for a given trajectory X. This statement is the famous *Pontryagin maximum principle*.

Sketch of proof: The starting point is the remark that one can express the derivative of H with respect to p in terms of $\tilde{u}$ and $\tilde{v}$. Namely,

$$\frac{\partial H}{\partial p}(x,p) = f(x, \tilde{u}(x,p), \tilde{v}(x,p)).$$

This result is known as the envelope theorem.

As a first consequence, we obtain $\dot{X}_t = \frac{\partial H}{\partial p}(X_t, P_t)$. By definition of P,

$$\dot{P}_t = D^2\mathbf{V}(X_t)\dot{X}_t = D^2\mathbf{V}(X_t) f(X_t, \overline{u}^*(X_t), \overline{v}^*(X_t))) = D^2\mathbf{V}(X_t)\frac{\partial H}{\partial p}(X_t, P_t).$$

Differentiating the HJI equation (8.3), we get

$$\frac{\partial H}{\partial x}(x, D\mathbf{V}(x)) + D^2\mathbf{V}(x)\frac{\partial H}{\partial p}(x, D\mathbf{V}(x)) = 0 \qquad \forall x \in \mathbb{R}^d \backslash C \ ,$$

from which we deduce the equation satisfied by P_t. □

Next we describe the behavior of the value function at the boundary of the target. For this we assume that the target is the closure of an open set with a smooth boundary, and we denote by ν_x the outward unit normal to C at a point $x \in \partial C$.

Proposition 1. *If* $\mathbf{V}$ *is of class* $\mathcal{C}^1$ *on* $\overline{\mathbb{R}^d \backslash C}$*, then*

$$\forall x \in \partial C,\ H(x, \nu_x) < 0 \text{ and } D\mathbf{V}(x) = -\frac{\nu_x}{H(x,\nu_x)}.$$

Note that the pair (X, P) in Theorem 2 has an initial condition for X (namely, $X_0 = x_0$) and a terminal condition for P (namely, $P_T = -\frac{\nu_{X_T}}{H(x,\nu_{X_T})}$ where $T = \mathbf{V}(x_0)$). We will see below that the condition $H(x, \nu_x) \leq 0$ in ∂C is necessary in order to ensure the value function to be continuous at a point $x \in \partial C$.

Indeed, if $x \in \partial C$ and since $\mathbf{V} = 0$ on ∂C and is nonnegative on $\mathbb{R}^d \backslash C$, it has a minimum on $\overline{\mathbb{R}^d \backslash C}$ at x. The Karush-Kuhn-Tucker condition then implies the existence of $\lambda \geq 0$ with $D\mathbf{V}(x) = \lambda \nu_x$. Recalling that $H(x, D\mathbf{V}(x)) = -1$ gives the result, thanks to the positive homogeneity of $H(x, \cdot)$. □

Construction of V: The above result implies that, if we solve the backward system

$$\begin{cases} \dot{X}_t = -\frac{\partial H}{\partial p}(X_t, P_t) \\ \dot{P}_t = \frac{\partial H}{\partial x}(X_t, P_t) \\ X_0 = \xi,\ P_0 = -\frac{\nu_\xi}{H(\xi,\nu_\xi)} \end{cases} \tag{8.5}$$

for $\xi \in \partial C$, then one should have

$$\mathbf{V}(X_t) = t \qquad \forall t \geq 0.$$

We can take the above equality as a definition of $\mathbf{V}$. Unfortunately such a definition seldom defines $\mathbf{V}$ unambiguously: there might be a point x_0, two initial conditions ξ and $\tilde{\xi}$ on the boundary of C, and two times t and $\tilde{t}$ such that $x_0 = X_t = \tilde{X}_{\tilde{t}}$, where $(\tilde{X}, \tilde{P})$ is the solution of (8.5) starting from $(\tilde{\xi}, -\nu_{\tilde{\xi}}/H(\tilde{\xi}, \nu_{\tilde{\xi}}))$. Then it is not easy to decide if $\mathbf{V}(x_0)$ is equal to t or $\tilde{t}$, or even some other value. In this case, the map $\mathbf{V}$ *cannot be* $\mathcal{C}^1$, and one has to study the *singularities* of $\mathbf{V}$, that is, the set of points at which $\mathbf{V}$ fails to be differentiable or even continuous. We explain ideas in this direction in the next part.

2.2.4 The Usable Part of the Boundary and Discontinuities

We now discuss the possible discontinuities of the value function $\mathbf{V}$. This will yield to two important notions: the usable part of the target and the Isaacs equation. These two notions play an important role in Sect. 5.

Let us start with the behavior at the boundary of the target. We have seen above (Proposition 1) that a necessary condition for the value to be of class $\mathcal{C}^1$ on $\overline{\mathbb{R}^d\backslash C}$ is that

$$\forall x \in \partial C,\ \ H(x, \nu_x) < 0,$$

where ν_x is the outward unit normal to C at x. This leads us to call **usable part of the boundary** ∂C the set of points $x \in \partial C$ such that $H(x, \nu_x) < 0$. We denote this set by UP.

Proposition 2. *If $x \in UP$, then $\mathbf{V}$ is continuous and vanishes at x. On the contrary, if $H(x, \nu_x) > 0$, then $\mathbf{V}$ is bounded below by a positive constant in a neighborhood of x in $\mathbb{R}^d\backslash C$.*

In particular, if the usable part is the whole boundary ∂C, then the HJI equation (8.3) can indeed be supplemented with a boundary condition: $\mathbf{V} = 0$ on ∂C. We will use this remark in the rigorous analysis of pursuit-evasion games (Sect. 5).

The proposition states that the pursuer can ensure an almost immediate capture in a neighborhood of the points of *UP*. On the contrary, in a neighborhood of a point $x \in \partial C$ such that $H(x, \nu_x) > 0$ holds, the evader can postpone the capture at least for a positive time. What happens at points $x \in \partial C$ such that $H(x, \nu_x) = 0$ is much more intricate. The set of such points – improperly called boundary of the usable part (*BUP*) – plays a central role in the computation of the boundary of the domain of the value function.

Sketch of proof: If $x \in UP$, then, from the definition of H,

$$H(x, \nu_x) = \inf_{u\in U} \sup_{v\in V} \langle f(x, u, v), \nu_x\rangle \ < \ 0.$$

Thus, there exists $u_0 \in U$ such that, for any $v \in V$, $\langle f(x, u_0, v), \nu_x)\rangle < 0$. If player 1 plays u_0 in a neighborhood of x, it is intuitively clear that the trajectory is going to reach the target C almost immediately whatever the action of player 2. The second part of the proposition can be proved symmetrically. □

We now discuss what can possibly happen when the value has a **discontinuity** outside of the target. Let us fix a point $x \notin C$ and let us assume that there is a (smooth) hypersurface described by $\{\psi = 0\}$ along which $\mathbf{V}$ is discontinuous. We suppose that ψ is a smooth map with $D\psi(x) \neq 0$, and we assume that there exist two maps $\mathbf{V}_1$ and $\mathbf{V}_2$ of class $\mathcal{C}^1$ such that, in a neighborhood of the point x,

$$\mathbf{V}(y) = \mathbf{V}_1(y) \text{ if } \psi(y) > 0, \qquad \mathbf{V}(y) = \mathbf{V}_2(y) \text{ if } \psi(y) < 0 \qquad \text{and} \qquad \mathbf{V}_1 > \mathbf{V}_2.$$

In other word, the surface $\{\psi = 0\}$ separates the part where $\mathbf{V}$ is large (i.e., when $\mathbf{V} = \mathbf{V}_1$) from the part where $\mathbf{V}$ is small (in which case $\mathbf{V} = \mathbf{V}_2$).

Proposition 3. *In the above configuration, one has* $H(x, D\psi(x)) = 0$.

One says that the set $\{x \mid \psi(x) = 0\}$ is a *barrier* for the game. It is a particular case of a semipermeable surface which have the property that each player can prevent the other one from crossing the surface in one direction.

Equation $H(x, D\psi(x)) = 0$ is called Isaacs' equation. It is a geometric equation, in the sense that one is interested not in the solution ψ itself but in the set $\{x \mid \psi(x) = 0\}$.

An important application of the proposition concerns the domain of the value function (i.e., the set $dom(\mathbf{V}) := \{x \mid \mathbf{V}(x) < +\infty\}$). If the boundary of the domain is smooth, then it satisfies Isaacs' equation.

Sketch of proof: If on the contrary one had $H(x, D\psi(x)) > 0$, then, from the very definition of H, there would exist $v_0 \in V$ for the second player such that

$$\inf_{u \in U} \langle f(x, u, v_0), D\psi(x)\rangle > 0.$$

This means that the second player can force (by playing v_0) the state of the system to cross the boundary $\{\psi = 0\}$ from the part where $\mathbf{V}$ is small (i.e., $\{\psi < 0\}$) to the part where $\mathbf{V}$ is large (i.e., $\{\psi > 0\}$). However, if player 1 plays in an optimal way, the value should not increase along the path and in particular cannot have a positive jump. This leads to a contradiction. □

3 Approach by Viscosity Solutions

In this section, which is the heart of the chapter, we discuss the existence of the value function and its characterization. The existence of the value – in pure strategies – holds under a condition on the structure of the game: the so-called Isaacs' condition.

Without this condition, the players have to play in random strategies[1]: this issue being a little tricky in continuous time, we describe precisely how to proceed. We mostly focus on the Bolza problem and then briefly explain how to adapt the ideas to the infinite horizon problem. The analysis of pursuit-evasion game is postponed to Sect. 5, as it presents specific features (discontinuity of the value).

3.1 The Bolza Problem: The Value Functions

From now on we concentrate on the Bolza problem. In this framework one can rigorously show the existence of a value and characterize this value as a viscosity solution of a Hamilton-Jacobi-Isaacs equation (HJI equation). The results in this section go back to Evans and Souganidis (1984) (see also Bardi and Capuzzo Dolcetta 1996, Chap. VIII).

3.1.1 Strategies

The notion of strategy describes how each player reacts to its environment and to its opponent's behavior. In this section we assume that the players have a perfect monitoring, i.e., not only they know perfectly the game (its dynamics and payoffs), but each one also perfectly observes the control played by its opponent up to that time. Moreover, they can react (almost) immediately to what they have seen.

Even under this assumption, there is no consensus on the definition of a strategy: the reason is that there is no concept of strategies which at the same time formalizes the fact that each player reacts immediately to its opponent's past action and allows the strategies of both players to be played together. This issue has led to the introduction of many different definitions of strategies, which in the end have been proved to provide the same value function.

Let U and V be metric spaces and $-\infty < t_0 < t_1 \leq +\infty$. We denote by $\mathcal{U}_{(t_0,t_1)}$ the set of bounded Lebesgue measurable maps $u : [t_0, t_1] \to U$. We set $\mathcal{U}_{t_0} := \mathcal{U}_{(t_0,+\infty)}$ (or, if the game has a fixed horizon T, $\mathcal{U}_{t_0} := \mathcal{U}_{(t_0,T)}$). Elements of $\mathcal{U}_{t_0}$ are called the *open-loop controls* of player 1. Similarly we denote by $\mathcal{V}_{(t_0,t_1)}$ the set of bounded Lebesgue measurable maps $v : [t_0, t_1] \to V$. We will systematically call player 1 the player playing with the open-loop control u and player 2 the player playing with the open-loop control v. If $u_1, u_2 \in \mathcal{U}_{t_0}$ and $t_1 \geq t_0$, we write $u_1 \equiv u_2$ on $[t_0, t_1]$ whenever u_1 and u_2 coincide almost everywhere on $[t_0, t_1]$.

A strategy for player 1 is a map α from $\mathcal{V}_{t_0}$ to $\mathcal{U}_{t_0}$. This means that player 1 answers to each control $v \in \mathcal{V}_{t_0}$ of player 2 by a control $u = \alpha(v) \in \mathcal{U}_{t_0}$. However, since we wish to formalize the fact that no player can guess in advance the future behavior of the other player, we require that such a map α is nonanticipative.

[1]Unless one allows an information advantage to one player, amounting to letting him know his opponent's control at each time (Krasovskii and Subbotin 1988).

Definition 3 (Nonanticipative strategy). A map $\alpha : V_{t_0} \to U_{t_0}$ is nonanticipative if, for any time $t_1 > t_0$ and any pair of controls $v_1, v_2 \in V_{t_0}$ such that $v_1 \equiv v_2$ on $[t_0, t_1]$, we have $\alpha(v_1) \equiv \alpha(v_2)$ on $[t_0, t_1]$.

The notion of nonanticipative strategies (as well as the notion of delay strategies introduced below) has been introduced in a series of works by Varaiya, Ryll-Nardzewski, Roxin, Elliott, and Kalton.

We denote by $\mathcal{A}(t_0)$ the set of player 1's nonanticipative strategies $\alpha : V_{t_0} \to U_{t_0}$. In a symmetric way we denote by $\mathcal{B}(t_0)$ the set of player 2's nonanticipative strategies, which are the nonanticipative maps $\beta : U_{t_0} \to V_{t_0}$.

In order to put the game under normal form, one should be able to say that, for any pair of nonanticipative strategies $(\alpha, \beta) \in \mathcal{A}(t_0) \times \mathcal{B}(t_0)$, there is a unique pair of controls $(u, v) \in U_{t_0} \times V_{t_0}$ such that

$$\alpha(v) \equiv u \text{ and } \beta(u) \equiv v. \tag{8.6}$$

This would mean that, when players play simultaneously the strategies α and β, the outcome should be the pair of controls (u, v). Unfortunately this is not possible: for instance, if $U = V = [-1, 1]$ and $\alpha(v)_t = -v_t$ if $v_t \neq 0$, $\alpha(v)_t = 1$ if $v_t = 0$ while $\beta(u)_t = u_t$, then there is no pair of control $(u, v) \in U_{t_0} \times V_{t_0}$ for which (8.6) holds. Another bad situation is described by a couple (α, β) with $\alpha(v) \equiv v$ and $\beta(u) \equiv u$ for all u and v: there are infinitely many solutions to (8.6).

For this reason we are lead to introduce a more restrictive notion of strategy, the nonanticipative strategies with delay.

Definition 4 (Delay strategies). A nonanticipative strategy with delay (in short delay strategy) for player 1 is a map $\alpha : V_{t_0} \to U_{t_0}$ for which there is a delay $\tau > 0$ such that, for any two controls $v_1, v_2 \in V_{t_0}$ and for any $t \geq t_0$, if $v_1 \equiv v_2$ on $[t_0, t]$, then $\alpha(v_1) \equiv \alpha(v_2)$ on $[t_0, t + \tau]$.

We denote by $\mathcal{A}_d(t_0)$ (resp. $\mathcal{B}_d(t_0)$) the set of delay strategies for player 1 (resp. player 2).

Note that the delay τ depends on the delay strategy. Delay strategies are nonanticipative strategies, but the converse is false in general. For instance, if $\sigma : V \to U$ is Borel measurable, then the map

$$\alpha(v)_t = \sigma(v_t) \qquad \forall t \in [t_0, +\infty), \ \forall v \in V_{t_0}$$

is a nonanticipative strategy but not a delay strategy, unless σ is constant.

The key property of delay strategies is given in the following Lemma:

Lemma 1. *Let $\alpha \in \mathcal{A}(t_0)$ and $\beta \in \mathcal{B}(t_0)$. Assume that either α or β is a delay strategy. Then there is a unique pair of controls $(u, v) \in U_{t_0} \times V_{t_0}$ such that*

$$\alpha(v) \equiv u \text{ and } \beta(u) \equiv v \qquad \text{on}[t_0, +\infty). \tag{8.7}$$

Sketch of proof: Let us assume, to fix the ideas, that α is a delay strategy with delay τ. We first note that the restriction of $u := \alpha(v)$ to the interval $[t_0, t_0 + \tau]$ is independent of v because any two controls $v_1, v_2 \in V_{t_0}$ coincide almost everywhere on $[t_0, t_0]$. Then, as β is nonanticipative, the restriction of $v := \beta(u)$ for $[t_0, t_0 + \tau]$ is uniquely defined and does not depend on the values on u on $[t_0 + \tau, T]$.

Now, on the time interval $[t_0 + \tau, t_0 + 2\tau]$, the control $u := \alpha(v)$ depends only on the restriction of v to $[t_0, t_0 + \tau]$, which is uniquely defined as we just saw. This means that $\alpha(v)$ is uniquely defined $[t_0, t_0 + 2\tau]$. We can proceed this way by induction on the full interval $[t_0, +\infty)$. □

3.1.2 Definition of the Value Functions

Let $T > 0$ be the finite horizon of the game, i.e., the time at which the game ends.

Dynamics: For a fixed initial position $(t_0, x_0) \in [0, T] \times \mathbb{R}^d$ and any pair of controls $(u, v) \in U_{t_0} \times V_{t_0}$, we consider the differential equation

$$\begin{cases} \dot{X}_t = f(t, X_t, u_t, v_t) & t \in [t_0, T], \\ X_{t_0} = x_0. \end{cases} \tag{8.8}$$

In order to ensure the existence and the uniqueness of the solution, we suppose that f satisfies the following conditions:

$$\begin{cases} (i) \; U \text{ and } V \text{ are compact metric spaces,} \\ (ii) \; \text{the map } f : [0, T] \times \mathbb{R}^d \times U \times V \text{ is bounded and continuous} \\ (iii) \; f \text{ is uniformly Lipschitz continuous with respect to the space variable :} \\ \qquad |f(t, x, u, v) - f(t, y, u, v)| \le \text{Lip}(f)|x - y| \\ \qquad\qquad \forall (t, x, y, u, v) \in [0, T] \times \mathbb{R}^d \times \mathbb{R}^d \times U \times V \end{cases} \tag{8.9}$$

The Cauchy-Lipschitz theorem then asserts that Eq. (8.8) has a unique solution, denoted $X^{t_0, x_0, u, v}$.

Payoffs: The payoff of the players depends on a running payoff $\ell : [0, T] \times \mathbb{R}^d \times U \times V \to \mathbb{R}$ and on a terminal payoff $g : \mathbb{R}^d \to \mathbb{R}$. Namely, if the players play the controls $(u, v) \in U_{t_0} \times V_{t_0}$, then the cost the first player is trying to minimize (it is a payoff for the second player who is maximizing) is given by

$$\mathcal{J}(t_0, x_0, u, v) = \int_{t_0}^{T} \ell(s, X_s^{t_0, x_0, u, v}, u_s, v_s) ds + g(X_T^{t_0, x_0, u, v}).$$

Here are our assumptions on ℓ and g:

$$\begin{cases} (i) \; g : \mathbb{R}^d \to \mathbb{R} \text{ is bounded and Lipschitz continuous,} \\ (ii) \; \ell : [0, T] \times \mathbb{R}^d \times U \times V \to \mathbb{R} \text{ is continuous, bounded,} \\ \qquad \text{and Lipschitz continuous with respect to the } x \text{ variable.} \end{cases} \tag{8.10}$$

Normal form: Let $(\alpha, \beta) \in \mathcal{A}_d(t_0) \times \mathcal{B}_d(t_0)$ a pair of strategies for each player. Following Lemma 1, we know that there exists a unique pair of controls $(u, v) \in U_{t_0} \times V_{t_0}$ such that

$$\alpha(v) \equiv u \text{ and } \beta(u) \equiv v. \tag{8.11}$$

We will always use the notation (α_s, β_s) for (u_s, v_s) and $X_t^{t_0,x_0,\alpha,\beta}$ for $X_t^{t_0,x_0,u,v}$. The payoff associated to the two strategies $(\alpha, \beta) \in \mathcal{A}_d(t_0) \times \mathcal{B}_d(t_0)$ is given by

$$\mathcal{J}(t_0, x_0, \alpha, \beta) = \int_{t_0}^{T} \ell(s, X_s^{t_0,x_0,\alpha,\beta}, \alpha_s, \beta_s) ds + g(X_T^{t_0,x_0,\alpha,\beta}).$$

Definition 5 (Value functions). The **upper value function** is given by

$$\mathbf{V}^+(t_0, x_0) := \inf_{\alpha \in \mathcal{A}_d(t_0)} \sup_{\beta \in \mathcal{B}_d(t_0)} \mathcal{J}(t_0, x_0, \alpha, \beta) \tag{8.12}$$

while the **lower value function** is

$$\mathbf{V}^-(t_0, x_0) := \sup_{\beta \in \mathcal{B}_d(t_0)} \inf_{\alpha \in \mathcal{A}_d(t_0)} \mathcal{J}(t_0, x_0, \alpha, \beta). \tag{8.13}$$

Remark 1. We note for later use that the following equality holds:

$$\mathbf{V}^-(t_0, x_0) = \sup_{\beta \in \mathcal{B}_d(t_0)} \inf_{u \in U_{t_0}} \mathcal{J}(t_0, x_0, u, \beta(u)).$$

Indeed, inequality $\leq$ is clear because $U_{t_0} \subseteq \mathcal{A}_d(t_0)$. For inequality $\geq$, we note that, given $\beta \in \mathcal{B}_d(t_0)$ and $\alpha \in \mathcal{A}_d(t_0)$, there exists a unique pair $(u, v) \in U_{t_0} \times V_{t_0}$ such that (8.11) holds. Then $\mathcal{J}(t_0, x_0, \alpha, \beta) = \mathcal{J}(t_0, x_0, u, \beta(u))$. Of course the symmetric formula holds for $\mathbf{V}^+$.

The main result in zero-sum differential games is the existence of a value:

Theorem 3. *Assume that* (8.9) *and* (8.10) *hold, as well as Isaacs' condition:*

$$\inf_{u \in U} \sup_{v \in V} \{\langle p, f(t, x, u, v)\rangle + \ell(t, x, u, v)\} = \sup_{v \in V} \inf_{u \in U} \{\langle p, f(t, x, u, v)\rangle + \ell(t, x, u, v)\} \tag{8.14}$$

for any $(t, x, p) \in [0, T] \times \mathbb{R}^d \times \mathbb{R}^d$. *Then the game has a value:*

$$\mathbf{V}^+(t, x) = \mathbf{V}^-(t, x) \qquad \forall (t, x) \in [0, T] \times \mathbb{R}^d.$$

Remark 2. From the very definition of the value functions, one easily checks that $\mathbf{V}^- \leq \mathbf{V}^+$. So the key point is to prove the reverse inequality. The proof is not easy. It consists in showing that:

1. both value functions satisfy a *dynamic programming property* and enjoy some weak regularity property (continuity, for instance);
2. deduce from the previous fact that both value functions satisfy a partial differential equation – the HJI equation – in a weak sense (the viscosity sense);
3. then infer from the uniqueness of the solution of the HJI equation that the value functions are equal.

This method of proof also provides a characterization of the value function (see Theorem 6 below). In the rest of the section, we give some details on the above points and explain how they are related.

Before doing this, let us compare the value defined in nonanticipative strategies with delay and without delay:

Proposition 4 (see Bardi and Capuzzo Dolcetta (1996)). *Let* $\mathbf{V}^-$ *and* $\mathbf{V}^+$ *be the lower and upper value functions as in Definition 5. Then*

$$\mathbf{V}^-(t_0, x_0) = \inf_{\alpha \in \mathcal{A}(t_0)} \sup_{v \in V_{t_0}} \mathcal{J}(t_0, x_0, \alpha(v), v)$$

and

$$\mathbf{V}^+(t_0, x_0) = \sup_{\beta \in \mathcal{B}(t_0)} \inf_{u \in U_{t_0}} \mathcal{J}(t_0, x_0, u, \beta(u)).$$

At a first glance the result might seem surprising since, following Remark 1, $\mathbf{V}^-$ is equal to:

$$\mathbf{V}^-(t_0, x_0) = \sup_{\beta \in \mathcal{B}_d(t_0)} \inf_{\alpha \in \mathcal{A}_d(t_0)} \mathcal{J}(t_0, x_0, \alpha, \beta) = \sup_{\beta \in \mathcal{B}_d(t_0)} \inf_{u \in U_{t_0}} \mathcal{J}(t_0, x_0, u, \beta(u)).$$

One sees here the difference between a nonanticipative strategy, which allows to synchronize exactly with the current value of the opponent's control and the delay strategies, where one cannot immediately react to this control: in the first case, one has an informational advantage, but not in the second one.

The proof of the proposition involves a dynamic programming for the value functions defined in the proposition in terms of nonanticipative strategies and derive from this that they satisfy the same HJI equation than $\mathbf{V}^+$ and $\mathbf{V}^-$. Whence the equality.

3.2 The Bolza Problem: Dynamic Programming

The dynamic programming expresses the fact that if one stops the game at an intermediate time, then one does not lose anything by restarting it afresh with the only knowledge of the position reached so far.

Theorem 4 (Dynamic programming). *Let* $(t_0, x_0) \in [0, T) \times \mathbb{R}^d$ *and* $h \in (0, T - t_0)$*. Then*

$$\mathbf{V}^+(t_0, x_0) = \inf_{\alpha \in \mathcal{A}_d(t_0)} \sup_{v \in V_{t_0}} \left\{ \int_{t_0}^{t_0+h} \ell(s, X_s^{t_0,x_0,\alpha(v),v}, \alpha(v)_s, v_s) ds \right.$$
$$\left. + \mathbf{V}^+(t_0 + h, X_{t_0+h}^{t_0,x_0,\alpha(v),v}) \right\} \qquad (8.15)$$

while

$$\mathbf{V}^-(t_0, x_0) = \sup_{\beta \in \mathcal{B}_d(t_0)} \inf_{u \in U_{t_0}} \left\{ \int_{t_0}^{t_0+h} \ell(s, X_s^{t_0,x_0,u,\beta(u)}, u_s, \beta(u)_s) ds \right.$$
$$\left. + \mathbf{V}^-(t_0 + h, X_{t_0+h}^{t_0,x_0,u,\beta(u)}) \right\}.$$

Sketch of proof: Following Remark 1, we have

$$\mathbf{V}^+(t_0, x_0) = \inf_{\alpha \in \mathcal{A}_d(t_0)} \sup_{v \in V_{t_0}} \left\{ \int_{t_0}^{T} \ell(X_s^{t_0,x_0,\alpha(v),v}, \alpha(v)_s, v_s) ds + g(X_T^{t_0,x_0,\alpha(v),v}) \right\}$$
$$= \inf_{\alpha \in \mathcal{A}_d(t_0)} \sup_{v \in V_{t_0}} \left\{ \int_{t_0}^{t_0+h} \ell(X_s^{t_0,x_0,\alpha(v),v}, \alpha(v_s), v_s) ds \right.$$
$$\left. + \int_{t_0+h}^{T} \ell(X_s^{t_0,x_0,\alpha(v),v}, \alpha(v)_s, v_s) ds + g(X_T^{t_0,x_0,\alpha(v),v}) \right\}.$$

By the semigroup property, one has, for $s \geq t_0 + h$ and any control pair $(u, v) \in U_{t_0} \times V_{t_0}$,

$$X_s^{t_0,x_0,u,v} = X_s^{t_0+h, X^{t_0,x_0,u,v}, \tilde{u}, \tilde{v}}$$

where $\tilde{u}$ (respectively $\tilde{v}$) is the restriction to $[t_0 + h, T]$ of u (respectively v). Hence, the above equality can be rewritten (loosely speaking) as

$$\mathbf{V}^+(t_0,x_0) = \inf_{\alpha\in\mathcal{A}_d(t_0)} \sup_{v\in V_{t_0}} \int_{t_0}^{t_0+h} \ell(X_s^{t_0,x_0,\alpha(v),v},\alpha(v)_s,v_s)ds$$

$$+ \inf_{\tilde{\alpha}\in\mathcal{A}_d(t_0+h)} \sup_{\tilde{v}\in V_{t_0+h}} \left\{ \int_{t_0+h}^{T} \ell(X_s^{t_0,X_{t_0+h}^{t_0,x_0,\alpha(v),v},\tilde{\alpha}(v),\tilde{v}},\tilde{\alpha}(v)_s,\tilde{v}_s)ds \right.$$

$$\left. + g(X_T^{t_0,X_{t_0+h}^{t_0,x_0,\alpha(v),v},\tilde{\alpha}(v),\tilde{v}}) \right\}$$

$$= \inf_{\alpha\in\mathcal{A}_d(t_0)} \sup_{v\in V_{t_0}} \left\{ \int_{t_0}^{t_0+h} \ell(X_s^{t_0,x_0,\alpha(v),v},\alpha(v)_s,v_s)ds + \mathbf{V}^+(t_0+h,X_{t_0+h}^{t_0,x_0,\alpha(v),v}) \right\}.$$

□

One can also show that the value functions enjoy some space-time regularity:

Proposition 5. *The maps* $\mathbf{V}^+$ *and* $\mathbf{V}^-$ *are bounded and Lipschitz continuous in all variables.*

The Lipschitz regularity in space relies on similar property of the flow of the differential equation when one translates the space. The time regularity is more tricky and uses the dynamic programming principle.

The regularity described in the proposition is quite sharp: in general, the value function has singularities and cannot be of class $\mathcal{C}^1$.

3.3 The Bolza Problem: HBI Equation and Viscosity Solutions

We now explain, in a purely heuristic way, how the dynamic programming property is related with a partial differential equation called Hamilton-Jacobi-Isaacs' equation (HJI equation). We work with $\mathbf{V}^+$. The dynamic programming (8.15) can be rewritten as

$$\inf_{\alpha\in\mathcal{A}_d(t_0)} \sup_{v\in V_{t_0}} \left\{ \frac{1}{h}\int_{t_0}^{t_0+h} \ell(s,X_s^{\alpha(v),v},\alpha(v)_s,v_s)ds \right.$$
$$\left. + \frac{\mathbf{V}^+(t_0+h,X_{t_0+h}^{\alpha,v}) - \mathbf{V}^+(t_0,x_0)}{h} \right\} = 0 \tag{8.16}$$

where we have used the notation $X_t^{\alpha,v} = X_t^{t_0,x_0,\alpha(v),v}$. As h tends to 0^+, $\frac{X_{t_0+h}^{\alpha,v}-x_0}{h}$ behaves as $f(t_0,x_0,\alpha(v)_{t_0},v_{t_0})$. Hence, $\frac{\mathbf{V}^+(t_0+h,X_{t_0+h}^{\alpha,v})-\mathbf{V}^+(t_0,x_0)}{h}$ is close to

$$\partial_t \mathbf{V}^+(t_0, x_0) + \langle D\mathbf{V}^+, f(t_0, x_0, \alpha(v)_{t_0}, v_{t_0})\rangle.$$

Moreover, $\frac{1}{h}\int_{t_0}^{t_0+h} \ell(s, X_s^{\alpha,\beta}, \alpha_s, v_s)ds$ "behaves" as $\ell(t_0, x_0, \alpha_{t_0}, v_{t_0})$. Therefore, equality (8.16) becomes

$$\inf_{u\in U}\sup_{v\in V}\left\{\ell(t_0, x_0, u, v) + \partial_t \mathbf{V}^+(t_0, x_0) + \langle D\mathbf{V}^+, f(t_0, x_0, u, v)\rangle\right\} = 0$$

(the difficult part of the proof is to justify that one can pass from the infimum over strategies to the infimum over the sets U and V). If we set, for $(t, x, p) \in [0, T] \times \mathbb{R}^d \times \mathbb{R}^d$,

$$H^+(t, x, p) := \inf_{u\in U}\sup_{v\in V}\{\langle p, f(t, x, u, v)\rangle + \ell(t, x, u, v)\}\ , \tag{8.17}$$

the map $\mathbf{V}^+$ should satisfy the HJI equation

$$\begin{cases} -\partial_t \mathbf{V}^+(t, x) - H^+(t, x, D\mathbf{V}^+(t, x)) = 0 & \text{in } (0, T) \times \mathbb{R}^d, \\ \mathbf{V}^+(T, x) = g(x) & \text{in } \mathbb{R}^d \end{cases} \tag{8.18}$$

(The sign convention is discussed below, when we develop the notion of viscosity solutions.) Applying the similar arguments for $\mathbf{V}^-$, we obtain that $\mathbf{V}^-$ should satisfy the symmetric equation

$$\begin{cases} -\partial_t \mathbf{V}^-(t, x) - H^-(t, x, D\mathbf{V}^-(t, x)) = 0 & \text{in } (0, T) \times \mathbb{R}^d, \\ \mathbf{V}^-(T, x) = g(x) & \text{in } \mathbb{R}^d, \end{cases} \tag{8.19}$$

where

$$H^-(t, x, p) := \sup_{v\in V}\inf_{u\in U}\{\langle p, f(t, x, u, v)\rangle + \ell(t, x, u, v)\}\,. \tag{8.20}$$

Now if Isaacs' condition holds, i.e., if $H^+ = H^-$, then $\mathbf{V}^+$ and $\mathbf{V}^-$ satisfy the same equation, and one can hope that this implies the equality $\mathbf{V}^+ = \mathbf{V}^-$. This is indeed the case, but we have to be careful with the sense we give to Eqs. (8.18) and (8.19).

Let us recall that, since $\mathbf{V}^+$ is Lipschitz continuous, Rademacher's theorem states that $\mathbf{V}^+$ is differentiable almost everywhere. In fact one can show that $\mathbf{V}^+$ indeed satisfies Eq. (8.18) at each point of differentiability. Unfortunately, this is not enough to characterize the value functions. For instance, one can show the existence of infinitely many Lipschitz continuous functions satisfying almost everywhere an equation of the form (8.18).

The idea of "viscosity solutions" is that one should look closely even at points where the function is not differentiable.

We now explain the proper meaning for equations of the form:

$$-\partial_t \mathbf{V}(t,x) - H(t,x,D\mathbf{V}(t,x)) = 0 \qquad \text{in } (0,T)\times\mathbb{R}^d \tag{8.21}$$

where $H : [0,T]\times\mathbb{R}^d\times\mathbb{R}^d \to \mathbb{R}$ is continuous. Let us point out that this equation is backward in time, meaning that the natural condition is that $\mathbf{V}$ is given at time T.

Definition 6 (Viscosity solution).

- A map $\mathbf{V} : [0,T]\times\mathbb{R}^d \to \mathbb{R}$ is a **viscosity supersolution** of (8.21) if $\mathbf{V}$ is lower semicontinuous (l.s.c.) in $(0,T)\times\mathbb{R}^d$ and if, for any test function $\phi \in \mathcal{C}^1([0,T]\times\mathbb{R}^d)$ such that $\mathbf{V}-\phi$ has a local minimum at some point $(t,x)\in(0,T)\times\mathbb{R}^d$, one has

$$-\partial_t\phi(t,x) - H(t,x,D\phi(t,x)) \geq 0.$$

- A map $\mathbf{V} : [0,T]\times\mathbb{R}^d \to \mathbb{R}$ is a **viscosity subsolution** of (8.21) if $\mathbf{V}$ is upper semicontinuous (u.s.c.) and if, for any test function $\phi \in \mathcal{C}^1([0,T]\times\mathbb{R}^d)$ such that $\mathbf{V}-\phi$ has a local maximum at some point $(t,x)\in(0,T)\times\mathbb{R}^d$, one has

$$-\partial_t\phi(t,x) - H(t,x,D\phi(t,x)) \leq 0.$$

- A **viscosity solution** to (8.21) is a map $\mathbf{V}$ which is a viscosity sub- and supersolution to (8.21).

Note that, with this definition, a solution is a continuous map. One can easily check that, if $\mathbf{V} \in \mathcal{C}^1([0,T]\times\mathbb{R}^d)$, then $\mathbf{V}$ is a supersolution (resp. subsolution) of (8.21) if and only if, for any $(t,x)\in(0,T)\times\mathbb{R}^d$,

$$-\partial_t\mathbf{V}(t,x) - H(t,x,D\mathbf{V}(t,x)) \geq 0 \qquad (\text{resp.} \leq 0).$$

Finally, note the sign convention: the equations are written in such a way that supersolutions satisfy the inequality with ≥ 0.

The main point in considering viscosity solution is the comparison principle, which implies that Eq. (8.21), supplemented with a terminal condition, has at most one solution. For this we assume that H satisfies the following conditions :

$$|H(t_1,x_1,p) - H(t_2,x_2,p)| \leq C(1+|p|)|(t_1,x_1)-(t_2,x_2)| \tag{8.22}$$

and

$$|H(t,x,p_1) - H(t,x,p_2)| \leq C|p_1-p_2| \tag{8.23}$$

for some constant C. Note that the Hamiltonians $H^{\pm}$ defined by (8.17) and (8.20) satisfy the above assumptions.

Theorem 5 (Comparison principle, Crandall et al. (1992)). *Under assumptions* (8.22) *and* (8.23)*, let* $\mathbf{V}_1$ *be a subsolution of* (8.21) *which is u.s.c. on* $[0,T]\times\mathbb{R}^d$ *and* $\mathbf{V}_2$ *be a supersolution of* (8.21) *which is l.s.c. on* $[0,T]\times\mathbb{R}^d$*. Let us assume that* $\mathbf{V}_1(T,x)\leq\mathbf{V}_2(T,x)$ *for any* $x\in\mathbb{R}^d$*. Then*

$$\mathbf{V}_1(t,x)\leq\mathbf{V}_2(t,x)\qquad\forall(t,x)\in[0,T]\times\mathbb{R}^d.$$

From this result one easily deduces:

Corollary 1. *Let* $g:\mathbb{R}^d\to\mathbb{R}$ *be continuous. Then Eq.* (8.21) *has at most one continuous viscosity solution which satisfies the terminal condition* $\mathbf{V}(T,x)=g(x)$ *for any* $x\in\mathbb{R}^d$*.*

Proof. Let $\mathbf{V}_1$ and $\mathbf{V}_2$ be two bounded and Lipschitz continuous viscosity solutions of (8.21) such that $\mathbf{V}_1(T,x)=\mathbf{V}_2(T,x)=g(x)$ for any $x\in\mathbb{R}^d$. Since, in particular, $\mathbf{V}_1$ is a subsolution and $\mathbf{V}_2$ a supersolution and $\mathbf{V}_1(T,\cdot)=\mathbf{V}_2(T,\cdot)$, we have by comparison $\mathbf{V}_1\leq\mathbf{V}_2$ in $[0,T]\times\mathbb{R}^d$. Reversing the roles of $\mathbf{V}_1$ and $\mathbf{V}_2$, one gets the opposite inequality, whence the equality. □

3.3.1 Existence and Characterization of the Value

We are now ready to state the main result of the section:

Theorem 6. *Under conditions* (8.9) *and* (8.10) *on* f*,* ℓ*, and* g *and if Isaacs' assumption holds,*

$$H^+(t,x,p)=H^-(t,x,p)\qquad\forall(t,x,p)\in[0,T]\times\mathbb{R}^d\times\mathbb{R}^d\ ,\tag{8.24}$$

the game has a value:

$$\mathbf{V}^+(t,x)=\mathbf{V}^-(t,x)\qquad\forall(t,x)\in[0,T]\times\mathbb{R}^d.$$

Moreover, $\mathbf{V}^+=\mathbf{V}^-$ *is the unique viscosity solution of Isaacs' equation* (8.18) = (8.19).

The key point of the proof of Theorem 6 is the following (half-)characterization of the value functions:

Lemma 2. *The upper and lower value functions* $\mathbf{V}^+$ *and* $\mathbf{V}^-$ *are respectively viscosity solutions of equation* (8.18) *and of* (8.19).

The proof, which is a little intricate, follows more or less the formal argument described in Sect. 3.3. Let us underline at this point that if Isaacs' condition does not hold, there is no reason in general that the upper value function and the lower one coincide. Actually they do not in general, in the sense that there exists a finite horizon $T > 0$ and a terminal condition g for which $\mathbf{V}^- \neq \mathbf{V}^+$. However, the statement is misleading: in fact the game can have a value, if we allow the players to play random strategies.

3.4 Without Isaacs' Condition

We just discussed the existence of a value for zero-sum differential games when the players play deterministic strategies. This result holds under Isaacs' condition, which expresses the fact that the infinitesimal game has a value.

We discuss here the existence of a value when Isaacs' condition does not hold. In this case, simple counterexamples show that there is no value in pure strategies. However, it is not difficult to guess that the game should have a value in random strategies, exactly as it is the case in classical game theory. Moreover, the game should have a dynamics and payoff in which the players randomize at each time their control. The corresponding Hamiltonian is then expected to be the one where the saddle point is defined over the sets of probabilities on the control sets, instead of the control sets themselves.

But giving a precise meaning to this statement turns out to be tricky. The reason is the continuous time: indeed, if it is easy to build a sequence of independent random variables with a given law, it is impossible to build a continuum of random variables which are at each time independent. So one has to discretize the game. The problem is then to prevent the player who knows his opponent's strategy to coordinate upon this strategy. The solution to this issue for games in positional strategies goes back to Krasovskii and Subbotin (1988). We follow here the presentation for games in nonanticipative strategies given by Buckdahn, Quincampoix, Rainer, and Xu ('15).

3.4.1 Statement of the Problem

We consider a differential game with deterministic dynamics of the form

$$\dot{X}_t = f(t, X_t, u_t, v_t),$$

and we denote by $(X_t^{t_0,x_0,u,v})_{t\geq t_0}$ the solution of this ODE with initial condition $X_{t_0} = x_0$. The cost of the first player is defined as usual by

$$\mathcal{J}(t_0, x_0, u, v) := \int_{t_0}^{T} \ell(t, X_t^{t_0,x_0,u,v}, u_t, v_t)\, dt + g(X_T^{t_0,x_0,u,v})$$

where the map u (resp. v) belongs to the set $\mathcal{U}_{t_0}$ of measurable controls with values in U (resp. $\mathcal{V}_{t_0}$ with values in V).

We now define the notion of random strategies. For this let us first introduce the probability space

$$(\Omega, \mathcal{F}, \mathbb{P}) = ([0,1], \mathcal{B}([0,1]), dx)$$

where $\mathcal{B}([0,1])$ is the Borel $\sigma-$field on $[0,1]$ and dx is the Lebesgue measure. Then, in contrast to the notion of strategies with delay in case where Isaacs' assumption holds, we have to fix a partition $\pi = \{0 = t_0 < t_1 < \cdots < t_N = T\}$, which is common to both players.

The idea developed here is to consider random nonanticipative strategies with delay, for which the randomness is reduced to the dependence on a finite number of independent random variables $(\zeta_{j,l})_{j=1,2,l\geq 1}$ which all have the same uniform law on $[0,1]$: one variable for each player on each interval $[t_l, t_{l+1})$:

Definition 7 (Random delay strategies along the partition π). Fix an arbitrary initial time $t \in [t_k, t_{k+1})$ for some $k \in \{0, \ldots, N-1\}$. A random delay strategy along the partition π for player 1 is a mapping $\alpha : \Omega \times V_t \to U_t$, such that, for each $(\omega, v) \in \Omega \times V_t$, on each time interval $[t_l, t_{l+1})$ with $k \leq l \leq N-1$, $\alpha(\omega, v)$ depends on ω only through the $l - k + 1$ first random variables $\zeta_k, \ldots, \zeta_l$ and on v through its restriction to $[t, t_{l-1})$. The set of random delay strategies along π for player 1 is denoted by $\mathcal{A}_r^\pi(t)$.

Random delay strategies for player 2 are defined in a similar way, and the set of such strategies is denoted by $\mathcal{B}_r^\pi(t)$.

The game can be put into normal form: Let $(\alpha, \beta) \in \mathcal{A}_r^\pi(t) \times \mathcal{B}_r^\pi(t)$. Then one can show that, for any $\omega \in \Omega$, there exists a unique pair of controls $(u_\omega, v_\omega) \in U_t \times V_t$ such that

$$\alpha(\omega, v_\omega) = u_\omega \qquad \text{and} \qquad \beta(\omega, u_\omega) = v_\omega \qquad \text{a.e.} \tag{8.25}$$

We may then extend the payoff to the set of strategies:

$$\mathcal{J}(t, x, \alpha, \beta) := \mathbb{E}\left[\int_t^T \ell(s, X_s^{t,x,u_\omega,v_\omega}, u_{\omega,s}, v_{\omega,s})\, ds + g(X_T^{t,x,u_\omega,v_\omega})\right]$$

where the pair (u_ω, v_ω) is defined by (8.25).

Finally, to each partition π can be associated an upper and a lower value functions:

$$\mathbf{V}^{\pi,+}(t, x) := \inf_{\alpha \in \mathcal{A}_r^\pi(t)} \sup_{\beta \in \mathcal{B}_r^\pi(t)} \mathcal{J}(t, x, \alpha, \beta)$$

and

$$\mathbf{V}^{\pi,-}(t_0, x_0) := \sup_{\beta \in \mathcal{B}_r^\pi(t)} \inf_{\alpha \in \mathcal{A}_r^\pi(t)} \mathcal{J}(t, x, \alpha, \beta).$$

3.4.2 Existence and Characterization of the Value

We now assume that conditions (8.9) and (8.10) on the dynamics and the payoff hold. Let $\Delta(U)$ (resp. $\Delta(V)$) denote the set of Borel probabilities on U (resp. V), and let us set

$$\begin{aligned} H(t,x,p) &:= \inf_{\mu\in\Delta(U)} \sup_{\nu\in\Delta(V)} \int_{U\times V} (\langle f(t,x,u,v), p\rangle + \ell(t,x,u,v))\ d\mu(u)d\nu(v) \\ &= \sup_{\nu\in\Delta(V)} \inf_{\mu\in\Delta(U)} \int_{U\times V} (\langle f(t,x,u,v), p\rangle + \ell(t,x,u,v))\ d\mu(u)d\nu(v). \end{aligned}$$

Remark that there is no need to suppose some supplementary Isaacs' condition here: the equality simply holds, thanks to the min-max theorem.

Theorem 7. *For all sequences of partitions* (π_n) *with* $|\pi_n| \to 0$, *the sequences* $(\mathbf{V}^{\pi_n,+})$ *and* $(\mathbf{V}^{\pi_n,-})$ *converge uniformly on compact sets to a same Lipschitz continuous function* $\mathbf{V}$, *which is the unique solution of the HJ equation*

$$\begin{cases} -\partial_t \mathbf{V}(t,x) - H(t,x,D\mathbf{V}(t,x)) = 0 & \text{in } (0,T)\times\mathbb{R}^d, \\ \mathbf{V}(T,x) = g(x) & \text{in } \mathbb{R}^d. \end{cases}$$

3.5 The Infinite Horizon Problem

The approach for the Bolza problem can be extended to many other classes of differential games. We concentrate here on the infinite horizon problem, for which the associate Hamilton-Jacobi-Isaacs equation is stationary.

3.5.1 Description of the Game

Dynamics: For a fixed initial position $x_0 \in \mathbb{R}^d$, we consider the differential equation

$$\begin{cases} \dot{X}_t = f(X_t, u_t, v_t) & t \in [0,+\infty) \\ X_0 = x_0 \end{cases} \tag{8.26}$$

where $f : \mathbb{R}^d \times U \times V \to \mathbb{R}^d$ satisfies the usual assumptions (8.9). The controls of player 1 and player 2 are now Lebesgue measurable maps $u : [0,+\infty) \to U$ and $v : [0,+\infty) \to V$. For any pair $(u,v) \in \mathcal{U}_0 \times \mathcal{V}_0$, Eq. (8.26) has a unique solution, denoted $X^{x_0,u,v}$.

Payoffs: The payoff of the players depends on a discount rate $\lambda > 0$ and on a running payoff $\ell : \mathbb{R}^d \times U \times V \to \mathbb{R}$. Namely, if the players play the controls $(u,v) \in \mathcal{U}_0 \times \mathcal{V}_0$, then the first player is trying to minimize his cost:

$$\mathcal{J}(x_0, u, v) = \int_0^{+\infty} e^{-\lambda s} \ell(X_s^{x_0,u,v}, u_s, v_s) ds \ ,$$

where ℓ is assumed to satisfy (8.10).

Strategies: As for the Bolza problem, we consider delay strategies: see Definition 4. Delay strategies for player 1 (resp. player 2) are denoted by $\mathcal{A}_d := \mathcal{A}_d(0)$ (resp. $\mathcal{B}_d := \mathcal{B}_d(0)$). We systematically use the fact that if $(\alpha, \beta) \in \mathcal{A}_d \times \mathcal{B}_d$ is a pair of strategies, then there is a unique pair of controls $(u, v) \in U_0 \times V_0$ such that

$$\alpha(v) = u \text{ and } \beta(u) = v \qquad \text{a.e. in } [0, +\infty). \tag{8.27}$$

In particular we always use the notation (α_s, β_s) for (u_s, v_s) and $X_t^{x_0,\alpha,\beta}$ for $X_t^{x_0,u,v}$, where (u_s, v_s) is defined by (8.27). The payoff associated to the two strategies $(\alpha, \beta) \in \mathcal{A}_d \times \mathcal{B}_d$ is given by

$$\mathcal{J}(x_0, \alpha, \beta) = \int_0^{+\infty} e^{-\lambda s} \ell(X_s^{x_0,\alpha,\beta}, \alpha_s, \beta_s) ds.$$

Definition 8 (Value functions). The **upper value function** is given by

$$\mathbf{V}^+(x_0) := \inf_{\alpha \in \mathcal{A}_d} \sup_{\beta \in \mathcal{B}_d} \mathcal{J}(x_0, \alpha, \beta) \tag{8.28}$$

while the **lower value function** is

$$\mathbf{V}^-(x_0) := \sup_{\beta \in \mathcal{B}_d} \inf_{\alpha \in \mathcal{A}_d} \mathcal{J}(x_0, \alpha, \beta). \tag{8.29}$$

Note that, in contrast with the Bolza problem, the value functions only depend on the space variable: the idea is that, because the dynamics and the running cost are independent of time and because the discount is of exponential type, it is no longer necessary to include time in the definition of the value functions to obtain a dynamic programming principle. Indeed, in the case of the upper value, for instance, one can show a dynamic programming principle of the following form: for any $h \geq 0$,

$$\mathbf{V}^+(x_0) = \inf_{\alpha \in \mathcal{A}_d} \sup_{\beta \in \mathcal{B}_d} \left\{ \int_0^h \ell(X_s^{x_0,\alpha,\beta}, \alpha_s, \beta_s) ds + e^{-\lambda h} \mathbf{V}^+ \left(X_h^{x_0,\alpha,\beta} \right) \right\} .$$

It is then clear that the associated Hamilton-Jacobi equation does not involve a time derivative of the value function. Let us now explicitly write this Hamilton-Jacobi equation.

3.5.2 Existence and Characterization of the Value

Because of the dynamic programming given above, the value functions are expected to satisfy the following Hamilton-Jacobi equations:

$$\lambda \mathbf{V}(t,x) - H^+(x, D\mathbf{V}(t,x)) = 0 \text{ in } \mathbb{R}^d \tag{8.30}$$

where H^+ is defined by

$$H^+(x,p) = \inf_{u\in U} \sup_{v\in V} \{\langle p, f(x,u,v)\rangle + \ell(x,u,v)\} \ , \tag{8.31}$$

and

$$\lambda \mathbf{V}(t,x) - H^-(x, D\mathbf{V}(t,x)) = 0 \text{ in } \mathbb{R}^d \tag{8.32}$$

where H^- is defined by

$$H^-(x,p) = \sup_{v\in V} \inf_{u\in U} \{\langle p, f(x,u,v)\rangle + \ell(x,u,v)\} \, . \tag{8.33}$$

Theorem 8. *Under the above conditions on f and ℓ and if Isaacs' assumption holds,*

$$H^+(x,p) = H^-(x,p) \qquad \forall (x,p) \in \mathbb{R}^d \times \mathbb{R}^d \ , \tag{8.34}$$

then the game has a value: $\mathbf{V}^+ = \mathbf{V}^-$, which is the unique bounded viscosity solution of Isaacs' equation (8.30) = (8.32).

Requiring a growth condition on the solution is mandatory for the uniqueness. For instance, in dimension $d = 1$, the equation $W + W_x = 0$ has a unique bounded solution: $W = 0$. However, $W(x) = ce^{-x}$ is of course a solution for any constant c.

Without Isaacs' assumption, the game has a value, provided one defines this value in terms of random strategies as in Sect. 3.4.

As for the Bolza problem, the proof is based on a dynamic programming property and on the regularity of the value functions.

4 Stochastic Differential Games

Several approaches have been developed to handle stochastic differential games (SDG). Indeed, for the same result – the existence and the characterization of the value – there exists not only various techniques of proofs, but even very different ways to define the actions of the players and therefore the values of the game. We discuss three of them: by nonanticipative strategies, by pathwise strategies, and by backward differential equations.

Even if, in each paragraph, we have to wait for the PDE characterization to be able to claim that the different definitions of the upper and lower value coincide, we denote them, from the start, by the same letters $\mathbf{V}^+$ and $\mathbf{V}^-$.

For simplicity, we restrict the presentation to the Bolza problem, the extension to the infinite horizon problem being straightforward.

4.1 The Value Functions

First, we present the seminal approach by Fleming and Souganidis (1989).

Dynamics: For a fixed initial position $(t_0, x_0) \in [0, T] \times \mathbb{R}^d$, we consider the stochastic differential equation

$$\begin{cases} dX_t = b(t, X_t, u_t, v_t)dt + \sigma(t, X_t, u_t, v_t)dB_t, & t \in [t_0, T], \\ X_{t_0} = x_0 \ . \end{cases} \tag{8.35}$$

where, for some $M \geq 1$, $(B_t)_{t\in[0,T]}$ is a standard M-dimensional Brownian motion defined on the canonical Wiener space. We denote by $(\mathcal{F}_t)_{t\in[0,T]}$ the filtration generated by $(B_t)_{t\in[0,T]}$ (completed by all sets of zero probability).

The assumptions on the control sets and the parameters of the system are the direct generalization of those required for the deterministic setting:

$$\begin{cases} (i) \ U \text{ and } V \text{ are compact metric spaces,} \\ (ii) \text{ the maps } b : [0, T] \times \mathbb{R}^d \times U \times V \to \mathbb{R}^d \text{ and } \sigma : [0, T] \times \mathbb{R}^d \times U \times V \to \\ \qquad \mathbb{R}^{d\times M} \text{ are bounded and continuous in all their variables} \\ (iii) \ b \text{ and } \sigma \text{ are uniformly Lipschitz continuous in the space variable.} \end{cases} \tag{8.36}$$

The controls of player 1 and player 2 are measurable maps $u : [t_0, T] \times \Omega \to U$ and $v : [t_0, T] \times \Omega \to V$ adapted to the filtration $(\mathcal{F}_t)$. This last assumption translate the intuitive idea that both players observe the Brownian motion, but their controls may only depend on its past. The sets of such controls are denoted by $\mathcal{U}(t_0)$ and $\mathcal{V}(t_0)$.

Under assumptions (8.36), it is well known that for any pair $(u, v) \in \mathcal{U}(t_0) \times \mathcal{V}(t_0)$, Eq. (8.35) has a unique solution, denoted by $X^{t_0,x_0,u,v}$.

Payoff: To any pair of controls $(u, v) \in \mathcal{U}(t_0) \times \mathcal{V}(t_0)$, we associate the payoff

$$\mathcal{J}(t_0, x_0, u, v) = \mathbb{E}\left[\int_{t_0}^{T} \ell(s, X_s^{t_0,x_0,u,v}, u_s, v_s)ds + g(X_T^{t_0,x_0,u,v})\right],$$

where we assume that ℓ and g satisfy the standard conditions (8.10).
Strategies: A **nonanticipative strategy** for player 1 is a mapping $\alpha : \mathcal{V}(t_0)$ to $\mathcal{U}(t_0)$ such that, if, with probability 1, two controls v and v' coincide Lebesgue-a.s. on a

time interval $[t_0, t]$, then the same holds for the controls $\alpha(v)$ and $\alpha(v')$. Strategies for player 2 are defined in a symmetric way. This assumption is coherent with the idea that the answer to a control of the opponent depends only on the past. We denote by $\mathcal{A}(t_0)$ the set of nonanticipative strategies for player 1 (resp. $\mathcal{B}(t_0)$ the set of nonanticipative strategies for player 2).

Definition 9 (Value functions). The **upper** and **lower value functions** are given by

$$\mathbf{V}^+(t_0, x_0) = \sup_{\beta \in \mathcal{B}(t_0)} \inf_{u \in \mathcal{U}(t_0)} \mathcal{J}(t_0, x_0, u, \beta(u)),$$

and

$$\mathbf{V}^-(t_0, x_0) = \inf_{\alpha \in \mathcal{A}(t_0)} \sup_{v \in \mathcal{V}(t_0)} \mathcal{J}(t_0, x_0, \alpha(v), v).$$

The main result in this section is that, under a suitable Isaacs' condition, the game has a value which can be characterized as the unique viscosity solution of an HJI equation *of second order*.

4.2 Viscosity Solutions of Second-Order HJ Equations

The appearance of a random term in the dynamics of the game corresponds to the addition of a second-order term in the functional equation which characterizes the value function. The main reference for the theory of second-order Hamilton-Jacobi equations is the User's Guide of Crandall-Ishii-Lions (1992). Let us summarize what we need in the framework of SDG.

We are concerned with the following (backward in time) Hamilton-Jacobi equation

$$-\partial_t \mathbf{V}(t,x) - H(t, x, D\mathbf{V}(t,x), D^2\mathbf{V}(t,x)) = 0 \ \text{ in } (0,T) \times \mathbb{R}^d, \tag{8.37}$$

where $H : [0,T] \times \mathbb{R}^d \times \mathbb{R}^d \times \mathcal{S}_d \to \mathbb{R}$ is a continuous Hamiltonian, which satisfies the following monotonicity assumption:

$$\begin{aligned} &\text{for all } (t,x,p) \in [0,T] \times \mathbb{R}^d \times \mathbb{R}^d \text{ , for all } A, B \in \mathcal{S}_d, \\ &A \le B \Rightarrow H(t,x,p,A) \le H(t,x,p,B) \end{aligned} \tag{8.38}$$

($\mathcal{S}_d$ denotes the set of symmetric $d \times d$-matrices).

For instance, we are interested in the following Hamiltonians:

$$H^-(t,x,p,A) = \sup_{v\in V}\inf_{u\in U}\Big\{\langle p, b(t,x,u,v)\rangle + \frac{1}{2}\mathrm{tr}[\sigma(t,x,u,v)\sigma^*(t,x,u,v)A] + \ell(t,x,u,v)\Big\},$$

and

$$H^+(t,x,p,A) = \inf_{u\in U}\sup_{v\in V}\Big\{\langle p, b(t,x,u,v)\rangle + \frac{1}{2}\mathrm{tr}[\sigma(t,x,u,v)\sigma^*(t,x,u,v)A] + \ell(t,x,u,v)\Big\}.$$

It is easy to see that H^+ and H^- both satisfy assumption (8.38).

Definition 10.

- A map $\mathbf{V} : [0,] \times \mathbb{R}^d \to \mathbb{R}$ is a **viscosity supersolution** of (8.37) if $\mathbf{V}$ is lower semicontinuous (l.s.c.) in $(0,T)\times\mathbb{R}^d$ and if, for any test function $\phi \in \mathcal{C}^{1,2}([0,T]\times\mathbb{R}^d)$ such that $\mathbf{V}-\phi$ has a local minimum at some point $(t,x)\in(0,T)\times\mathbb{R}^d$, one has

$$-\partial_t\phi(t,x) - H(t,x,D\phi(t,x),D^2\phi(t,x)) \geq 0.$$

- A map $\mathbf{V} : [0,T] \times \mathbb{R}^d \to \mathbb{R}$ is a **viscosity subsolution** of (8.37) if $\mathbf{V}$ is upper semicontinuous (u.s.c.) in $(0,T)\times\mathbb{R}^d$ and if, for any test function $\phi \in \mathcal{C}^{1,2}([0,T]\times\mathbb{R}^d)$ such that $\mathbf{V}-\phi$ has a local maximum at some point $(t,x)\in(0,T)\times\mathbb{R}^d$, one has

$$-\partial_t\phi(t,x) - H(t,x,D\phi(t,x),D^2\phi(t,x)) \leq 0.$$

- A **viscosity solution** of (8.37) is a map $\mathbf{V}$ which is both a sub- and a supersolution of (8.37).

Theorem 9 (Comparison principle, Crandall et al. (1992)). *Suppose that the Hamiltonian H is given by H^+ or H^- defined above. Let $\mathbf{V}_1$ (resp. $\mathbf{V}_2$) be a subsolution (resp. a supersolution) of (8.37). If, for all $x\in\mathbb{R}^d$, $\mathbf{V}_1(T,x)\leq\mathbf{V}_2(T,x)$, then $\mathbf{V}_1\leq\mathbf{V}_2$ on $[0,T]\times\mathbb{R}^d$.*

4.3 Existence and Characterization of the Value

The key point is the following statement:

Proposition 6. *The lower value function* $\mathbf{V}^-$ *is a viscosity subsolution of the HJI equation*

$$\begin{cases} -\partial_t \mathbf{V} - H^-(t,x,D\mathbf{V},D^2\mathbf{V}) = 0, & (t,x) \in (0,T)\times\mathbb{R}^d, \\ \mathbf{V}(T,x) = g(x), & x \in \mathbb{R}^d, \end{cases} \tag{8.39}$$

and the symmetric result holds for $\mathbf{V}^+$.

Note that Eq. (8.39) has a unique viscosity solution, thanks to Theorem 9. When the volatility σ is nondegenerate (i.e., $\sigma\sigma^* \geq \delta I$ for some $\delta > 0$) and the Hamiltonian is sufficiently smooth, the value function turns out to be a classical solution of Eq. (8.39).

The proof of the proposition is extremely tricky and technical: a sub-dynamic programming principle for the lower value function can be obtained, provided that the set of strategies for player 1 is replaced by a smaller one, where some additional measurability condition is required. From this sub-dynamic programming principle, we deduce in a standard way that this new lower value function – which is larger than $\mathbf{V}^-$ – is a subsolution of (8.39). A supersolution of (8.39), which is smaller than $\mathbf{V}^-$, is obtained by discretizing the controls of player 2. The result follows finally by comparison. Of course, similar arguments lead to the analogous result for $\mathbf{V}^+$.

Once the hard work has been done for the upper and lower values, the existence and characterization of the value of the game follow immediately.

Corollary 2. *Suppose that Isaacs' condition holds: for all* $(t,x,p,A) \in [0,T]\times\mathbb{R}^d\times\mathbb{R}^d\times\mathcal{S}_d$,

$$H^-(t,x,p,A) = H^+(t,x,p,A) := H(t,x,p,A).$$

Then the game has a value: $\mathbf{V}^+ = \mathbf{V}^- = \mathbf{V}$ *which is the unique viscosity solution of the HJI equation*

$$\begin{cases} -\partial_t \mathbf{V} - H(t,x,D\mathbf{V},D^2\mathbf{V}) = 0, & (t,x) \in (0,T)\times\mathbb{R}^d, \\ \mathbf{V}(T,x) = g(x), & x \in \mathbb{R}^d. \end{cases} \tag{8.40}$$

4.4 Pathwise Strategies

When one looks on stochastic differential games with incomplete information, the precise definition and analysis of what each player knows about the actions of his opponent becomes crucial. More precisely, in this framework, it is definitively not possible to allow the players to observe a whole control process, but only its ω-wise realization. This has motivated the definition of *pathwise strategies* by Cardaliaguet-Rainer ('13):

We place us here still in the stochastic game setting defined in Sect. 4.1, except that we need to consider, for each $t_0 \in [0, T]$, $(\Omega_{t_0}, \mathcal{F}_{t_0}, \mathbb{P}_{t_0})$, the Wiener space restricted to the time interval $[t_0, T]$. In the following definition, as in the deterministic setting, the delay enables the players to play strategy against strategy.

Definition 11. A **pathwise nonanticipative strategy with delay** at time t_0 for player 1 is a Borel measurable map $\alpha : \Omega_{t_0} \times U_{t_0} \to V_{t_0}$ for which there exists a time grid $t_0 < t_1 < \ldots < t_N = T$ such that, for all $k \leq N - 1$, for all $v_1, v_2 \in V_{t_0}$, and for $\mathbb{P}_{t_0}$-almost any $\omega_1, \omega_2 \in \Omega_{t_0}$, if $(\omega_1, v_1) \equiv (\omega_2, v_2)$ on $[t_0, t_k]$, then $\alpha(\omega_1, v_1) \equiv \alpha(\omega_2, v_2)$ on $[t_0, t_{k+1}]$.

Pathwise nonanticipative strategies with delay for player 2 are defined in a symmetric way. We denote by $\mathcal{A}(t_0)$ (resp. $\mathcal{B}(t_0)$) the set of nonanticipative strategies with delay for player 1 (resp. player 2).

Note that the time grid is a part of the strategy. In other words, to define a strategy, one chooses a time grid and then the controls on this grid.

For each pair $(\alpha, \beta) \in \mathcal{A}(t_0) \times \mathcal{B}(t_0)$, there exists a pair of stochastic controls $(u, v) \in \mathcal{U}(t_0) \times \mathcal{V}(t_0)$ such that

$$u \equiv \alpha(v) \text{ and } v \equiv \beta(u), \ \mathbb{P}_{t_0}\text{-a.s., on } [t_0, T]. \tag{8.41}$$

This allows us to define the value functions as

$$\mathbf{V}^+(t_0, x_0) = \inf_{\alpha} \sup_{\beta} \mathcal{J}(t_0, x_0, \alpha, \beta),$$

and

$$\mathbf{V}^-(t_0, x_0) = \sup_{\beta} \inf_{\alpha} \mathcal{J}(t_0, x_0, \alpha, \beta)$$

where $\mathcal{J}(t_0, x_0, \alpha, \beta) = \mathcal{J}(t_0, x_0, u, v)$, for (u, v) realizing the fix point relation (8.41). This symmetric definition of the value functions simplifies considerably the proof of the existence of the value and its characterization as a solution of a HJI equation:

We avoid here the most difficult part in the approach of Fleming-Souganidis, namely, the inequality which is not covered by comparison: it follows here from the universal relation between sup inf and inf sup. However, the technicality of their proof is replaced here by fine measurable issues upstream.

4.5 Link with Backward Stochastic Differential Equations

In this section we present an approach introduced by Hamadène-Lepeltier ('95), for the case where the volatility is uniformly elliptic and uncontrolled: using a weak formulation (i.e., the players add a controlled drift to the uncontrolled dynamics by using a change of probability) and backward stochastic differential equation (BSDE) techniques, the players can use feedback strategies. Its advantage is firstly its simplicity, given that one is familiar with the BSDE setting. Moreover, it allows to express the value of the game and some optimal controls in terms of the solution of a BSDE. Finally, exploiting the wide range of the actual knowledge on BSDEs, it gives rise to a large class of extensions, which are not directly obtainable by viscosity methods.

We keep the same notations and assumptions on the dynamics and on the cost functions as above, but suppose that $M = d$ (i.e., the Brownian motion has the same dimension as the dynamics).

We start with an uncontrolled stochastic differential equation (SDE): for any initial position $(t_0, x_0) \in [0, T] \times \mathbb{R}^d$, let $(X_s^{t_0,x_0})_{s\in[t_0,T]}$ be the unique solution of the SDE

$$X_s = x_0 + \int_{t_0}^{s} \sigma(r, X_r)dB_r,\ s \in [t_0, T], \tag{8.42}$$

where $\sigma : [0, T] \times \mathbb{R}^d \to \mathbb{R}^{d\times d}$, in addition to assumptions (8.36), satisfies, for all $x \in \mathbb{R}^d$,

$$\frac{1}{\alpha} I \le \sigma(t, x)\sigma^*(t, x) \le \alpha I,$$

with I the identity matrix on $\mathbb{R}^d$ and $\alpha \ge 1$ fixed.

Given a pair of controls $(u, v) \in \mathcal{U}(t_0)\times\mathcal{V}(t_0)$, the players act on the dynamics by changing the probability on the state space Ω: let $\mathbb{P}^{u,v}$ be the probability on $(\Omega, \mathcal{F})$ defined by

$$\frac{d\mathbb{P}^{u,v}}{d\mathbb{P}} = \exp\left\{ \int_{t_0}^{T} \sigma^{-1}(s, X_s^{t_0,x_0})b(s, X_s^{t_0,x_0}, u_s, v_s)ds \right. \\ \left. - \frac{1}{2}\int_{t_0}^{T} |\sigma^{-1}(s, X_s^{t_0,x_0})b(s, X_s^{t_0,x_0}, u_s, v_s)|^2 ds \right\}.$$

For $s \in [t_0, T]$ set

$$B_s^{u,v} = B_s - B_{t_0} - \int_{t_0}^{s} \sigma^{-1}(r, X_r^{t_0,x_0})b(r, X_r^{t_0,x_0}, u_r, v_r)dr.$$

According to Girsanov's theorem, $B^{u,v}$ is a Brownian motion under $P^{u,v}$, and X^{t_0,x_0} is solution to the – now controlled – SDE:

$$dX_s = b(s, X_s, u_s, v_s)ds + \sigma(s, X_s)dB_s^{u,v}, \ s \in [t_0, T].$$

Denoting by $E^{u,v}$ the expectation with respect to $P^{u,v}$, the payoff is now defined by

$$\mathcal{J}(t_0, x_0, u, v) = \mathbb{E}^{u,v}\left[g(X_T^{t_0,x_0}) + \int_{t_0}^T \ell(s, X_s^{t_0,x_0}, u_s, v_s)ds\right],$$

and the upper and lower functions can be simply expressed as

$$\mathbf{V}^+(t_0, x_0) = \inf_{u\in\mathcal{U}(t_0)} \sup_{v\in\mathcal{V}(t_0)} J(t_0, x_0, u, v),$$

and

$$\mathbf{V}^-(t_0, x_0) = \sup_{v\in\mathcal{V}(t_0)} \inf_{u\in\mathcal{U}(t_0)} J(t_0, x_0, u, v).$$

The crucial remark here is that, if one considers the family of conditional payoffs

$$Y_t^{u,v} := \mathbb{E}^{u,v}\left[g(X_T^{t_0,x_0}) + \int_t^T \ell(s, X_s^{t_0,x_0}, u_s, v_s)ds|\mathcal{F}_t\right], \ t \in [t_0, T],$$

then, together with some associated $(Z_t^{u,v})$, the process $(Y_t^{u,v})_{t\in[t_0,T]}$ constitutes the solution of the following BSDE:

$$\begin{aligned} Y_t^{u,v} = Y_T^{u,v} + \textstyle\int_t^T \left(\ell(s, X_s^{t_0,x_0}, u_s, v_s) + \langle Z_s^{u,v}, b(s, X_s^{t_0,x_0}, u_s, v_s)\rangle\right) ds \\ - \textstyle\int_t^T \langle Z_s^{u,v}, \sigma(s, X_s^{t_0,x_0})dB_s\rangle, \ t \in [t_0, T]. \end{aligned}$$

Now let us suppose that Isaacs' condition holds. Since the volatility σ does not depend on controls, it can be rewritten as, for all $(t, x, p) \in |0, T] \times \mathbb{R}^d \times \mathbb{R}^d$,

$$\begin{aligned} &\inf_{u\in U} \sup_{v\in V} \{\langle p, b(t, x, u, v)\rangle + \ell(t, x, u, v)\} \\ &\qquad = \sup_{v\in V} \inf_{u\in U} \{\langle p, b(t, x, u, v)\rangle + \ell(t, x, u, v)\} := G(t, x, p). \end{aligned} \tag{8.43}$$

Since U and V are compact, this assumption implies the existence of a saddle point: there exists a couple of measurable maps (u^*, v^*) such that

$$G(t, x, p) = \langle p, b(t, x, u^*(x, p), v^*(x, p))\rangle + \ell(t, x, u^*(x, p), v^*(x, p)). \tag{8.44}$$

Consider finally the BSDE

$$-dY_t = G(t, X_t^{t_0,x_0}, Z_t)dt - \langle Z_t, \sigma(X_t^{t_0,x_0})dB_t\rangle, \ t \in [t_0, T], \tag{8.45}$$

and define the couple of strategies $(\overline{u}_t, \overline{v}_t) \in \mathcal{U}(t_0) \times \mathcal{V}(t_0)$ by

$$\overline{u}_t = u^*(X_t^{t_0,x_0}, Z_t) \text{ and } \overline{v}_t = v^*(X_t^{t_0,x_0}, Z_t), \ t \in [t_0, T].$$

It is then easy to show that $(Y_t^{\overline{u},\overline{v}}, Z_t)$ is a solution to (8.45), and, from the comparison theorem between solutions of BSDEs, it follows that, under Isaacs' condition (8.43), the couple of controls $(\overline{u}, \overline{v})$ is optimal for the game problem, i.e., $\mathbf{V}^+(t_0, x_0) = \mathbf{V}^-(t_0, x_0) = \mathcal{J}(t_0, x_0, \overline{u}, \overline{v})$. Finally, the characterization of the value of the game as a viscosity solution of the HJI equation (8.37) corresponds to a Feynman-Kac formula for BSDEs.

Let us finally mention other approaches to stochastic zero-sum differential games: Nisio ('88) discretizes the game, Swiech ('96) smoothens the HJ equation and proceeds by verification, and Buckdahn-Li ('08) uses the stochastic backward semigroups introduced by Peng ('97) to establish dynamic programming principles for the value functions – and therefore the existence and characterization of a value in the case when Isaacs' condition holds. This later approach allows to extend the framework to games involving systems of forward-backward SDEs.

5 Games of Kind and Pursuit-Evasion Games

In this section we discuss differential games in which the continuity of the value function is not ensured by the "natural conditions" on the dynamics and the payoff. The first class is the games of kind (introduced by Isaacs in contrast with the "games of degree" described so far): the solution is a set, the victory domains of the players. The second class is the pursuit-evasion games, for which we saw in Sect. 2 that the value is naturally discontinuous if the usable part of the target is not the whole boundary of the target.

5.1 Games of Kind

In this part we consider a zero-sum differential game in which one of the players wants the state of the system to reach an open target, while the other player wants the state of the system to avoid this target forever. In contrast with the differential games with a payoff, the solution of this game is a set: the set of positions from which the first player (or the second one) can win. In Isaacs' terminology, this is *a game of kind*. The problems described here are often described as a viability problem.

The main result is that the victory domains of the players form a partition of the complement of the target and that they can be characterized by the mean of geometric conditions (as a "discriminating domain"). Besides, the common boundary of the victory domains enjoys a semipermeability property. These results go back to Krasovskii and Subbotin (1988). We follow here Cardaliaguet ('96) in order to use nonanticipative strategies.

5.1.1 The Victory Domains

Dynamics: For an initial position $x_0 \in \mathbb{R}^d$, we consider the differential equation

$$\begin{cases} \dot{X}_t = f(X_t, u_t, v_t) & t \in [0, +\infty), \\ X_0 = x_0. \end{cases} \tag{8.46}$$

We assume that f satisfies the standard assumptions (8.9), and we use the notations U_0 and V_0 for the time measurable controls of the players introduced in Sect. 3.1. For any pair $(u, v) \in U_0 \times V_0$, Eq. (8.26) has a unique solution, denoted $X^{x_0,u,v}$.

Strategies: A nonanticipative strategy for the first player is a map $\alpha : V_0 \to U_0$ such that for any two controls $v_1, v_2 \in V_0$ and for any $t \geq 0$, if $v_1 = v_2$ a.e. in $[0, t]$, then $\alpha(v_1) = \alpha(v_2)$ a.e. in $[0, t]$. Nonanticipative strategies for player 1 are denoted $\mathcal{A}$. Nonanticipative strategies for the second player are defined in a symmetric way and denoted $\mathcal{B}$.

Let O be an open subset of $\mathbb{R}^d$: it is the target of player 2. It will be convenient to quantify the points of O which are at a fixed distance of the boundary: for $\epsilon > 0$, let us set

$$O_\epsilon := \{x \in O,\ d_{\partial O}(x) > \epsilon\}.$$

Definition 12. The victory domain of player 1 is the set of initial configurations $x_0 \in \mathbb{R}^d \backslash O$ for which there exists a nonanticipative strategy $\alpha \in \mathcal{A}$ (depending on x_0) of player 1 such that, for any control $v \in V_0$ played by player 2, the trajectory $X^{x_0,\alpha,v}$ avoids O forever:

$$X_t^{x_0,\alpha(v),v} \notin O \qquad \forall t \geq 0.$$

The victory domain of player 2 is the set of initial configurations $x_0 \in \mathbb{R}^d \backslash O$ for which there exists a finite time $T > 0$, $\epsilon > 0$ and a nonanticipative strategy $\beta \in \mathcal{B}$ (all depending on x_0) of player 2 such that, for any control $u \in U_0$ played by player 1, the trajectory $X^{x_0,u,\beta}$ enters O_ϵ within time less than T:

$$\exists t \in [0, T],\ X_t^{x_0,u,\beta(u)} \in O_\epsilon.$$

Note that the definition of victory domain is slightly dissymmetric: however, the fact that the target is open is already dissymmetric.

5.1.2 The Alternative Theorem

In order to proceed, we need a structure condition on the dynamics:

$$\text{The set } \{f(x, u, v),\ u \in U\} \text{ is convex for any } (x, v) \in \mathbb{R}^d \times V. \tag{8.47}$$

Such a convexity condition is standard in the calculus of variation: it guaranties the existence of optimal solution. It plays the same role in our context, ensuring the existence of optimal strategies.

We now introduce sets with a particular property which will help us to characterize the victory domains: these are sets in which the first player can ensure the state of the system to remain. There are two notions, according to which player 1 plays a strategy or reacts to a strategy of his opponent.

Definition 13. A closed subset K of $\mathbb{R}^d$ is a discriminating domain if, for any $x_0 \in K$, there exists a nonanticipative strategy $\alpha \in \mathcal{A}$ of player 1 such that, for any control $v \in V_0$ played by player 2, the trajectory $X^{x_0,\alpha,v}$ remains in K forever:

$$X_t^{x_0,\alpha,v} \in K \qquad \forall t \geq 0.$$

A closed subset K of $\mathbb{R}^d$ is a leadership domain if, for any $x_0 \in K$, for any $\epsilon, T > 0$, and for any nonanticipative strategy $\beta \in \mathcal{B}(0)$, there exists a control $u \in U_0$ such that the trajectory $X^{x_0,u,\beta}$ remains in an $\epsilon-$neigbourhood of K on the time interval $[0, T]$:

$$X_t^{x_0,u,\beta} \in K_\epsilon \qquad \forall t \in [0, T],$$

where $K_\epsilon := \{x \in \mathbb{R}^d,\ \exists y \in K \text{ with } \|x - y\| \leq \epsilon\}$.

The notions of discriminating and leadership domains are strongly related with the *stable bridges* of Krasovskii and Subbotin (1988). The interest of the notion is that, if a discriminating domain K is contained in $\mathbb{R}^d \backslash O$, then clearly K is contained in the victory domain of the first player.

Discriminating or leadership domains are in general not smooth. In order to characterize these sets, we need a suitable notion of generalized normal, the proximal normal. Let K be a closed subset of $\mathbb{R}^d$. We say that $\nu \in \mathbb{R}^d$ is a proximal normal to K at x if the distance to K of $x + \nu$ is $\|\nu\|$:

$$\inf_{y \in K} \|x + \nu - y\| = \|\nu\|.$$

In other words, x belongs to the projection of $x + \nu$ onto K. If K is the closure of an open set with a smooth boundary, then the notion of proximal normal coincides (up to a positive scalar factor) with the classical notion of normal.

Theorem 10 (Characterization of discriminating and leadership domains). *Under the above assumptions, a closed subset K of $\mathbb{R}^d$ is a discriminating domain (resp. leadership domain) if and only if, for any $x \in K$ and for any proximal normal $\nu \in N_K(x)$, one has*

$$\sup_{v \in V} \inf_{u \in U} \langle f(x, u, v), \nu \rangle \leq 0 \qquad (\text{resp. } \inf_{u \in U} \sup_{v \in V} \langle f(x, u, v), \nu \rangle \leq 0). \tag{8.48}$$

A given closed set contains the largest discriminating domain:

Proposition 7 (Discriminating kernel). *The union of all discriminating domains contained in $\mathbb{R}^d \backslash O$ is itself a closed set and a discriminating domain. It is called the discriminating kernel of $\mathbb{R}^d \backslash O$.*

It turns out that, under Isaacs' condition:

$$\sup_{v \in V} \inf_{u \in U} \langle f(x,u,v), p \rangle = \inf_{u \in U} \sup_{v \in V} \langle f(x,u,v), p \rangle \qquad \forall (x,p) \in \mathbb{R}^d \times \mathbb{R}^d , \tag{8.49}$$

the victory domains can be characterized in terms of discriminating kernel.

Theorem 11 (Krasovskii-Subbotin alternative theorem). *Under the above assumptions, the victory domain of player 1 is equal to the discriminating kernel of $\mathbb{R}^d \backslash O$, while the victory domain of player 2 is its complementary in $\mathbb{R}^d \backslash O$.*

In particular, the victory domains of the players form a partition of $\mathbb{R}^d \backslash O$.

5.1.3 Semipermeable Surfaces

From now on we set $K := \mathbb{R}^d \backslash O$ and denote by Σ the common part of the boundary of the victory domains contained in the interior of K. It will be convenient to consider (at least formally) that the (proximal) normal vectors to Σ point toward the victory domain of player 2. Because of its maximality, the discriminating kernel (i.e., the victory domain of the first player) enjoys a particular property. As explained in Sect. 2.2, one expects that its boundary Σ satisfies Isaacs' equation

$$\inf_{u \in U} \sup_{v \in V} \langle f(x,u,v), \nu_x \rangle = 0$$

for any $x \in \Sigma$, where ν_x is the outward unit normal. This can be made rigorous in the following way:

Proposition 8 (Geometric formulation). *Let $Ker_f(K)$ be the discriminating kernel of K. For any $x \in \overline{K \backslash Ker_f(K)}$ which lies in the interior of K and for any ν proximal normal to $\overline{K \backslash Ker_f(K)}$, one has*

$$\inf_{u \in U} \sup_{v \in V} \langle f(x,u,v), -\nu \rangle \geq 0.$$

As explained below, this inequality formally means that player 2 can prevent the state of the system to enter the interior of the discriminating kernel. Recalling that the discriminating kernel is a discriminating domain, thus satisfying the geometric condition (8.48), one concludes that the boundary of the victory domains is a weak solution of Isaacs' equation.

Heuristically, one can also interpret Isaacs' equation as an equation for a semipermeable surface, i.e., each player can prevent the state of the system from crossing Σ in one direction. As the victory domain of the first player is a discriminating domain, Theorem 10 states that the first player can prevent the state of the system from leaving it. The existence of a strategy for the second player is the aim of the following proposition:

Proposition 9. *Assume that the set $f(x,u,V)$ is convex for any (x,u). Let x_0 belong to the interior of K and to the boundary of $Ker_f(K)$. Then there exists a nonanticipative strategy $\beta \in \mathcal{B}(0)$ and a time $T > 0$ such that, for any control $u \in U_0$, the trajectory $(X_t^{x_0,u,\beta})$ remains in the closure of $K \backslash Ker_f(K)$ on* $[0,T]$.

In other words, the trajectory $(X_t^{x_0,u,\beta})$ does not cross Σ for a while.

5.2 Pursuit-Evasion Games

In this section, we revisit, in the light of the theory of viscosity solutions, the pursuit-evasion games. In a first part, we consider general dynamics, but without state constraints and assuming a controllability condition on the boundary of the target. Then we discuss problems with state constraints and without controllability condition.

5.2.1 Pursuit-Evasion Games Under a Controllability Condition

We first consider pursuit-evasion games under condition which guaranty the value function to be continuous. We follow here Soravia ('93).

As in the previous section, the dynamics of the game is given by the ordinary differential equation (8.46). Let $C \subseteq \mathbb{R}^d$ be a closed target. Given a continuous trajectory $X = (X_t)$ in $\mathbb{R}^d$, we denote by $\theta_C(X)$ the minimal time for X to reach C:

$$\theta_C(X) = \inf\{t \geq 0,\ X(t) \in C\}$$

and set $\theta_C(X) = +\infty$ if X avoids C forever. Given a continuous and positive cost function $\ell : \mathbb{R}^d \times U \times V \to \mathbb{R}$, the first player wants to minimize the quantity $\int_0^{\theta_C(X)} \ell(X_t, u_t, v_t)dt$. We denote as usual by $\mathbf{V}^+$ and $\mathbf{V}^-$ the corresponding value functions (written here for nonanticipative strategies):

$$\mathbf{V}^+(x_0) = \sup_{\beta \in \mathcal{B}(0)} \inf_{u \in U_0} \int_0^{\theta_C(X^{x_0,u,\beta(u)})} \ell(X_t^{x_0,u,\beta(u)}, u_t, \beta(u)_t)dt$$

and

$$\mathbf{V}^-(x_0) = \inf_{\alpha \in \mathcal{A}} \sup_{v \in V_0} \int_0^{\theta_C(X^{x_0,\alpha(v),v})} \ell(X_t^{x_0,\alpha(v),v}, \alpha(v)_t, v)dt.$$

Note that in this game, the first player is the pursuer and the second one the evader.

We suppose that f and ℓ satisfy the standard regularity condition (8.9) and (8.10). Moreover, we suppose that ℓ is bounded below by a positive constant:

$$\exists \delta > 0 \text{ such that } \ell(x,u,v) \geq \delta.$$

This condition formalizes the fact that the minimizer is a pursuer, i.e., that he wants the capture to hold quickly. Finally, we assume that Isaacs' condition holds:

$$\begin{aligned} H(x,p) &:= \inf_{u\in U}\sup_{v\in V}\{\langle p, f(x,u,v)\rangle + \ell(x,u,v)\} \\ &= \sup_{v\in V}\inf_{u\in U}\{\langle p, f(x,u,v)\rangle + \ell(x,u,v)\} \end{aligned} \tag{8.50}$$

for any $(x,p) \in \mathbb{R}^d \times \mathbb{R}^d$.

However, without additional assumption, little is known on the problem. Indeed, because of the exit time, the value functions might be discontinuous (and actually even not finite), and the standard approach by viscosity solution does not apply. Recalling the analysis of Sect. 2, we know that we cannot expect the value function to be continuous at the boundary of the target unless the *usable part* of the boundary is the whole set ∂C. Namely, we suppose that the target C is the closure of an open subset of $\mathbb{R}^d$ with a smooth boundary, and we also assume that

$$\inf_{u\in U}\sup_{v\in V}\langle f(x,u,v), \nu_x\rangle \; < \; 0$$

for any $x \in \partial C$, where ν_x is the outward unit normal to C at x. In other words, the pursuer can guaranty the immediate capture when the state of the system is at the boundary of the target. This assumption is often called *a controllability condition*.

We denote by $\mathrm{dom}(\mathbf{V}^{\pm})$ the domain of the maps $\mathbf{V}^{\pm}$, i.e., the set of points where $\mathbf{V}^{\pm}$ are finite.

Theorem 12. *Under the above assumptions, the game has a value:* $\mathbf{V} := \mathbf{V}^+ = \mathbf{V}^-$, *which is continuous on its domain* $\mathrm{dom}(\mathbf{V})$ *and is the unique viscosity of the following HJI equation:*

$$\begin{cases} -H(x, D\mathbf{V}(x)) = 0 & \text{in } \mathrm{dom}(\mathbf{V})\backslash C, \\ \mathbf{V}(x) = 0 & \text{on } \partial C, \\ \mathbf{V}(x) \to +\infty & \text{as } x \to \partial\mathrm{dom}(\mathbf{V})\backslash C. \end{cases} \tag{8.51}$$

Sketch of proof: Thanks to the controllability condition, one can check that the $\mathbf{V}^{\pm}$ are continuous on ∂C and vanish on this set. The next step involves using the regularity of the dynamics and the cost to show that the domains of the $\mathbf{V}^{\pm}$ are open and that $\mathbf{V}^{\pm}$ are continuous on their domain. Then one can show that the $\mathbf{V}^{\pm}$ satisfy a dynamic programming principle and the HJI equation (8.51). The remaining issue is to prove that this equation has a unique solution. There are two difficulties for this: the first one is that the HJI equation does not contain 0-order term (in contrast with the equation for the infinite horizon problem), so even in a formal way, the comparison principle is not straightforward. Second, the domain of the solution is also an unknown of the problem, which technically complicates the proofs.

To overcome these two difficulties, one classically uses the *Kruzhkov transform*:

$$W^{\pm}(x) = 1 - \exp\left\{-\mathbf{V}^{\pm}(x)\right\}.$$

The main advantage of the $W^{\pm}$ compared to the $\mathbf{V}^{\pm}$ is that they take finite values (between 0 and 1). Moreover, the $W^{\pm}$ satisfy a HJI equation with a zero-order term, for which the standard tools of viscosity solutions apply. □

5.2.2 Pursuit-Evasion Games with State Constraints

We now turn to the more difficult issue where there are state constraints and no controllability condition on the boundary of the target. The reader might have in mind the lion and man game, described at the very beginning.

To fix the ideas, we consider here a game with separate dynamics: we suppose that the state can be written as a pair (X_t, Y_t), where (X_t) belongs to $\mathbb{R}^{d_1}$ and is controlled by the first player (say, the pursuer), while (Y_t) is in $\mathbb{R}^{d_2}$ and is controlled by the second player (the evader). There is a constraint on each state: (X_t) (resp. (Y_t)) is restricted to stay in the set $K_X \subseteq \mathbb{R}^{d_1}$ (resp. $K_Y \subseteq \mathbb{R}^{d_2}$). The dynamics becomes

$$\begin{cases} \dot{X}_t = f_1(X_t, u_t), & X_t \in K_X \qquad t \in [0, +\infty), \\ \dot{Y}_t = f_2(X_t, v_t), & Y_t \in K_Y \qquad t \in [0, +\infty), \\ X_0 = x_0 \in K_X, \; Y_0 = y_0 \in K_Y. \end{cases}$$

In contrast with differential games in the whole space, the controls and the strategies of the players depend here on the initial position. For $x_0 \in K_X$, let us denote by $\mathcal{U}_{x_0}$ the set of time measurable controls $u : [0, +\infty) \to U$ such that the corresponding solution (X_t) remains in K_X. We use the symmetric notion $\mathcal{V}_{y_0}$ for the second player. A nonanticipative strategy for the first player is then a nonanticipative map $\alpha : \mathcal{V}_{y_0} \to \mathcal{U}_{x_0}$. We denote by $\mathcal{A}(x_0, y_0)$ and $\mathcal{B}(x_0, y_0)$ the set of nonanticipative strategies for the first and second player, respectively. Under the assumptions stated below, the sets $\mathcal{U}_{x_0}$, $\mathcal{V}_{y_0}$, $\mathcal{A}(x_0, y_0)$ and $\mathcal{B}(x_0, y_0)$ are nonempty.

Let $C \subseteq \mathbb{R}^{d_1+d_2}$ be the target set, which is assumed to be closed. For a given trajectory $(X, Y) = (X_t, Y_t)$, we set

$$\theta_C(X, Y) = \inf\{t \geq 0, \; (X_t, Y_t) \in C\}$$

(with $\theta_C(X, Y) = +\infty$ if there is no time at which $(X_t, Y_t) \in C$). The value functions of the game are defined as follows:

$$\mathbf{V}^-(x_0, y_0) = \inf_{\alpha \in \mathcal{A}(x_0,y_0)} \sup_{v \in V_{y_0}} \theta_C(X^{x_0,\alpha(v)}, Y^{y_0,v})$$

and

$$\mathbf{V}^+(x_0, y_0) = \lim_{\epsilon \to 0^+} \sup_{\beta \in \mathcal{B}(x_0,y_0)} \inf_{u \in U_{x_0}} \theta_{C_\epsilon}(X^{x_0,u}, Y^{y_0,\beta(u)}),$$

where C_ϵ is the set of points which are at a distance of C not larger than ϵ. Note that, as for the games of kind described in the previous section, the definition of the upper and lower value functions is not symmetric: this will ensure the existence of a value.

Besides the standard assumption on the dynamics of the game that we do not recall here, we now assume the sets $f_1(x, U)$ and $f_2(y, V)$ are convex for any x, y. Moreover, we suppose of an inward pointing condition on the vector fields f_1 and f_2 at the boundary of the constraint sets, so that the players can ensure the state of their respective system to stay in these sets. We explain this assumption for the first player, the symmetric one is assumed to hold for the second one. We suppose that the set K_X is the closure of an open set with a C^1 boundary and that there exists a constant $\delta > 0$ such that

$$\inf_{u \in U} \langle f_1(x, u), \nu_x \rangle \leq -\delta \qquad \forall x \in \partial K_X,$$

where ν_x is the outward unit normal to K_X at x. This condition not only ensures that the set U_{x_0} is non empty for any $x_0 \in K_X$ but also that it depends on a Lipschitz continuous way of the initial position x_0. For $x \in K_X$, we denote by $U(x)$ the set of actions that the first player has to play at the point x to ensure that the direction $f_1(x, u)$ points inside K_X:

$$U(x) := \{u \in U,\ \langle f_1(x, u), \nu_x \rangle \leq 0\}.$$

We use the symmetric notation $V(y)$ for the second player. Note that the set-valued maps $x \rightsquigarrow U(x)$ and $y \rightsquigarrow V(y)$ are discontinuous on K_X and K_Y.

Theorem 13 (Cardaliaguet-Quincampoix-Saint Pierre ('01)). *Under the above assumptions, the game has a value,* $\mathbf{V} := \mathbf{V}^+ = \mathbf{V}^-$*, which is the smallest viscosity supersolution of the HJI inequality:*

$$\inf_{u \in U(x)} \langle f_1(x, y), D_x \mathbf{V}(x, y) \rangle + \sup_{v \in V(y)} \langle f_2(y, v), D_y \mathbf{V}(x, y) \rangle \geq -1 \qquad \text{in } K_X \times K_Y.$$

Moreover, the value is lower semicontinuous, and there exists an optimal strategy for the first player.

There are two technical difficulties in this result: the first one is the state constraints for both players. As a consequence, the HJI equation has discontinuous coefficients, the discontinuity occurring at the boundary of the constraint sets. The second difficulty comes from the fact that, since no controllability at the boundary is assumed, the value functions can be discontinuous (and are discontinuous in general). To overcome these issues, one rewrites the problem as a game of kind for a game in higher dimensions.

The result can be extended to the Bolza problems with state constraints and to the infinite horizon problem.

6 Differential Games with Information Issues

In this section we consider differential games in which at least one of the players has not a complete knowledge of the state or of the control played so far by his opponent. This class of game is of course very relevant in practical applications, where it is seldom the case that the players can observe their opponent's behavior in real time or are completely aware of the dynamics of the system.

However, in contrast with the case of complete information, there is no general theory for this class of differential games: although one could expect the existence of a value, no general result in this direction is known. More importantly, no general characterization of the upper and lower values is available. The reason is that the general methodology described so far does not work, simply because the dynamic programming principle does not apply: indeed, if the players do not have a complete knowledge of the dynamics of the system or of their opponent's action, they cannot update the state of the system at later times, which prevents a dynamic programming to hold.

If no general theory is known so far, one has nevertheless identified two important classes of differential games for which a value is known to exist and for which one can describe this value: the search games and the game with incomplete information.

In the *search games*, the players do not observe each other at all. The most typical example is Isaacs' princess and monster game in which the monster tracks the princess in a dark room. In this setting, the players play *open-loop controls*, and because of this simple information structure, the existence of a value can be (rather) easily derived. The difficult part is then to characterize the value or, at least, to obtain qualitative properties of this value.

In *differential games with incomplete information*, the players observe each other perfectly, but at least one of the players has a private knowledge of some parameter of the game. For instance, one could think of a pursuit-evasion game in which the pursuer can be of two different types: either he has a bounded speed (but no bound on the acceleration) or he has a bounded acceleration (but an unbounded speed). Moreover, we assume that the evader cannot not know a priori which type the pursuer actually is: he only knows that the pursuer has fifty-fifty chance to be one of them. He can nevertheless try to guess the type of the pursuer by observing his behavior. The pursuer, on the contrary, knows which type he is, but has probably interest to hide this information in order to deceive the evader. The catch here is to understand how the pursuer uses (and therefore discloses) his type along the game.

6.1 Search Games

Search games are pursuit-evasion games in the dark: players do not observe each other at all. The most famous example of such a game is the princess and monster game, in which the princess and the monster are in a circular room plunged in total darkness and the monster tries to catch the princess in a minimal time. This part is borrowed from Alpern and Gal's monograph (2003).

Let us first fix the notation. Let (Q, d) be a metric compact set in which the game takes place (for instance, in the princess and monster game, Q is the room). We assume that the pursuer can move in Q within a speed less than a fixed constant, set to 1 to fix the ideas. So a pure strategy for the pursuer is just a curve (X_t) on Q which satisfies $d(X_s, X_t) \le |t - s|$. The evader has a maximal speed denoted by w, so a pure strategy for the evader is a curve (Y_t) such that $d(Y_s, Y_t) \le w|t - s|$.

Given a radius $r > 0$, the capture time τ for a pair of trajectories $(X_t, Y_t)_{t \ge 0}$ is the smallest time (if any) such that the distance between the evader and pursuer is not larger than r:

$$\tau(X, Y) = \inf\{t \ge 0,\ d(X_t, Y_t) \le r\}$$

(as usual we set $\tau = +\infty$ if there is no time t at which $d(X_t, Y_t) \le r$).

The set of pure strategies for the pursuer (resp. the evader) is denoted by $\mathcal{X}$ (resp. $\mathcal{Y}$) and is endowed with the topology of uniform convergence on any finite interval. Note that this is a Hausdorff compact space.

A mixed strategy for the pursuer (resp. the evader) is a regular probability measure on $\mathcal{X}$ (resp. $\mathcal{Y}$), and we denote by $\Delta(\mathcal{X})$ (resp. $\Delta(\mathcal{Y})$) this set of strategies.

A variant of Sion's minimax theorem implies the existence of the value:

Theorem 14 (Alpern-Gal ('88)). *The game has a value:*

$$\inf_{\mu \in \Delta(\mathcal{X})} \sup_{\nu \in \Delta(\mathcal{Y})} \int_{\mathcal{X} \times \mathcal{Y}} \tau(X, Y)\, d\mu(X) d\nu(Y)$$
$$= \sup_{\nu \in \Delta(\mathcal{Y})} \inf_{\mu \in \Delta(\mathcal{X})} \int_{\mathcal{X} \times \mathcal{Y}} \tau(X, Y)\, d\mu(X) d\nu(Y).$$

Moreover, there exists an optimal strategy $\overline{\mu}$ for the pursuer.

However, the description of this value does not seem to be known. An interesting question is the behavior of the value as the radius r tends to 0. Here is an answer:

Theorem 15 (Gal ('79, '80)). *Assume that Q is a compact convex subset of $\mathbb{R}^2$, and let $v(r)$ be the value of the game for the radius r. Then, as $r \to 0^+$,*

$$v(r) \sim \frac{|Q|}{2r},$$

where $|Q|$ is the Lebesgue measure of Q.

6.2 Differential Games with Incomplete Information

Differential games with incomplete information are a class of differential games in which (at least) one of the players has a private information on the structure of the game: for instance, he knows perfectly the cost function or the dynamics, while the opponent has only a probabilistic knowledge on these data ("a belief"). Since in these games the players observe each other perfectly, the point is to understand how the players use their information in an optimal way.

This class of problems is the transposition to differential games for Aumann-Maschler analysis of repeated games with incomplete information (see ▶ Chap. 2, "Nonzero-Sum Differential Games").

To summarize, we study here classes of zero-sum differential games in which:

- at least one of the players has some private knowledge on the structure of the game: for instance, he may know precisely some random parameter of the game, while his opponent is only aware of the law of this parameter.
- the players observe each other's control perfectly. In this way they can try to guess their missing information by observing the behavior of his opponent.

We present here a typical result of this class of games in a simple framework: the dynamics is deterministic and the lack of information is only on the payoff. To fix the ideas we consider a finite horizon problem.

Dynamics: As usual, the dynamics is of the form

$$\dot{X}_t = f(X_t, u_t, v_t),$$

and we denote by $(X_t^{t_0,x_0,u,v})_{t\geq t_0}$ the solution of this ODE with initial condition $X_{t_0} = x_0$ when the controls played by the players are (u_t) and (v_t), respectively.

Payoff: We assume that the payoff depends on a parameter $i \in \{1, \dots, I\}$ (where $I \geq 2$): namely, the running payoff $\ell_i = \ell_i(x, u, v)$ and the terminal payoff $g_i = g_i(x)$ depend on this parameter. When the parameter is i, the cost for the first player (the gain for the second one) is then

$$J_i(t_0, x_0, u, v) := \int_{t_0}^{T} \ell_i(X_t, u_t, v_t)\, dt + g_i(X_T).$$

The key point is that the parameter i is known to the first player, but not to the second one: player 2 has only a probabilistic knowledge of i. We denote by $\Delta(I)$ the set of probability measures on $\{1, \ldots, I\}$. Note that elements of this set can be written as $p = (p_1, \ldots, p_I)$ with $p_i \geq 0$ for any i and $\sum_{i=1}^{I} p_i = 1$.

The game is played in two steps. At the initial time t_0, the random parameter i is chosen according to some probability $p = (p_1, \ldots, p_I) \in \Delta(I)$, and the result is told to player 1 but not to player 2. Then the game is played as usual; player 1 is trying to minimize his cost. Both players know the probability p, which can be interpreted as the *a priori belief* of player 2 on the parameter i.

Strategies: In order to hide their private information, the players are naturally led to play *random strategies*, i.e., they choose randomly their strategies. In order to fix the ideas, we assume that the players build their random parameters on the probability space $([0,1], \mathcal{B}([0,1]), \mathcal{L})$, where $\mathcal{B}([0,1])$ is the family of Borel sets on $[0,1]$ and where $\mathcal{L}$ is the Lebesgue measure on $\mathcal{B}([0,1])$. The players choose their random parameter independently. To underline that $[0,1]$ is seen here as a probability space, we denote by ω_1 an element of the set $\Omega_1 = [0,1]$ for player 1 and use the symmetric notation for player 2.

A random nonanticipative strategy with delay (in short random strategy) for player 1 is a (Borel measurable) map $\alpha : \Omega_1 \times V_{t_0} \to U_{t_0}$ for which there is a delay $\tau > 0$ such that for any $\omega_1 \in \Omega_1$ and any two controls $v_1, v_2 \in V_{t_0}$ and for any $t \geq t_0$, if $v_1 \equiv v_2$ on $[t_0, t]$, then $\alpha(\omega_1, v_1) \equiv \alpha(\omega_1, v_2)$ on $[t_0, t + \tau]$.

Random strategies for player 2 are defined in a symmetric way, and we denote by $\mathcal{A}_r(t_0)$ (resp. $\mathcal{B}_r(t_0)$) the set of random strategies for player 1 (resp. player 2).

As for delay strategies, one can associate with a pair of strategies a pair of controls, but, this time, the control is random. More precisely, one can show that, for any pair $(\alpha, \beta) \in \mathcal{A}_r(t_0) \times \mathcal{B}_r(t_0)$, there exists a unique pair of Borel measurable control $(u, v) : (\omega_1, \omega_2) \to U_{t_0} \times V_{t_0}$ such that

$$\alpha(\omega_1, v(\omega_1, \omega_2)) = u(\omega_1, \omega_2) \qquad \text{and} \qquad \beta(\omega_2, u(\omega_1, \omega_2)) = v(\omega_1, \omega_2).$$

This leads us to define the cost associated with the strategies (α, β) as

$$\mathcal{J}_i(t_0, x_0, \alpha, \beta) = \int_0^1 \int_0^1 J_i(t_0, x_0, u(\omega_1, \omega_2), v(\omega_1, \omega_2)) d\omega_1 d\omega_2.$$

Value functions: Let us recall that the first player can choose his strategy in function of i, so as an element of $[\mathcal{A}_r(t_0)]^I$, while the second player does not know i. This leads to the definition of the upper and lower value functions:

$$\mathbf{V}^+(t_0, x_0, p) := \inf_{(\alpha^i) \in (\mathcal{A}_r(t_0))^I} \sup_{\beta \in \mathcal{B}_r(t_0)} \sum_{i=1}^{I} p_i \mathcal{J}_i(t_0, x_0, \alpha^i, \beta)$$

and

$$\mathbf{V}^{-}(t_0, x_0, p) := \sup_{\beta \in \mathcal{B}_r(t_0)} \inf_{(\alpha^i) \in (\mathcal{A}_r(t_0))^I} \sum_{i=1}^{I} p_i \mathcal{J}_i(t_0, x_0, \alpha^i, \beta).$$

Note that the sum $\sum_{i=1}^{I} p_i \dots$ corresponds to the expectation with respect to the random parameter i.

We assume that dynamics and cost functions satisfy the usual regularity properties (8.9) and (8.10). We also suppose that a generalized form of Isaacs' condition holds: for any $(x, \xi) \in \mathbb{R}^d \times \mathbb{R}^d$ and for any $p \in \Delta(I)$,

$$\inf_{u \in U} \sup_{v \in V} \left(\sum_{i=1}^{I} p_i \ell_i(t, x, u, v) + \langle f(t, x, u, v), \xi \rangle \right)$$
$$= \sup_{v \in V} \inf_{u \in U} \left(\sum_{i=1}^{I} p_i \ell_i(t, x, u, v) + \langle f(t, x, u, v), \xi \rangle \right),$$

and we denote by $H(x, p, \xi)$ the common value. Isaacs' assumption can be relaxed, following ideas described in Sect. 3.4.

Theorem 16. *Under the above assumption, the game has a value:* $\mathbf{V}^+ = \mathbf{V}^-$. *This value* $\mathbf{V} := \mathbf{V}^+ = \mathbf{V}^-$ *is convex with respect to the* p *variable and solves in the viscosity sense the Hamilton-Jacobi equation*

$$\begin{cases} \max\left\{ -\partial_t \mathbf{V}(t, x, p) - H(t, x, p, D_x \mathbf{V}(t, x, p))\, ;\, -\Lambda_{\max}(D^2_{pp}\mathbf{V}(t, x, p), p) \right\} = 0 \\ \qquad\qquad\qquad\qquad \text{in } (0, T) \times \mathbb{R}^d \times \Delta(I), \\ V(T, x, p) = \sum_{i=1}^{I} p_i g_i(x) \qquad \text{in } \mathbb{R}^d \times \Delta(I), \end{cases} \tag{8.52}$$

where $\Lambda_{\max}(X, p)$ *is the largest eigenvalue of the symmetric matrix restricted to the tangent space of* $\Delta(I)$ *at* p.

The result heuristically means that the value function – which is convex in p – is either "flat" with respect to p or solves the Hamilton-Jacobi equation. Let us point out that the value function explicitly depends on the parameter p, which becomes a part of the state space.

The above result has been extended (through a series of works by Cardaliaguet and Rainer) in several directions: when both players have a private information (even when this information is correlated; cf. Oliu-Barton ('15)), when there is an incomplete information on the dynamics and on the initial condition as well, when the dynamics is driven by a stochastic differential equation, and when the set of parameters is a continuum.

Sketch of proof. The main issue is that the dynamic principle does not apply in a standard way: indeed, the non-informed player (i.e., player 2) might learn at a part of his missing information by observing his opponent's control. So one cannot start afresh the game without taking into account the fact that the information has changed. Unfortunately, the dynamics of information is quite subtle to quantify in a continuous time.

One way to overcome this issue is first to note that the upper and lower values are convex with respect to p: this can be interpreted as the fact that the value of information is positive.

Then one seeks at showing that the upper value is a subsolution, while the lower value is a supersolution of the HJ equation (8.52). The first point is easy, since, if the first player does not use his private information on the time interval $[t_0, t_1]$, then he plays in a suboptimal way, but, on the other hand, he does not reveal any information, so that one can start the game afresh at t_1 with the new initial position: one thus gets a subdynamic programming for $\mathbf{V}^+$ and the fact that $\mathbf{V}^+$ is a subsolution.

For the supersolution case, one follows ideas introduced by De Meyer ('95) for repeated games and considers the convex conjugates of $\mathbf{V}^-$ with respect to p. It turns out that this conjugate satisfies a subdynamic programming principle, which allows in turn to establish that $\mathbf{V}^-$ is a supersolution to the HJ equation (8.52). One can conclude thanks to a comparison principle for (8.52). □

As in the case of complete information, it would be interesting to derive from the value function the optimal strategies of the players. This issue is a little intricate to formalize in general, and no general answer is known. Heuristically one expects the set of points at which $\mathbf{V}$ is "flat" with respect to p as the set of point for which the first player can use his information (and therefore can reveal it). On the contrary, on the points where the HJ equation is fulfilled, the first player should not use his information at all and play as if he only knew the probability of the parameter.

6.2.1 Optimal Strategies in a Simple Game

This heuristics has been rigorously established in a simple game in which there is no dynamics and where, for simplicity, the terminal cost vanishes. With the notations above, the cost functional becomes

$$J_i(t_0, u, v) := \int_{t_0}^{T} \ell_i(t, u_t, v_t)\, dt.$$

The expected payoff $\mathcal{J}_i$ against random strategies and the value function (which exists according to Theorem 16) are defined accordingly:

$$\mathbf{V}(t_0, p) = \inf_{(\alpha^i)\in(\mathcal{A}_r(t_0))^I} \sup_{\beta\in\mathcal{B}_r(t_0)} \sum_{i=1}^{I} p_i \mathcal{J}_i(t_0, \alpha^i, \beta)$$

$$= \sup_{\beta\in\mathcal{B}_r(t_0)} \inf_{(\alpha^i)\in(\mathcal{A}_r(t_0))^I} \sum_{i=1}^{I} p_i \mathcal{J}_i(t_0, \alpha^i, \beta).$$

Note that there is no longer an $x-$dependence in J_i and $\mathbf{V}$. The map $\mathbf{V}$ is characterized by the Hamilton-Jacobi equation

$$\begin{cases} \max\{-\partial_t \mathbf{V}(t,p) - H(t,p);\ -\Lambda_{\max}(D^2_{pp}\mathbf{V}(t,p), p)\} = 0 & \text{in } (0,T)\times\Delta(I) \\ V(T,p) = 0 \quad \text{in } \Delta(I) \end{cases} \tag{8.53}$$

with

$$H(t,p) := \inf_{u\in U}\sup_{v\in V}\sum_{i=1}^{I} p_i \ell_i(t,u,v) = \sup_{v\in V}\inf_{u\in U}\sum_{i=1}^{I} p_i \ell_i(t,u,v).$$

Let us now note that, if there was no information issue (or, more precisely, if the informed player did not use his information), the value of the game starting from p_0 at time t_0 would simply be given by $\int_{t_0}^{T} H(t, p_0)dt$. Suppose now that player 1 uses – at least partially – his information. Then, if player 2 knows player 1's strategy, he can update at each time the conditional law of the unknown parameter i: this leads to a martingale $(\mathbf{p}_t)$, which lives in $\Delta(I)$. Somehow the game with incomplete information can be interpreted as a pure manipulation of the information through this martingale:

Theorem 17. *The following equality holds:*

$$\mathbf{V}(t_0, p_0) = \min \mathbb{E}\left[\int_{t_0}^{T} H(s, \mathbf{p}(s))ds\right] \qquad \forall (t_0, p_0) \in [0,T]\times\Delta(I)\ , \tag{8.54}$$

where the minimum is taken over all the (càdlàg) martingales $\mathbf{p}$ *on* $\Delta(I)$ *starting from* p_0 *at time* t_0.

The martingale process can be interpreted as the information on the index i the informed player discloses along the time. One can show that the informed player can use the martingale which is optimal in (8.54) in order to build his strategy. In particular, this optimal martingale plays a key role in the analysis of the game.

For this reason, it is interesting to characterize such an optimal martingale. For this let us assume, for simplicity, that the value function is sufficiently smooth. Let us denote by $\mathcal{R}$ the subset of $[0,T]\times\Delta(I)$ at which the following equality holds:

$$-\partial_t \mathbf{V}(t,p) - H(t,p) = 0.$$

Let $\mathbf{p}$ be a martingale starting from p_0 at time t_0. Then one can show, under technical assumptions on $\mathbf{p}$, that $\mathbf{p}$ is optimal for (8.54) if and only if, for any $t \in [t_0, T]$, the following two conditions are satisfied:

(i) $(t, \mathbf{p}_t)$ belongs to $\mathcal{R}$,
(ii) $\mathbf{V}(t,\mathbf{p}(t)) - \mathbf{V}(t,\mathbf{p}(t^-)) - \langle D_p \mathbf{V}(t, p(t^-)), \mathbf{p}(t) - \mathbf{p}(t^-)\rangle = 0.$

In some sense the set $\mathcal{R}$ is the set on which the information changes very slowly (or not at all), while it can jump only on the flat parts of $\mathbf{V}$.

We complete this section by discussing two examples.

Example 1. Assume that $H = H(p)$ does not depend on t. Then we claim that

$$\mathbf{V}(t,p) = (T-t)\mathrm{Vex}H(p) \qquad \forall (t,p) \in [0,T] \times \Delta(I),$$

where $\mathrm{Vex}H(p)$ is the convex hull of the map H. Indeed let us set $W(t,p) = (T-t)\mathrm{Vex}H(p)$. Then (at least in a formal way), W is convex in p and satisfies

$$-\partial_t W(t,p) - H(p) \le -\partial_t W(t,p) - \mathrm{Vex}H(p) = 0.$$

So W is a subsolution to (8.53). In order to prove that it is a supersolution, we consider two cases: either $\mathrm{Vex}H(p) = H(p)$ or $\mathrm{Vex}H(p) < H(p)$, so that $\mathrm{Vex}H(p)$ – and thus W – is "flat" at p. In the first case, one has $-\partial_t W(t,p) - H(p) = 0$; in the second one, $D^2 W(t,p) = 0$. So W "satisfies" the HJ equation (8.53) with terminal condition 0 and is therefore equal to the value $\mathbf{V}$.

In this case, an optimal martingale just needs to jump once at time t_0 from the initial position p_0 to one of the positions $p_0^k \in \Delta(I)$ with probability λ^k, where the p_0^k and $(\lambda_k) \in \Delta(I)$ are such that $p_0 = \sum_k \lambda_k p_0^k$ and $\mathrm{Vex}(H)(p_0) = \sum_k \lambda_k H(p_0^k)$.

Let us emphasize that the above statement is very close to the "$Cav\ u$ theorem" by Aumann and Maschler.

Example 2. Assume that $I = 2$, and let us identify $p \in [0,1]$ with the pair $(p, 1-p) \in \Delta(2)$. We suppose that there exists $h_1, h_2 : [0,T] \to [0,1]$ continuous, $h_1 \le h_2$, h_1 decreasing, and h_2 increasing, such that

$$\mathrm{Vex}H(t,p) = H(t,p) \ \Leftrightarrow \ p \in [0, h_1(t)] \cup [h_2(t), 1]$$

(where $\mathrm{Vex}H(t,p)$ denote the convex hull of $H(t,\cdot)$) and

$$\frac{\partial^2 H}{\partial p^2}(t,p) > 0 \qquad \forall (t,p) \text{ with } p \in [0, h_1(t)) \cup (h_2(t), 1].$$

Then one can show that $\mathbf{V}$ can be explicitly computed:

$$\mathbf{V}(t, p) = \int_t^T \text{Vex}H(s, p)ds \qquad \forall (t, p) \in [0, T] \times \Delta(I).$$

Moreover, the optimal martingale is unique and can be described as follows: as long as p_0 does not belong to $[h_1(t), h_2(t)]$, one has $\mathbf{p}_t = p_0$. As soon as p_0 belongs to the interval $[h_1(t_0), h_2(t_0)]$, the martingale $(\mathbf{p}_t)$ switches randomly between the graphs of h_1 and h_2.

7 Long-Time Average and Singular Perturbation

We complete the chapter by a quick presentation of the singular perturbation problems in differential games. This is a large and fascinating topic, which remains open in full generality. We present here two typical results: the first one is on the long-time average and the second one on homogenization.

7.1 Long-Time Average

When one considers differential games with a (large) time horizon or differential games in infinite horizon but small discount rate, one may wonder to what extent the value and the strategies strongly depend on the horizon or on the discount factor. The investigation of this kind of issue started with a seminal paper by Lions-Papanicolaou-Varadhan (circa '86).

We discuss here deterministic differential games in which one of the players can control the game in a suitable sense. To fix the ideas, we work under Isaacs' condition (but this plays no role in the analysis). Let us denote by $\mathbf{V}_T$ the value of the Bolza problem for a horizon $T > 0$ and $\mathbf{V}_\lambda$ the value of the infinite horizon problem with discount factor $\lambda > 0$. We know that $\mathbf{V}_T$ is a viscosity solution to

$$\begin{cases} -\partial_t \mathbf{V}_T(t, x) - H(t, x, D\mathbf{V}_T(t, x)) = 0 & \text{in } (0, T) \times \mathbb{R}^d \\ \mathbf{V}_T(T, x) = g(x) & \text{in } \mathbb{R}^d, \end{cases} \tag{8.55}$$

while $\mathbf{V}_\lambda$ solves

$$\lambda \mathbf{V}_\lambda(x) - H(x, D\mathbf{V}_\lambda(x)) = 0 \qquad \text{in } \mathbb{R}^d \tag{8.56}$$

where H is defined by

$$\begin{aligned} H(x, p) &= \inf_{u \in U} \sup_{v \in V} \{\langle p, f(x, u, v)\rangle + \ell(x, u, v)\} \\ &= \sup_{v \in V} \inf_{u \in U} \{\langle p, f(x, u, v)\rangle + \ell(x, u, v)\}. \end{aligned} \tag{8.57}$$

We seek convergence of the quantity $\mathbf{V}_T(0,x)/T$ or $\lambda\mathbf{V}_\lambda(x)$ as $T \to +\infty$ or $\lambda \to 0$. As this convergence issue is related to ergodicity, one needs conditions which "confine" the state of the system into a compact set. To fix the ideas, we assume here that the maps f, g, and ℓ are periodic in space of period 1: for instance,

$$f(x+k,u,v) = f(x,u,v) \qquad \forall k \in \mathbb{Z}^d,\ \forall (x,u,v) \in \mathbb{R}^d \times U \times V.$$

Then H is also periodic in space:

$$H(x+k,p) = H(x,p) \qquad \forall k \in \mathbb{Z}^d,\ \forall (x,p) \in \mathbb{R}^d \times \mathbb{R}^d.$$

This means that the game is actually played in the torus $\mathbb{R}^d/\mathbb{Z}^d$. Moreover, we assume that the Hamiltonian is coercive with respect to the second variable:

$$\lim_{|p|\to+\infty} \inf_{x\in\mathbb{R}^d} H(x,p) = +\infty. \tag{8.58}$$

This can be interpreted as a strong control of the second player on the game: indeed one can show that the second player can drive the system wherever he wants to and with a finite cost. Note, however, that H need not be convex with respect to p, so we really deal with a game problem.

Theorem 46 (Lions-Papanicolaou-Varadhan ('86)). *Under the above assumptions, there exists a constant $\overline{c} \in \mathbb{R}$ such that*

$$\lim_{T\to+\infty} \frac{\mathbf{V}_T(0,x)}{T} = \lim_{\lambda\to 0^+} \lambda\mathbf{V}_\lambda(x) = \overline{c},$$

the convergence being uniform with respect to x. Moreover, $\overline{c}$ is the unique constant for which the following equation has a continuous, periodic solution:

$$-H(x, D\chi(x)) = \overline{c} \qquad \text{in } \mathbb{R}^d. \tag{8.59}$$

The map χ is often called a *corrector*. Equation (8.59) is the *cell-problem* or the *corrector equation*.

As we will see in the next paragraph, the above result plays a key role in homogenization, i.e., in situations where the dynamics and cost depend on a fast variable and a slow variable. One can then show that the problem is equivalent to a differential game with a dynamics and cost depending on the slow variable only: this approach is called *model reduction*.

Theorem 46 has been extended to stochastic differential games by Evans ('89): in this setting, the controllability condition can be replaced by a uniform ellipticity assumption on the diffusion matrix.

Even for first-order problems, Theorem 46 actually holds in more general frameworks, and we refer to the monograph by Alvarez and Bardi (2010) for details. However, when neither controllability nor uniform ellipticity holds, the above convergence is not expected: in particular, the limit – even when it exists – might depend on the space variable. Identifying general configurations for which a limit exists is however far from understood.

Sketch of proof. We start with $\mathbf{V}_\lambda$ and first note that $\mathbf{V}_\lambda$ is $\mathbb{Z}^d$−periodic. We also note that, thanks to the maximum principle and Eq. (8.56),

$$\inf_{x\in\mathbb{R}^d} H(x,0) \le \lambda \mathbf{V}_\lambda(x) \le \sup_{x\in\mathbb{R}^d} H(x,0).$$

In particular, still thanks to Eq. (8.56), the quantity $H(x, D\mathbf{V}_\lambda(x))$ is bounded independently of λ. The coercivity condition (8.58) then ensures that $D\mathbf{V}_\lambda$ is uniformly bounded, i.e., $\mathbf{V}_\lambda$ is Lipschitz continuous, uniformly with respect to λ. We now set $\chi_\lambda(x) = \mathbf{V}_\lambda(x) - \mathbf{V}_\lambda(0)$ and note that χ_λ is $\mathbb{Z}^d$−periodic and uniformly Lipschitz continuous and vanishes at 0. By the Arzela-Ascoli theorem, there is a sequence (λ_n), which converges to 0, such that χ_{λ_n} uniformly converges to some continuous and $\mathbb{Z}^d$−periodic map χ. We can also assume that the bounded sequence $(\lambda_n \mathbf{V}_{\lambda_n}(0))$ converges to some constant $\overline{c}$. Then it is not difficult to pass to the limit in (8.56) to get

$$-H(x, D\chi(x)) = \overline{c} \qquad \text{in } \mathbb{R}^d.$$

The above equation has several solutions, but the constant $\overline{c}$ turns out to be unique.

Let us now explain the convergence of $\mathbf{V}_T(0,\cdot)/T$. It is not difficult to show that the map $(t,x) \to \chi(x) + \overline{c}(T-t) + \|g\|_\infty + \|\chi\|_\infty$ is a supersolution to (8.55). Thus, by comparison, we have

$$\mathbf{V}_T(t,x) \le \chi(x) + \overline{c}(T-t) + \|g\|_\infty.$$

Dividing by T and letting $T \to +\infty$, we get

$$\limsup_{T\to+\infty} \sup_{x\in\mathbb{R}^d} \frac{\mathbf{V}_T(0,x)}{T} \le \overline{c}.$$

In the same way, we obtain

$$\liminf_{T\to+\infty} \inf_{x\in\mathbb{R}^d} \frac{\mathbf{V}_T(0,x)}{T} \ge \overline{c}.$$

This proves the uniform convergence of $\mathbf{V}_T(0,\cdot)/T$ to $\overline{c}$. The convergence of $\lambda\mathbf{V}_\lambda$ relies on the same kind of argument. □

7.2 Singular Perturbation and Homogenization

We can apply the results of the previous section to singular perturbation problems. We consider situations in which dynamics and cost game depend on several scales: the state can be split between a variable (the so-called fast variable) which evolves quickly and on another variable with a much slower motion (the so-called *slow variable*). The idea is that one can then break the problem into two simpler problems, in smaller dimensions.

Let us fix a small parameter $\epsilon > 0$ and let us assume that one can write the differential game as follows: the state of the system is of the form $(X, Y) \in \mathbb{R}^{d_1} \times \mathbb{R}^{d_2}$, with dynamics

$$\begin{cases} \dot{X}_t^\epsilon = f_1(X_t^\epsilon, Y_t^\epsilon, u_t, v_t) \\ \epsilon \dot{Y}_t^\epsilon = f_2(X_t^\epsilon, Y_t^\epsilon, u_t, v_t) \\ X_0^\epsilon = x_0, \; Y_0^\epsilon = y_0 \end{cases}$$

and a cost of the form

$$\int_0^T \ell(X_t^\epsilon, Y_t^\epsilon, u_t, v_t)dt + g^\epsilon(X_T^\epsilon, Y_T^\epsilon).$$

As ϵ is a small parameter, the variable Y^ϵ evolves in a much faster scale than the slow variable X^ϵ. The whole point is to understand the limit system as $\epsilon \to 0$. It turns out that the limit *does not necessarily consist in taking* $\epsilon = 0$ *in the system.*

We illustrate this point when the dependence with respect to the fast variable is periodic: the resulting problem is of *homogenization type*. In order to proceed, let us make a change of variable by setting $\tilde{Y}_t^\epsilon = \epsilon Y_t^\epsilon$. In these variables the state of the system has for dynamics:

$$\begin{cases} \dot{X}_t^\epsilon = f_1(X_t^\epsilon, \tilde{Y}_t^\epsilon/\epsilon, u_t, v_t) \\ \dot{\tilde{Y}}_t^\epsilon = f_2(X_t^\epsilon, \tilde{Y}_t^\epsilon/\epsilon, u_t, v_t) \\ X_0^\epsilon = x_0, \; \tilde{Y}_0^\epsilon = y_0/\epsilon \end{cases}$$

We define the Hamiltonian of the problem as follows:

$$H(x, y, p_x, p_y) = \inf_{u \in U} \sup_{v \in V} \big(\langle f(x, y, u, v), (p_x, p_y)\rangle + \ell(x, y, u, v)\big),$$

where $f = (f_1, f_2)$ and for $(x, y, p_x, p_y) \in \mathbb{R}^{d_1} \times \mathbb{R}^{d_2} \times \mathbb{R}^{d_1} \times \mathbb{R}^{d_2}$. We also assume that the terminal cost can be written in the form $g^\epsilon(x, y) = g(x, \epsilon y)$. Then, if we write the value function $\mathbf{V}^\epsilon$ of the game in terms of the variables $(X, \tilde{Y})$, it satisfies the HJI equation

$$\begin{cases} -\partial_t \mathbf{V}^\epsilon(t, x, y) - H(x, y/\epsilon, D_x \mathbf{V}^\epsilon, D_y \mathbf{V}^\epsilon) = 0 \qquad \text{in } (0, T) \times \mathbb{R}^{d_1} \times \mathbb{R}^{d_2} \\ \mathbf{V}^\epsilon(T, x, y) = g(x, y) \qquad \text{in } \mathbb{R}^{d_1} \times \mathbb{R}^{d_2} \end{cases}$$

Besides the standard assumption on the dynamics and cost, we suppose that H is $\mathbb{Z}^d$−periodic with respect to the y variable and satisfies the coercivity condition

$$\lim_{|p_y|\to+\infty} \inf_{y\in\mathbb{R}^d} H(x,y,p_x,p_y) = +\infty,$$

locally uniformly in the (x, p_x) variables.

If we freeze the variables (x, p_x), we know, from Theorem 46, that, for any p_y, there exists a constant $\bar{c} =: -\overline{H}(x, p_x, p_y)$ for which the cell problem

$$-H(x, y, p_x, D\chi(y) + p_y) = -\overline{H}(x, p_x, p_y)$$

has a periodic viscosity solution $\chi = \chi(y)$ in $\mathbb{R}^{d_2}$.

Theorem 47 (Lions-Papanicolaou-Varadhan ('86)). *As* $\epsilon \to 0, \mathbf{V}^\epsilon$ *converges locally uniformly to* $\overline{\mathbf{V}}$ *solution to*

$$\begin{cases} -\partial_t\overline{\mathbf{V}}(t,x,y) - \overline{H}(x, D_x\overline{\mathbf{V}}, D_y\overline{\mathbf{V}}) = 0 & \text{in } (0,T)\times\mathbb{R}^{d_1}\times\mathbb{R}^{d_2} \\ \overline{\mathbf{V}}(T,x,y) = g(x,y) \quad \text{in } \mathbb{R}^{d_1}\times\mathbb{R}^{d_2} \end{cases}$$

As a particular case, if g does not depend on y, then the system is actually in dimension d_1: so there is a reduction of dimension, the price being to compute the averaged Hamiltonian $\overline{H}$. One can also prove a convergence when the terminal condition g depends on the fast variable Y (and not on ϵY) and is periodic with respect to that variable. Then there appears a boundary layer at time T.

We described here only a particular instance of singular perturbation problems. A general introduction to the subject can be found in the monograph by Alvarez-Bardi (2010).

The heuristics behind the result is the following. Let us assume that $\mathbf{V}^\epsilon$ uniformly converges to a map $\overline{\mathbf{V}}$. In order to guess the equation satisfied by $\overline{\mathbf{V}}$, let us try to write a first-order expansion in ϵ of $\mathbf{V}^\epsilon$:

$$\mathbf{V}^\epsilon(t,x,y) = \overline{\mathbf{V}}(t,x,y) + \epsilon\chi^\epsilon(t,x,y/\epsilon)$$

Then, recalling the equation for $\mathbf{V}^\epsilon$ and neglecting the terms of order ϵ in the expansion, one should expect:

$$-\partial_t\overline{\mathbf{V}}(t,x) - H(x, y/\epsilon, D_x\overline{\mathbf{V}}(t,x), D_y\overline{\mathbf{V}}(t,x,y) + D_y\chi^\epsilon(t,x,y/\epsilon)) \approx 0.$$

Since y/ϵ evolves in a faster scale than x and y, the quantity in the Hamiltonian should be almost constant in y/ϵ. If we denote by $-\overline{H}(x, D_x\overline{\mathbf{V}}(t,x), D_y\overline{\mathbf{V}}(t,x,y))$ this constant quantity, we find the desired result.

Of course, the actual proof of Theorem 47 is more subtle than the above reasoning. As the maps are at most Lipschitz continuous, one has to do the proof at the level of viscosity solution. For this, a standard tool is *Evans' perturbed test function method*, which consists in perturbing the test function by using in a suitable way the corrector.

8 Conclusion

The analysis of two-person, zero-sum differential games started in the middle of the twentieth century with the works of Isaacs (1965) and of Pontryagin (1968). We have described in Sect. 2 some of Isaacs' ideas: he introduced tools to compute explicitly the solution of some games (homicidal chauffeur game, princess and monster game, the war of attrition and attacks, etc.) by solving the equation denoted in this text "Hamilton-Jacobi-Isaacs" equation. He also started the analysis of the possible discontinuities and singularities of the value function. The technique of resolution was subsequently carried out by several authors (such as Breakwell, Bernhard, Lewin, Melikyan, and Merz, to cite only a few names).

The first general result on the existence of a value goes back to the early 1960s with the pioneering work of Fleming (1961). There were subsequently (in the 1960s–1970s) many contributions on the subject, by Berkovitz, Elliot, Friedman, Kalton, Roxin, Ryll-Nardzewski, and Varaiya (to cite only a few names). Various notions of strategies were introduced at that time. However, at that stage, the connection with the HJI equations was not understood. Working with the notion of positional strategy, Krasovskii and Subbotin (1988) managed to characterize the value in terms of stable bridges (described in Sect. 5), and Subbotin (1995) later used Dini derivatives to write the conditions in terms of the HJI equation.

But it is only with the introduction of the notion of viscosity solution (Crandall and Lions ('81)) that the relation between the value function and HJI equation became completely clear, thanks to the works of Evans and Souganidis (1984) for deterministic differential games and of Fleming and Souganidis (1989) for stochastic ones. A detailed account of the large literature on the "viscosity solution approach" to zero-sum differential games can be found in Bardi and Capuzzo Dolcetta (1996). We reported on these ideas in Sects. 3 and 4: there we explained that zero-sum differential games with complete information and perfect observation have a value under fairly general conditions; we also characterized this value as the unique viscosity solution of the Hamilton-Jacobi-Isaacs equation. This characterization allows us to understand in a crisp way several properties of the value (long-time average, singular perturbation, as in Sect. 7). Similar ideas can be used for pursuit-evasion games, with the additional difficulty that the value function might be discontinuous (Sect. 5) and, to some extent, to games with information issues (Sect. 6).

We now discuss possible extensions and applications of the problems presented above.

- *Other classes of dynamics:* We have chosen to discuss differential games with classical dynamics (ODEs or SDEs). In many practical applications, there is no clear distinction between systems in discrete time and in continuous time: the state may alternatively be driven by a continuous motion (an ODE) and a discontinuous one (jump terms); this leads to the analysis of hybrid systems. Closely related are the problems with switching controls, where the controllers can change their controls only at discrete instants of time. These games naturally lead to the HJI equation with an obstacle problem or to monotone systems of HJI equations.
- *Boundary conditions:* For pursuit-evasion games, we explained that the value function has to vanish at the boundary of the domain. This is only an instance of the various boundary conditions that can be encountered in differential game theory. For example, in stochastic problems, the state is often restricted to remain in a bounded set by a "reflection" term at the boundary. Then the associated HJI equation has a boundary term, the Neumann boundary condition.
- *Applications :* Zero-sum differential games are mainly used in optimal control problems with uncertainty, where the second player is seen as disturbance against which the controller wants to guaranty himself completely; this leads to the notion of worst-case design or robust control. Pursuit-evasion games formalize problems in air traffic management systems and flight control or ground transportation systems: see ▶ Chap. 22, "Pursuit-Evasion Games" on aeronautics. Besides these applications, zero-sum differential games have also been used in less expected areas. We only provide a few examples. Risk-sensitive control problems study the singular perturbation for a stochastic optimal control problem when the noise vanishes. It turns out that, at the small noise limit, the system behaves as if the noise was an opponent: this leads naturally to a zero-sum differential game (see ▶ Chap. 9, "Robust Control and Dynamic Games"). Zero-sum differential games also appear in front propagation problems when the so-called Wulff shape is not convex: they are used to provide an insight on the evolution of the front. They are also utilized in nonzero-sum differential games to describe Nash equilibria in games with memory strategies (▶ Chap. 2, "Nonzero-Sum Differential Games").

Further Reading: The interested reader will find in the Bibliography much more material on zero-sum differential games. Isaacs' approach has been the subject of several monographs Başar and Olsder (1999), Blaquière et al. (1969), Isaacs (1965), and Melikyan (1998); the early existence theories for the value function are given in Friedman (1971); Krasovskii and Subbotin explained their approach of positional differential games in Krasovskii and Subbotin (1988); a classical reference for the theory of viscosity solutions is Crandall et al. (1992), while the analysis of (deterministic) control and differential games by the viscosity solution approach is developed in detail in Bardi and Capuzzo Dolcetta (1996); a reference on pursuit games is Petrosjan (1993) and Alvarez and Bardi (2010) collects many interesting results on the viscosity solution approach to singularly perturbed zero-sum differential games; finally, more recent results on differential games can be found in the survey paper (Buckdahn et al. 2011).

References

Alpern S, Gal S (2003) The theory of search games and rendezvous, vol 55. Springer science and business media, New York

Alvarez O, Bardi M (2010) Ergodicity, stabilization, and singular perturbations for Bellman-Isaacs equations, vol 204, no. 960. American Mathematical Society, Providence

Başar T, Olsder GJ (1999) Dynamic noncooperative game theory. Reprint of the second (1995) edition. Classics in applied mathematics, vol 23. Society for Industrial and Applied Mathematics (SIAM), Philadelphia

Bardi M, Capuzzo Dolcetta I (1996) Optimal control and viscosity solutions of Hamilton-Jacobi-Bellman equations. Birkhäuser, Basel

Blaquière A, Gérard F, Leitman G (1969) Quantitative and qualitative games. Academic Press, New York

Buckdahn R, Cardaliaguet P, Quincampoix M (2011) Some recent aspects of differential game theory. Dyn Games Appl 1(1):74–114

Crandall MG, Ishii H, Lions P-L (1992) User's guide to viscosity solutions of second-order partial differential equations. Bull Am Soc 27:1–67

Evans LC, Souganidis PE (1984) Differential games and representation formulas for solutions of Hamilton-Jacobi equations. Indiana Univ Math J 282:487–502

Fleming WH (1961) The convergence problem for differential games. J Math Anal Appl 3: 102–116

Fleming WH, Souganidis PE (1989) On the existence of value functions of two-player, zero-sum stochastic differential games. Indiana Univ Math J 38(2):293–314

Friedman A (1971) Differential games. Wiley, New York

Isaacs R (1965) Differential games. Wiley, New York

Lewin J (1994) Differential games. Theory and methods for solving game problems with singular surfaces. Springer, London

Krasovskii NN, Subbotin AI (1988) Game-theorical control problems. Springer, New York

Melikyan AA (1998) Generalized characteristics of first order PDEs. Applications in optimal control and differential games. Birkhäuser, Boston

Petrosjan L (1993) A. Differential games of pursuit. Translated from the Russian by J. M. Donetz and the author. Series on optimization, vol 2. World Scientific Publishing Co., Inc., River Edge

Pontryagin LS (1968) Linear differential games I and II. Soviet Math Doklady 8(3 and 4):769–771 and 910–912

Subbotin AI (1995) Generalized solutions of first order PDEs: the dynamical optimization perspective. Birkäuser, Boston

Robust Control and Dynamic Games

9

Pierre Bernhard

Contents

Abstract

We describe several problems of "robust control" that have a solution using game theoretical tools. This is by no means a general overview of robust control theory beyond that specific purpose nor a general account of system theory with set description of uncertainties.

Keywords

Robust control · $\mathcal{H}^\infty$-optimal control · Nonlinear $\mathcal{H}^\infty$ control · Robust control approach to option pricing

P. Bernhard (✉)
Emeritus Senior Scientist, Biocore Team, Université de la Côte d'Azur-INRIA, Sophia Antipolis Cedex, France
e-mail: Pierre.Bernhard@inria.fr

T. Başar, G. Zaccour (eds.), *Handbook of Dynamic Game Theory*,
https://doi.org/10.1007/978-3-319-44374-4_24

1 Introduction

Control theory is concerned with the design of control mechanisms for dynamic systems that compensate for (a priori) unknown disturbances acting on the system to be controlled. While early servomechanism techniques did not make use of much modeling, neither of the plant to be controlled nor of the disturbances, the advent of multi-input multi-output systems and the drive to more stringent specifications led researchers to use mathematical models of both. In that process, and most prominently with the famous Linear Quadratic Gaussian (LQG) design technique, disturbances were described as unknown inputs to a known plant and usually high-frequency inputs.

"Robust control" refers to control mechanisms of dynamic systems that are designed to counter unknowns in the system equations describing the plant rather than in the inputs. This is also a very old topic. It might be said that a PID controller is a "robust" controller, since it makes little assumption on the controlled plant. Such techniques as gain margin and phase margin guarantees and loop shaping also belong in that class. However, the gap between "classic" and "robust" control designs became larger with the advent of "modern" control designs.

As soon as the mid-1970s, some control design techniques appeared, often based upon some sort of Lyapunov function, to address that concern. A most striking feature is that they made use of bounded uncertainties rather than the prevailing Gaussian-Markovian model. One may cite Gutman and Leitmann (1976), Gutman (1979), and Corless and Leitman (1981). This was also true with the introduction of the so-called $\mathcal{H}^\infty$ control design technique in 1981 (Zames 1981) and will become such a systematic feature that "robust control" became in many cases synonymous to control against bounded uncertainties, usually with no probabilistic structure on the set of possible disturbances.

We stress that this is *not* the topic of this chapter. It is not an account of system theory for systems with set description of the uncertainties. Such comprehensive theories have been developed using set-valued analysis and/or elliptic approximations. See, e.g., Aubin et al. (2011) and Kurzhanski and Varaiya (2014). We restrict our scope to *control* problems, thus ignoring pure state estimation such as developed in the above references, or in Başar and Bernhard (1995, Chap. 7) or Rapaport and Gouzé (2003), for instance, and to game theoretic methods in control design, thus ignoring other approaches of robust control such as the gap metric (Georgiou and Smith 1997; Qiu 2015; Zames and El-Sakkary 1980), an input-output approach in the spirit of the original $\mathcal{H}^\infty$ approach of Zames (1981), using transfer functions and algebraic methods in the ring of rational matrices, or linear matrix inequality (LMI) approaches (Kang-Zhi 2015), and many other approaches found, e.g., in the same Encyclopedia of Systems and Control (Samad and Ballieul 2015).

We will first investigate in Sect. 2 the issue of modeling of the disturbances, show how non-probabilistic set descriptions naturally arise, and describe some specific tools we will use to deal with them. Section 2.1 is an elementary introduction to these questions, while Sect. 2.2 is more technical, introducing typical control theoretic issues and examples in the discussion. From there on, a good control of linear systems theory, optimal control—specifically the Hamilton-Jacobi-Caratheodory-Isaacs-Bellman (HJCIB) theory—and dynamic games theory is required. Then, in Sect. 3, we will give a rather complete account of the linear $\mathcal{H}^\infty$-optimal control design for both continuous time and discrete-time systems. Most of the material of these two sections can be found in more detail in Başar and Bernhard (1995). In Sect. 4, we will cover in less detail the so-called "nonlinear $\mathcal{H}^\infty$" control design, mainly addressing engineering problems, and a nonlinear example in the very different domain of mathematical finance.

2 Min-Max Problems in Robust Control

2.1 Decision Theoretic Formulation

2.1.1 Worst Case Design

We start this section with a very general setup in terms of decision theory. Let U be a decision space and W a disturbance space. We may think of W as being a bounded set in some ad hoc norm, but this is not necessary at this level of generality. Let $J : U \times W \to \mathbb{R}$ be a performance index, depending on the decision $u \in U$ and on an unknown disturbance $w \in W$. The decision maker, choosing u, wants to make J as small as possible in spite of the a priori unknown disturbance $w \in W$. (One may think of J as a measure, such as the L^2 norm, of an error signal.)

The so-called *worst case design* method is as follows. Let us first emphasize that there is no "malicious adversary" manipulating the disturbance in the following description, contrary to many accounts of this method.

The basic concept is that of *guaranteed performance* of a given decision $u \in U$: any number g such that

$$\forall w \in W\,, \quad J(u,w) \le g.$$

Clearly, the best (smallest) guaranteed performance for a given decision u is

$$G(u) = \sup_{w \in W} J(u,w)\,.$$

Hence the phrase "worst case", which has to do with a guarantee, not any malicious adversary.

Now, the problem of finding the best possible decision in this context is to find the smallest guaranteed performance, or

$$\inf_{u \in U} G(u) = \inf_{u \in U} \sup_{w \in W} J(u,w)\,. \tag{9.1}$$

If the infimum in u is reached, then the minimizing decision $u^\star$ deserves the name of optimal decision in that context.

2.1.2 Robust Disturbance Rejection or Gain Control

The following approach is mostly justified in the following linear setup, but can be extended to a nonlinear one (see van der Schaft 1996). We let U and W be normed vector spaces, Z be an auxiliary normed vector space, and $z \in Z$ be the output whose norm is to be kept small in spite of the disturbances w. We assume that for each decision $u \in U$, z depends linearly on w. Therefore, one has an (possibly nonlinear) application $P : U \to \mathcal{L}(W \to Z)$ and

$$z = P(u)w\,.$$

Clearly, z cannot be kept bounded if w is not. A natural formalization of the problem of keeping it small is to try and make the operator norm of $P(u)$ as small as possible. This may be expressed as follows: let $B(W)$ be the unit ball of W, one seeks

$$\inf_{u\in U}\ \sup_{w\in B(W)} \|P(u)w\|\,. \tag{9.2}$$

But we will often prefer another formulation, again in terms of guaranteed performance, which can easily be extended to nonlinear systems. Start from the problem of ensuring that $\|P(u)\| \le \gamma$ for a given positive *attenuation level* γ. This is equivalent to

$$\forall w \in W\,, \quad \|z\| \le \gamma\|w\|\,,$$

or equivalently (but leading to a smoother problem)

$$\forall w \in W\,, \quad \|z\|^2 \le \gamma^2\|w\|^2\,, \tag{9.3}$$

or equivalently again

$$\sup_{w\in W}[\|P(u)w\|^2 - \gamma^2\|w\|^2] \le 0\,.$$

Now, given a number γ, this has a solution (there exists a decision $u \in U$ satisfying that inequality) *if* the infimum hereafter is reached or is negative and *only if*

$$\inf_{u\in U}\sup_{w\in W}[\|P(u)w\|^2 - \gamma^2\|w\|^2] \le 0\,. \tag{9.4}$$

This is the method of $\mathcal{H}^\infty$-optimal control. Notice, however, that equation (9.3) has a meaning even for a nonlinear operator $P(u)$, so that the problem (9.4) is used in the so-called "nonlinear $\mathcal{H}^\infty$" control.

We will see a particular use of the control of the operator norm in feedback control when using the "small gain theorem." Although this could be set in the abstract context of nonlinear decision theory (see Vidyasagar 1993), we will rather show it in the more explicit context of control theory.

Remark 1. The three problems outlined above as equations (9.1), (9.2), and (9.4) are all of the form $\inf_u \sup_w$, and therefore, in a dynamic context, are amenable to dynamic game machinery. They involve no probabilistic description of the disturbances.

2.1.3 Set Description vs Probabilistic

Two main reasons have driven robust control toward set theoretic, rather than probabilistic, descriptions of disturbances. On the one hand, the traditional Gauss-Markov representation is geared toward high-frequency disturbances. When the main disturbance is, say, a misevaluated parameter in a dynamic equation, it is the exact opposite of high frequency: zero frequency. And this is typically the type of unknowns that robust control is meant to deal with.

On the other hand, concerning the coefficients of a differential equation, using a probabilistic description requires that one deals with differential equations with random coefficients, a much more difficult theory than that of differential equations driven by a stochastic process.

Thus, "robust control approaches" to decision problems involve a description of the unknown disturbances in terms of a set containing all possible disturbances, often a bounded set, and hence hard bounds on the possible disturbances. As mentioned in the introduction, systematic theories have been developed to deal with set descriptions of uncertainties under the name of *viability theory* (Aubin et al. 2011)—particularly well suited to deal with non-smooth problems in engineering, environment, economy, and finance (Bernhard et al. 2013, part V)—or *trajectory tubes* which, with the use of Hamiltonian formalism and elliptic approximations, recover more regularity and also lead to a theory of robust control in the presence of corrupted information (Kurzhanski and Varaiya 2014, Chap. 10). We emphasize here somewhat different, and more classical, methods, making explicit use of dynamic game theory.

To practitioners accustomed to Gauss-Markov processes (such as a filtered "white noise"), hard bounds may seem a rather severe limitation, providing the mathematical model with too much information. Yet, a probability law is an extremely rich information itself; in some sense, much richer than a simple set description. As an example, note that if we model a disturbance function as an ergodic stochastic process, then we tell the mathematical model that the longtime average of the disturbance is exactly known. In the case of a Brownian motion, we say that the total quadratic variation of (almost) all realizations is exactly known. And these are only instances of the rich information we provide, that the mathematical machinery will use, may be far beyond what was meant by the modeler.

2.2 Control Formulation

We want to apply the above ideas to control problems, i.e., problems where the decision to be taken is a control of a dynamic system. This raises the issues of causality and open-loop versus closed-loop control. No control theoretician would want to rely on open-loop control to overcome uncertainties and disturbances in a system. But then, the issue of the available information to form one's control becomes crucial.

2.2.1 Closed-Loop Control

Let a dynamic system be represented by a differential equation in $\mathbb{R}^n$ (we will therefore ignore the existing extensions of these results to infinite-dimensional systems; see Bensoussan and Bernhard (1993) for a game theoretic approach):

$$\dot{x} = f(t, x, u, w) \tag{9.5}$$

and two outputs, y the observed output and z to be controlled:

$$z(t) = g(t, x, u)\,, \tag{9.6}$$

$$y(t) = h(t, x, w)\,, \tag{9.7}$$

Here, as in the sequel, we have $x \in \mathbb{R}^n$, $u \in \mathsf{U} \subset \mathbb{R}^m$, $w \in \mathsf{W} \subset \mathbb{R}^\ell$, $y \in \mathbb{R}^p$, $z \in \mathbb{R}^q$, and $u(\cdot) \in \mathcal{U}$, $w(\cdot) \in \mathcal{W}$ the sets of measurable time functions into U and W, respectively. Likewise, we let $y(\cdot) \in \mathcal{Y}$ and $z(\cdot) \in \mathcal{Z}$. Moreover, we assume that f enjoys regularity and growth properties that insure existence and uniqueness of the solution to (9.5) for any initial condition and any pair of controls $(u(\cdot), w(\cdot)) \in \mathcal{U} \times \mathcal{W}$.

Since we are concerned with causality, we introduce the following notation: For any time function $v(\cdot)$, let v^t denote its restriction $[v(s)|s \le t]$. And if $\mathcal{V}$ is the set of functions $v(\cdot)$, let $\mathcal{V}^t$ be the set of their restrictions v^t:

$$\begin{gathered}\mathcal{V} = \{v(\cdot) : [t_0, t_1] \to \mathsf{V} : s \mapsto v(s)\} \Longrightarrow \\ \forall t \in (t_0, t_1)\,,\ \mathcal{V}^t = \{v^t : [t_0, t] \to \mathsf{V} : s \mapsto v(s)\}\,.\end{gathered} \tag{9.8}$$

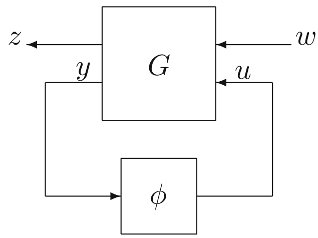
The control $u(t)$ will be synthesized as a closed-loop control

$$u(t) = \phi(t, y^t)\,, \tag{9.9}$$

typically by a dynamic compensator driven by the observed output y. Therefore, the operator $\inf_{u \in U}$ in equations (9.1), (9.2), and (9.4) must be replaced by $\inf_{\phi \in \Phi}$. But then, we must specify the class Φ of admissible controllers. However, we postpone that discussion until after we deal with the disturbance w.

The question arises as to whether one should allow closed-loop "disturbance laws," some $w(t) = \psi(t, u^t)$, or be content with open-loop disturbances $w(\cdot) \in \mathcal{W}$. This question is answered by the following lemma, occasionally attributed to Leonard Berkowitz (1971):

Fig. 9.1 Disturbance rejection problem



Lemma 1. *Let a criterion* $J(u(\cdot), w(\cdot)) = K(x(\cdot), u(\cdot), w(\cdot))$ *be given. Let a control function* ϕ *be given and* Ψ *be a class of closed-loop disturbance strategies* compatible with ϕ *(i.e., such that the differential equations of the dynamic system have a solution when* u *is generated by* ϕ *and* w *by any* $\psi \in \Psi$*). Then, (with a transparent abuse of notation)*

$$\sup_{w(\cdot)\in\mathcal{W}} J(\phi, w(\cdot)) \geq \sup_{\psi\in\Psi} J(\phi, \psi) .$$

Therefore, it is never necessary to consider closed-loop disturbance laws. This greatly simplifies the discussion of the class Φ of admissible controllers. It only has to be such that solutions to the differential equations exist against all open-loop measurable disturbances. This is an important difference with "true" (dynamic) game theory, arising from the fact that we are not concerned with the existence of a Value, but only with inf sup operators.

The control form of the disturbance rejection problem of Sect. 2.1.2 may therefore be stated, for the system G given by (9.5), (9.6), (9.7), and (9.9) over a time span $\mathcal{T}$, as follows: let Φ be the class of all causal controllers ensuring the existence in $L^2(\mathcal{T})$ and uniqueness of the solution of the system equations (Fig. 9.1).

Standard problem of $\mathcal{H}^\infty$-optimal control with attenuation level γ: *Does the following inequality hold:*

$$\inf_{\phi\in\Phi} \sup_{w\in L^2(\mathcal{T})} \left(\|z\|^2_{L^2} - \gamma^2\|w\|^2_{L^2}\right) \leq 0\,? \tag{9.10}$$

If the infimum is reached or is negative, find an admissible controller ϕ *ensuring the inequality* (9.3) *in* L^2 *norms.*

2.2.2 Minimax Certainty Equivalence Principle

The standard problem as formulated above amounts to a min-sup differential game with a real-time minimizer's information both partial and corrupted by the maximizer, a nonclassical problem. In some favorable cases, including that of linear systems, the information problem can be solved via the following certainty equivalence theorem (Başar and Bernhard 1995; Bernhard and Rapaport 1996). Let a differential game be specified by (9.5) and a criterion:

$$J(x_0; u(\cdot), w(\cdot)) = M(x(t_1)) + \int_{t_0}^{t_1} L(t, x(t), u(t), w(t))\, \mathrm{d}t + N(x_0)\,. \qquad (9.11)$$

Assume that the full state information min-sup problem with $N \equiv 0$ has a state feedback solution:

$$u(t) = \varphi^\star(t, x(t))\,, \qquad (9.12)$$

leading to a Value function $V(t,x)$ of class C^1.

We now define a partial information mechanism. Since we assume that the minimizer does not know $x(t)$, it is consistent to assume that he does not know x_0 either. We therefore allow, in the partial information problem, the added cost $N(x_0)$. Let $\omega \in \Omega = \mathbb{R}^n \times \mathcal{W}$ be the complete disturbance, i.e., the pair $(x_0, w(\cdot))$. Recall notation (9.8). An *observation process* is a device that, at each time instant t, defines a subset $\Omega_t \subset \Omega$ function of $u(\cdot)$ and ω, which enjoys the following three properties:

1. It is *consistent*: $\forall u \in \mathcal{U}\,,\ \forall \omega \in \Omega\,,\ \forall t\,,\ \omega \in \Omega_t$.
2. It is *perfect recall*: $\forall u \in \mathcal{U}\,,\ \forall \omega \in \Omega\,,\ \forall t' \geq t\,,\ \Omega_{t'} \subset \Omega_t$.
3. It is *nonanticipative*: $\omega^t \in \Omega_t^t \Rightarrow \omega \in \Omega_t$.

In the case of a system (9.5), (9.6), and (9.7),

$$\Omega_t(u(\cdot), \omega) = \{(x_0, w(\cdot)) \mid \forall s \leq t\,,\ h(s, x(s), w(s)) = y(s)\}.$$

We seek a controller

$$u(t) = \phi(t, \Omega_t)\,. \qquad (9.13)$$

Define the *auxiliary criterion* :

$$G_t(u^t, \omega^t) = V(t, x(t)) + \int_{t_0}^{t} L(s, x(s), u(s), w(s))\, \mathrm{d}s + N(x_0)\,,$$

and the *auxiliary problem*:

$$\max_{\omega^t \in \Omega_t^t} G_t(u^t, \omega^t)\,.$$

Note that in this problem, u^t is known, as our own past control.

Theorem 1. *If the auxiliary problem admits one or several solutions leading to a* unique *state $\hat{x}(t)$ at time t, then, if the controller*

$$u(t) = \varphi^\star(t, \hat{x}(t))$$

is admissible, it is a min-sup controller in partial information.

Conversely, if there exists a time t *such that* $\sup_{\omega^t \in \Omega_t^t} G_t = \infty$*, then the criterion* (9.11) *has an infinite supremum in* ω *for any admissible feedback controller* (9.13).

Remark 2. The state $\hat{x}(t)$ may be considered as the *worst possible state* given the available information.

One way to proceed in the case of an information such as (9.7) is to solve by forward dynamic programming (forward Hamilton-Jacobi-Caratheodory-Bellman equation) the constrained problem:

$$\max_{\omega} \left[\int_{t_0}^{t} L(s, x(s), u(s), w(s))\, \mathrm{d}s + N(x_0) \right]$$

subject to the control constraint $w(s) \in \{w \mid h(s, x(s), w) = y(s)\}$ and the terminal constraint $x(t) = \xi$. It is convenient to call $-W(t, \xi)$ the corresponding Value (or Bellman function). Assuming that, under the strategy $\varphi^\star$, for every t, the whole space is reachable by some $\omega \in \Omega$, $\hat{x}(t)$ is obtained via

$$\max_{x \in \mathbb{R}^n} [V(t, x) - W(t, x)] = V(t, \hat{x}(t)) - W(t, \hat{x}(t))\,. \tag{9.14}$$

If this max is reached at a unique $x = \hat{x}(t)$, then the uniqueness condition of the theorem is satisfied.

Remark 3. In the duality between probability and optimization (see Akian et al. 1998; Baccelli et al. 1992), the function $-W(t, \cdot)$ is the *conditional cost measure* of the state for the measure $\int_{t_0}^{t} L\mathrm{d}s + N(x_0)$ knowing $y(\cdot)$. The left-hand side of formula (9.14) is the dual of a mathematical expectation.

2.2.3 Small Gain Theorem

We aim to show classical linear control problems that can be cast into a standard problem of $\mathcal{H}^\infty$-optimal control. We develop some preliminary tools.

Linear operators and norms

In the case of linear systems, operator norms have concrete forms.
Matrix: A $p \times m$ matrix M represents a linear operator $\mathcal{M} : \mathbb{R}^m \to \mathbb{R}^p$ whose norm $\|\mathcal{M}\|$ is the maximum singular value $\sigma_{\max}(M)$ of the matrix.
Dynamic system: A stable stationary linear dynamic system with input $u(t) \in \mathbb{R}^m$ and output $y(t) \in \mathbb{R}^p$ will be considered as a linear operator, say $\mathcal{G}$, from $L^2(\mathbb{R} \to \mathbb{R}^m)$ to $L^2(\mathbb{R} \to \mathbb{R}^p)$. It may be represented by its transfer function $G(s)$, always a rational proper matrix since we confine ourselves to finite-dimensional state systems. Moreover, the system being stable (otherwise it would not map L^2 into L^2), the transfer function has all its poles in the left half complex plane. Thus it belongs to the Hardy space $\mathcal{H}^\infty$ of functions holomorphic in an open set of the

complex plane containing the half-plane $\mathrm{Re}(s) \geq 0$. In that case, it follows from Parseval's equality that

$$\|\mathcal{G}\| = \sup_{\omega\in\mathbb{R}} \sigma_{\max} G(\mathrm{j}\omega) =: \|G(\cdot)\|_\infty \,.$$

The norm $\|G(\cdot)\|_\infty$ (or $\|G\|_\infty$) is the norm of the transfer function in $\mathcal{H}^\infty$.
Block operator: If the input and output spaces of a linear operator $\mathcal{G}$ are represented as product of two spaces each, the operator takes a block form

$$\mathcal{G} = \begin{pmatrix} \mathcal{G}_{11} & \mathcal{G}_{12} \\ \mathcal{G}_{21} & \mathcal{G}_{22} \end{pmatrix}.$$

We will always consider the norm of product spaces as the Euclidean combination of the norms in the component spaces. In that case, it holds that $\|\mathcal{G}_{ij}\| \leq \|\mathcal{G}\|$. Furthermore, whenever the following operators are defined

$$\left\| \begin{pmatrix} \mathcal{G}_1 \\ \mathcal{G}_2 \end{pmatrix} \right\| \leq \sqrt{\|\mathcal{G}_1\|^2 + \|\mathcal{G}_2\|^2}\,, \qquad \|(\mathcal{G}_1 \; \mathcal{G}_2)\| \leq \sqrt{\|\mathcal{G}_1\|^2 + \|\mathcal{G}_2\|^2}\,.$$

Hence

$$\left\| \begin{pmatrix} \mathcal{G}_{11} & \mathcal{G}_{12} \\ \mathcal{G}_{21} & \mathcal{G}_{22} \end{pmatrix} \right\| \leq \sqrt{\|\mathcal{G}_{11}\|^2 + \|\mathcal{G}_{12}\|^2 + \|\mathcal{G}_{21}\|^2 + \|\mathcal{G}_{22}\|^2}\,.$$

We finally notice that for a block operator $\mathcal{G} = (\mathcal{G}_{ij})$, its norm is related to the matrix norm of its matrix of norms $(\|\mathcal{G}_{ij}\|)$ by $\|\mathcal{G}\| \leq \|(\|\mathcal{G}_{ij}\|)\|$.

Small gain theorem

Let a two input-two output linear system G_0 be given by

$$\begin{pmatrix} z \\ y \end{pmatrix} = \begin{pmatrix} \mathcal{G}_{11} & \mathcal{G}_{10} \\ \mathcal{G}_{01} & \mathcal{G}_{00} \end{pmatrix} \begin{pmatrix} w \\ u \end{pmatrix}. \tag{9.15}$$

Assume that it is connected in feedback by a linear operator Δ according to $w = \Delta z$, leading to a composite system $y = \mathcal{G}_\Delta u$. (See Fig. 9.2.)

Then, an easy consequence of Banach's fixed point theorem is

Theorem 2 (Small gain theorem). *If* $\|\Delta\mathcal{G}_{11}\| = \alpha < 1$ *or* $\|\mathcal{G}_{11}\Delta\| = \alpha < 1$, *then the combined system*

$$\mathcal{G}_\Delta = \mathcal{G}_{01}(I - \Delta\mathcal{G}_{11})^{-1}\Delta\mathcal{G}_{10} + \mathcal{G}_{00} \quad \text{or} \quad \mathcal{G}_\Delta = \mathcal{G}_{01}\Delta(I - \mathcal{G}_{11}\Delta)^{-1}\mathcal{G}_{10} + \mathcal{G}_{00}$$

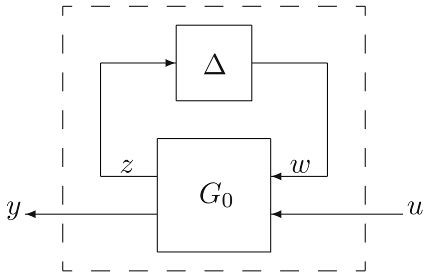


Fig. 9.2 The perturbed system $\mathcal{G}_\Delta$

is well defined and is stable. Moreover, it holds that

$$\|\mathcal{G}_\Delta\| \le \|\mathcal{G}_{00}\| + \|\mathcal{G}_{01}\|\,\|\mathcal{G}_{10}\|\frac{\|\Delta\|}{1-\alpha}\,.$$

Corollary 1. *If* $\|\mathcal{G}_{11}\| < 1$*, then* $\mathcal{G}_\Delta$ *is stable for any* Δ *with* $\|\Delta\| \le 1$.

This will be a motivation to try and solve the problem of making the norm of an operator smaller than 1, or smaller than a given number γ.

2.3 Robust Servomechanism Problem

2.3.1 Model Uncertainty

Assume we deal with a linear system

$$\dot{x} = Ax + Bu\,,$$
$$y = Cx + Du\,.$$

But the system matrices are not precisely known. We have an estimate or "nominal system" (A_0, B_0, C_0, D_0), and we set $A = A_0 + \Delta_A$, $B = B_0 + \Delta_B$, $C = C_0 + \Delta_C$, and $D = D_0 + \Delta_D$. The four matrices Δ_A, Δ_B, Δ_C, and Δ_D are unknown, but we know bounds δ_A, δ_B, δ_C, and δ_D on their respective norms. We rewrite the same system as

$$\dot{x} = A_0x + B_0u + w_1\,,$$
$$y = C_0x + D_0y + w_2\,,$$
$$z = \begin{pmatrix} x \\ u \end{pmatrix},$$

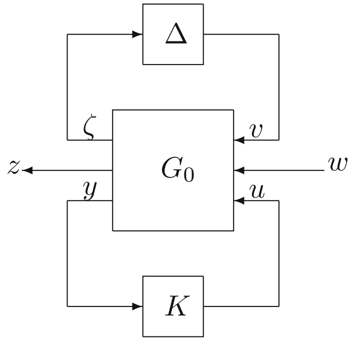


Fig. 9.3 The perturbed system and its compensator

and obviously

$$w = \begin{pmatrix} w_1 \\ w_2 \end{pmatrix} = \begin{pmatrix} \Delta_A & \Delta_B \\ \Delta_C & \Delta_D \end{pmatrix} \begin{pmatrix} x \\ u \end{pmatrix} = \Delta z \,.$$

We now have a known system G_0 connected in feedback by the unknown system Δ for which we have a norm bound $\|\Delta\| \le (\delta_A^2 + \delta_B^2 + \delta_C^2 + \delta_D^2)^{1/2}$, to which we can apply the small gain theorem.

2.3.2 Robust Stabilization

Indeed, that representation of a partially unknown system can be extended to a dynamic unknown perturbation Δ. We rename w and z as v and ζ, respectively, and we add a disturbance input w, an output z to be controlled, and a dynamic compensator $u = Ky$ (Fig. 9.3). From now on, to make things simpler, we assume that the signals v and ζ are scalar (see below a short discussion of vector signals). Assume that a frequency-dependent bound of the modulus of Δ's transfer function is known as $|\Delta(\mathrm{j}\omega)| \le \delta(\omega)$. Then, in a classical "loop-shaping" fashion, we devise a dynamic filter $W_1(s)$, such that $|W_1(\mathrm{j}\omega)| \ge \delta(\omega)$, and consider the fictitious control system with output $\tilde{\zeta} = W_1\zeta$. Stabilization of the system will be guaranteed if we can keep the transfer function norm from v to $\tilde{\zeta}$ less than one. We may also specify the disturbance rejection objective as keeping a fictitious output $\tilde{z} = W_0 z$ small for some prescribed dynamic filter W_0. The control objective of the compensator K will now be to keep the modulus of the transfer function from v to $\tilde{\zeta}$ less than one at all frequencies, while holding the norm of the transfer function from w to $\tilde{z}$ as small as possible. This may be cast into a standard problem, for instance, by choosing the pair $((1/\gamma)v, w)$ as the input and $(\tilde{\zeta}, \beta\tilde{z})$ as the output and finding β and γ as small as possible while holding inequality (9.10) true. But more clever weightings may also be tried.

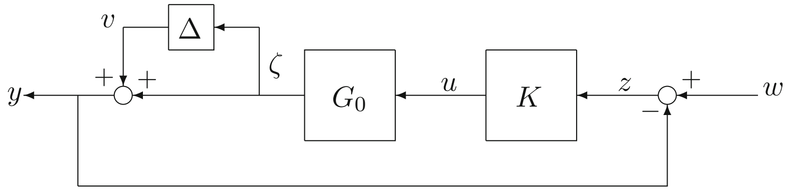


Fig. 9.4 The servomechanism

Some remarks are in order.

Remark 4.

1. We might as well have placed the shaping filter W_1 on the input channel v. The two resulting control problems are not equivalent. One may have a solution and the other one none. Furthermore, the weight $\delta(\omega)$ might be divided between two filters, one on the input and one on the output. The same remark applies to the filter W_0 and to the added weights β and γ. There is no known simple way to arbitrate these possibilities.
2. In case of vector signals v and ζ, one can use diagonal filters with the weight W_1 on each channel. But this is often inefficient. There exists a more elaborate way to handle that problem, called "μ-synthesis." (See Doyle 1982; Zhou et al. 1996.)

2.3.3 Robust Servomechanism

We describe a particular case of the above problem which was at the inception of the $\mathcal{H}^\infty$-optimal control problem in Zames (1981). All signals are scalar.

An uncertain system's transfer function G is described by a multiplicative uncertainty: $G = (I + \Delta)G_0$. We know G_0 and a bound $|\Delta(\mathrm{j}\omega)| \le \delta(\omega)$. However, it holds that as $\omega \to \infty$, $G_0(\mathrm{j}\omega) \to 0$, and because of a possible fixed offset in the true plant, $\delta(\omega) \to \infty$. The aim of the control system is to have the output y follow an unpredictable, but low frequency, reference signal w thanks to a dynamic compensator $u = K(w - y)$. We name z the error signal $w - y$, ζ the (fictitious) output of the (fictitious) system G_0, and $v = \Delta\zeta$. (See Fig. 9.4.) We can cast this problem as a standard problem with input (v, w) and output (ζ, z). However, adding frequency-dependent weights is now unavoidable. We have

$$
\begin{aligned}
z &= -G_0 u - v + w\,,\\
\zeta &= G_0 u\,,\\
u &= K z\,,
\end{aligned}
$$

hence

$$z = S(w - v)\,, \qquad S = (I + G_0 K)^{-1}\,,$$
$$\zeta = T(w - v)\,, \qquad T = G_0 K (I + G_0 K)^{-1}\,.$$

The control aim is to keep the *sensitivity transfer function* S small, while robust stability requires that the *complementary sensitivity transfer function* T be small. However, it holds that $S + T = I$, and hence, both cannot be kept small simultaneously. We need that

$$\forall \omega \in \mathbb{R}\,, \qquad \delta(\omega)|T(\mathrm{j}\omega)| \le 1\,.$$

And if $\delta(\omega) \to \infty$ as $\omega \to \infty$, this imposes that $|T(\mathrm{j}\omega)| \to 0$, and therefore $|S(\mathrm{j}\omega)| \to 1$. Therefore we cannot follow a high-frequency reference input w. However, it may be possible to keep $|S(\mathrm{j}\omega)|$ small at low frequencies, where $\delta(\omega)$ is small. The solution is to work with dynamic filters and fictitious outputs $\tilde{z} = W_0 z$, $\tilde{\zeta} = W_1 \zeta$, with $|W_1(\mathrm{j}\omega)| \ge \delta(\omega)$ (a generalized differentiator). If we can keep the transfer function from v to $\tilde{\zeta}$ less than one at all frequencies, we ensure stability of the control system. And we choose a weight W_0 large at low frequencies (a generalized integrator) and investigate the standard problem with the output $(\tilde{z}, \tilde{\zeta})$.

If $W_0 G_0$ is strictly proper, as it will usually be, there is no throughput from u to the output $\tilde{z}$. In that case, it is necessary to add a third component $\tilde{u} = W_3 u$ to the regulated output, with W_3 proper but not strictly (may be a multiplicative constant R). This is to satisfy the condition $R > 0$ of the next section.

3 $\mathcal{H}^\infty$-Optimal Control

Given a linear system with inputs u and w and outputs y and z, and a desired attenuation level γ, we want to know whether there exist causal control laws $u(t) = \phi(t, y^t)$ guaranteeing inequality (9.3) and, if yes, find one. This is the *standard problem* of $\mathcal{H}^\infty$-optimal control. We propose here an approach of this problem based upon dynamic game theory, following Başar and Bernhard (1995). Others exist. See Doyle et al. (1989), Stoorvogel (1992), and Glover (2015).

We denote by an accent the transposition operation.

3.1 Continuous Time

We start with a state space description of the system. Here $x \in \mathbb{R}^n$, $u \in \mathbb{R}^m$, $w \in \mathbb{R}^\ell$, $y \in \mathbb{R}^p$, and $z \in \mathbb{R}^q$. This prescribes the dimensions of the various matrices in the system equations. In the finite horizon problem, they may all be time dependent, say piecewise continuously. These equations are

$$\dot{x} = Ax + Bu + Dw\,, \quad x(t_0) = x_0\,, \tag{9.16}$$

$$y = Cx \qquad + Ew\,, \tag{9.17}$$

$$z = Hx + Gu\,. \tag{9.18}$$

We will use the *system matrix*

$$\mathcal{S} = \begin{pmatrix} A & B & D \\ C & 0 & E \\ H & G & 0 \end{pmatrix}. \tag{9.19}$$

Remark 5.

1. The fact that "the same" input w drives the dynamics (9.16) and corrupts the output y in (9.17) is not a restriction. Indeed different components of w may enter the different equations. (Then $DE' = 0$.)
2. The fact that we allowed no throughput from u to y is not restrictive. y is the measured output used to control the system. If there were a term $+Fu$ on the r.h.s. of (9.17), we could always use $\tilde{y} = y - Fu$ as measured output.
3. A term $+Fw$ in (9.18) would create cross terms in wx and wu in the linear quadratic differential game to be solved. The problem would remain feasible, but the equations would be more complicated, which we would rather avoid.

We further let

$$\begin{pmatrix} H' \\ G' \end{pmatrix} (\, H \;\; G\,) = \begin{pmatrix} Q & P \\ P' & R \end{pmatrix} \text{ and } \begin{pmatrix} D \\ E \end{pmatrix} (\, D' \;\; E'\,) = \begin{pmatrix} M & L' \\ L & N \end{pmatrix}. \tag{9.20}$$

All the sequel makes use of the following hypotheses:

Assumption 1.

1. *G is one to one or, equivalently, $R > 0$,*
2. *E is onto or, equivalently, $N > 0$.*

Duality

Notice that changing $\mathcal{S}$ into its transpose $\mathcal{S}'$ swaps the two block matrices in (9.20) and also the above two hypotheses. This operation (together with reversal of time), that we will encounter again, is called duality.

3.1.1 Finite Horizon

As in Sect. 2.2.2, we include initial and terminal costs. For any symmetric nonnegative matrix Z, we write $\|x\|_Z^2 := x'Zx$ and likewise for other vectors. Let Q_0 be a symmetric positive definite matrix and Q_1 a nonnegative definite one, and

$$J_\gamma(x_0, u(\cdot), w(\cdot)) = \|x(t_1)\|^2_{Q_1} +$$
$$\int_{t_0}^{t_1} \left[(\, x'(t)\; u'(t)\,) \begin{pmatrix} Q & P \\ P' & R \end{pmatrix} \begin{pmatrix} x(t) \\ u(t) \end{pmatrix} - \gamma^2 \|w(t)\|^2 \right] \mathrm{d}t - \gamma^2 \|x_0\|^2_{Q_0} \,. \tag{9.21}$$

We introduce two Riccati equations for the symmetric matrices S and Σ, where we use the feedback gain F and the Kalman gain K appearing in the LQG control problem:

$$F = -R^{-1}(B'S + P')\,, \qquad K = (\Sigma C' + L')N^{-1}\,, \tag{9.22}$$

$$-\dot S = SA + A'S - F'RF + \gamma^{-2} SMS + Q\,, \qquad S(t_1) = Q_1\,, \tag{9.23}$$
$$\dot\Sigma = \Sigma A' + A\Sigma - KNK' + \gamma^{-2}\Sigma Q \Sigma + M\,, \qquad \Sigma(t_0) = Q_0^{-1}\,. \tag{9.24}$$

Theorem 3. *For a given γ, if both equations (9.23) and (9.24) have solutions over $[t_0, t_1]$ satisfying*

$$\forall t \in [t_0, t_1]\,, \quad \rho(\Sigma(t)S(t)) \le \gamma^2\,, \tag{9.25}$$

the finite horizon standard problem has a solution for that value of γ and any larger one. In that case, a (nonunique) min-sup controller is given by $u = u^\star$ as defined by either of the following two systems:

$$\dot{\hat x} = (A + \gamma^{-2}MS)\hat x + Bu^\star + (I - \gamma^{-2}\Sigma S)^{-1}K[y - (C + \gamma^{-2}LS)\hat x], \;\; \hat x(t_0) = 0\,,$$
$$u^\star = F\hat x\,,$$

or

$$\dot{\check x} = (A + \gamma^{-2}\Sigma Q)\check x + (B + \gamma^{-2}\Sigma P)u^\star + K(y - C\check x)\,, \qquad \check x(t_0) = 0\,,$$
$$u^\star = F(I - \gamma^{-2}\Sigma S)^{-1}\check x\,.$$

If any one of the above conditions fails to hold, then, for any smaller γ, the criterion (9.21) has an infinite supremum for any causal controller $u(t) = \phi(t, y^t)$.

Remark 6.

1. The notation $\rho(X)$ stands for the spectral radius of the matrix X. If condition (9.25) holds, then, indeed, the matrix $(I - \gamma^{-2}\Sigma S)$ is invertible.

2. The two Riccati equations are dual of each other as in the LQG control problem.
3. The above formulas coincide with the LQG formulas in the limit as $\gamma \to \infty$.
4. The first system above is the "certainty equivalent" form. The second one follows from placing $\check{x} = (I - \gamma^{-2}\Sigma S)\hat{x}$. The symmetry between these two forms seems interesting.
5. The solution of the $\mathcal{H}^\infty$-optimal control problem is highly nonunique. The above one is called the *central controller*. But use of Başar's representation theorem (Başar and Olsder 1982) yields a wide family of admissible controllers. See Başar and Bernhard (1995).

3.1.2 Infinite Horizon Stationary Problem

In this section, we assume that the matrix $\mathcal{S}$ is a constant. The system is then said to be stationary. The spaces L^2 considered are to be understood as $L^2(\mathbb{R} \to \mathbb{R}^d)$ with the appropriate dimension d. The dynamic equation (9.16) is to be understood with zero initial condition at $-\infty$, and we will be interested in asymptotically stable solutions. There is no room for the terms in Q_0 and Q_1 of the finite time criterion (9.21), and its integral is to be taken from $-\infty$ to ∞.

We further invoke a last pair of dual hypotheses:

Assumption 2. *The pair (A, D) is stabilizable and the pair (A, H) is detectable.*

The Riccati equations are replaced by their stationary variants, still using (9.22):

$$SA + A'S - F'RF + \gamma^{-2}SMS + Q = 0\,, \tag{9.26}$$

$$\Sigma A' + A\Sigma - KNK' + \gamma^{-2}\Sigma Q\Sigma + M = 0\,. \tag{9.27}$$

Theorem 4. *Under condition 2, if the two algebraic Riccati equations (9.26) and (9.27) have positive definite solutions, the minimal such solutions $S^\star$ and $\Sigma^\star$ can be obtained as the limit of the Riccati equations (9.23) when integrating from $S(0) = 0$ backward and, respectively, (9.24) when integrating from $\Sigma(0) = 0$ forward. If these solutions satisfy the condition $\rho(\Sigma^\star S^\star) < \gamma^2$, then the same formula as in Theorem 3 replacing S and Σ by $S^\star$ and $\Sigma^\star$, respectively, provides a solution to the stationary standard problem. If moreover the pair (A, B) is stabilizable and the pair (A, C) is detectable, there exists such a solution for sufficiently small γ.*

If the existence or the spectral radius condition fails to hold, there is no solution to the stationary standard problem for any smaller γ.

The condition of positive definiteness of $S^\star$ and Σ^* can be slightly weakened, leading to a slightly more precise theorem (see Başar and Bernhard 1995). But this does not seem to be very useful in practice.

3.2 Discrete Time

We consider now the discrete-time system where $t \in \mathbb{N}$:

$$x_{t+1} = A_t x_t + B_t u_t + D_t w_t \,, \quad x_{t_0} = x_0 \tag{9.28}$$

$$y_t = C_t x_t \qquad\qquad + E_t w_t \,, \tag{9.29}$$

$$z_t = H_t x_t + G_t u_t \,. \tag{9.30}$$

where the system matrices may depend on the time t. We use notation (9.20), still with assumption 1 for all t, and

$$\bar{A}_t = A_t - B_t R_t^{-1} P_t' \,, \qquad \widetilde{A}_t = A_t - L_t' N_t^{-1} C_t \,. \tag{9.31}$$

We also invoke all along the following two dual hypotheses:

Assumption 3.

$$\forall t \,, \quad \operatorname{rank} \begin{pmatrix} A_t \\ H_t \end{pmatrix} = n \,, \qquad \operatorname{rank} (\, A_t \;\; D_t \,) = n \,.$$

We want to control the system with a *strictly causal* controller

$$u_t = \phi_t(t, y^{t-1}) \,.$$

(See Başar and Bernhard (1995) for a nonstrictly causal controller or delayed information controllers.)

3.2.1 Finite Horizon

We introduce two positive definite symmetric $n \times n$ matrices X and Y. The augmented criterion is now

$$J = \|x_{t_1}\|_X^2 + \sum_{t=t_0}^{t_1-1} (\|z_t\|^2 - \gamma^2 \|w_t\|^2) - \gamma^2 \|x_0\|_Y^2 \,. \tag{9.32}$$

We will not attempt to describe here the (nonlinear) discrete-time certainty equivalence theorem used to solve this problem. (See Başar and Bernhard 1995; Bernhard 1994.) We go directly to the solution of the standard problem.

The various equations needed may take quite different forms. We choose one. We need the following notation (note that Γ_t and $\bar{S}_t$ involve S_{t+1}):

$$\begin{aligned}
\Gamma_t &= (S_{t+1}^{-1} + B_t R_t^{-1} B_t' - \gamma^{-2} M_t)^{-1}\,, \\
\bar{S}_t &= \bar{A}_t' (S_{t+1} - \gamma^{-2} M_t)^{-1} \bar{A}_t + Q_t - P_t R_t^{-1} P_t'\,, \\
\Delta_t &= (\Sigma_t^{-1} + C_t' N_t^{-1} C_t - \gamma^{-2} Q_t)^{-1}\,. \\
\widetilde{\Sigma}_{t+1} &= \widetilde{A}_t (\Sigma_t^{-1} - \gamma^{-2} Q_t)^{-1} \widetilde{A}_t' + M_t - L_t' N_t^{-1} L_t\,.
\end{aligned} \tag{9.33}$$

The two discrete Riccati equations may be written as

$$S_t = \bar{A}_t' \Gamma_t \bar{A}_t + Q_t - P_t R_t^{-1} P_t'\,, \qquad S_{t_1} = X\,, \tag{9.34}$$

$$\Sigma_{t+1} = \widetilde{A}_t \Delta_t \widetilde{A}_t' + M_t - L_t' N_t^{-1} L_t\,, \qquad \Sigma_{t_0} = Y^{-1}\,. \tag{9.35}$$

Theorem 5. *Under the hypothesis 3, if both discrete Riccati equations (9.34) and (9.35) have solutions satisfying either $\rho(M_t S_{t+1}) < \gamma^2$ and $\rho(\widetilde{\Sigma}_{t+1} S_{t+1}) < \gamma^2$ or $\rho(\Sigma_t Q_t) < \gamma^2$ and $\rho(\Sigma_t \bar{S}_t) < \gamma^2$, then the standard problem has a solution for that value of γ and any larger one, given by*

$$u_t^\star = -R_t^{-1} (B_t' \Gamma_t \bar{A}_t + P_t')(I - \gamma^{-2} \Sigma_t S_t)^{-1} \check{x}_t\,, \tag{9.36}$$

$$\check{x}_{t+1} = A_t \check{x}_t + B_t u_t^\star + \gamma^{-2} \widetilde{A}_t \Delta_t (Q_t \check{x}_t + P_t u_t^\star) + (\widetilde{A}_t \Delta_t C_t' + L_t') N_t^{-1} (y_t - C_t \check{x}_t), \tag{9.37}$$

$$\check{x}_{t_0} = 0\,. \tag{9.38}$$

If any one of the above conditions fails to hold, then for any smaller γ, the criterion (9.32) has an infinite supremum for any strictly causal controller.

Remark 7.

1. Equation (9.37) is a *one-step predictor* allowing one to get the *worst possible state* $\hat{x}_t = (I - \gamma^{-2} \Sigma_t S_t)^{-1} \check{x}_t$ as a function of past y_s up to $s \le t-1$.
2. Controller (9.36) is therefore a strictly causal, *certainty equivalent* controller.

3.2.2 Infinite Horizon Stationary Problem

We consider the same system as above, with all system matrices constant and with the strengthened hypothesis (as compared to Assumption 2):

Assumption 4. *The pair (A, D) is controllable and the pair (A, H) is observable.*

The dynamics is to be understood with zero initial condition at $-\infty$, and we seek an asymptotically stable and stabilizing controller. The criterion (9.32) is replaced by a sum from $-\infty$ to ∞, thus with no initial and terminal terms.

The stationary versions of all the equations in the previous subsection are obtained by removing the index t or $t+1$ to all matrix-valued symbols.

Theorem 6. *Under Assumption 4, if the stationary versions of* (9.34) *and* (9.35) *have positive definite solutions, the smallest such solutions $S^\star$ and $\Sigma^\star$ are obtained as, respectively, the limit of S_t when integrating* (9.34) *backward from $S_0 = 0$ and the limit of Σ_t when integrating* (9.35) *forward from $\Sigma_0 = 0$. If these limit values satisfy either of the two spectral conditions in Theorem 5, the standard problem has a solution for that value of γ and any larger one, given by the stationary versions of equations* (9.36) *and* (9.37).

If any one of these conditions fails to hold, the supremum of J is infinite for all strictly causal controllers.

4 Nonlinear Problems

4.1 Nonlinear $\mathcal{H}^\infty$ Control

There are many different ways to extend $\mathcal{H}^\infty$-optimal control theory to nonlinear systems. A driving factor is how much we are willing to parametrize the system and the controller. If one restricts its scope to a parametrized class of system equations (see below) and/or to a more or less restricted class of compensators (e.g., finite dimensional), then some more explicit results may be obtained, e.g., through special Lyapunov functions and LMI approaches (see Coutinho et al. 2002; El Ghaoui and Scorletti 1996), or passivity techniques (see Ball et al. 1993).

In this section, partial derivatives are denoted by indices.

4.1.1 A Finite Horizon Problem

As an example, we specialize Didinsky et al. (1993) for a problem slightly more natural than the standard problem (9.10) in a nonlinear setup (see van der Schaft 1996), i.e., finding a control law such that the controlled system has *finite L^2-gain*:

Definition 1. A system $\omega \mapsto z$ is said to have L^2-gain less than or equal to γ if there exists a number β such that

$$\forall \omega \in \Omega\,, \quad \|z\|^2 \le \gamma^2 \|\omega\|^2 + \beta^2\,. \tag{9.39}$$

We consider a nonlinear system defined by

$$\begin{aligned} \dot{x} &= a(t,x,u) + b(t,x,u)w\,, \quad x(0) = x_0\,,\\ z(t) &= g(t,x,u)\,,\\ y(t) &= h(t,x) + v\,. \end{aligned}$$

We work over the time interval $[0, T]$. We use three symmetric matrices, R positive definite and P and Q nonnegative definite, and take as the norm of the disturbance:

$$\|\omega\|^2 = \int_0^T \left(\|w(t)\|_R^2 + \|v(t)\|_P^2\right) \mathrm{d}t + \|x_0\|_Q^2 \,.$$

Let also

$$B(t,x,u) := \frac{1}{4} b(t,x,u) R^{-1} b'(t,x,u) \,.$$

We use the following two HJCIB equations, where $\hat{\varphi} : [0,T] \times \mathbb{R}^n \to \mathsf{U}$ is a state feedback and $\hat{u}(t)$ stands for the control actually used:

$$\begin{aligned} &V_t(t,x) + V_x(t,x) a(t,x,\hat{\varphi}(t,x)) + \gamma^{-2} V_x(t,x) B(t,x,\hat{\varphi}(t,x)) V_x'(t,x) \\ &+ \|g(t,x,\hat{\varphi}(t,x))\|^2 = -p(t,x) \,, \qquad V(T,x) = 0. \end{aligned} \tag{9.40}$$

$$\begin{aligned} &W_t(t,x) + W_x(t,x) a(t,x,\hat{u}(t)) + \gamma^{-2} W_x(t,x) B(t,x,\hat{u}(t)) W_x'(t,x) \\ &+ \|g(t,x,\hat{u}(t))\|^2 = \gamma^2 \|y(t) - h(t,x)\|^2 \,, \qquad W(0,x) = \gamma^2 \|x\|_Q^2. \end{aligned} \tag{9.41}$$

Theorem 7.

1. *If there exists an admissible state feedback $\hat{\varphi}$, a nonnegative C^1 function p and a C^1 function V satisfying equation (9.40), then the strategy $u(t) = \hat{\varphi}(t, x(t))$ solves the state feedback finite gain problem (9.39).*
2. *If furthermore, V is C^2, and there exists for all pairs $(\hat{u}(\cdot), y(\cdot))$ a C^2 function W solution of equation (9.41), and $\hat{x}(t)$ in (9.14) is always unique, then the strategy $u(t) = \hat{\varphi}(t, \hat{x}(t))$ solves the output feedback finite gain problem (9.39). Moreover, $\hat{x}(t)$ may be computed recursively according to the equation*

$$\begin{aligned} \dot{\hat{x}} &= a(t,\hat{x},\hat{\varphi}(t,\hat{x})) + 2\gamma^{-2} B(t,\hat{x},\hat{\varphi}(t,\hat{x})) V_x(t,\hat{x})' \\ &\quad - [V_{xx}(t,\hat{x}) - W_{xx}(t,\hat{x})]^{-1} \left[2\gamma^2 h_x(t,\hat{x})' P(y - h(t,\hat{x})) - p_x(t,\hat{x})'\right], \\ \hat{x}(0) &= \operatorname{Arg}\min_x \left[V(0,x) - \gamma^2 \|x\|_Q^2\right]. \end{aligned}$$

Remark 8. Although $\hat{x}$ is given by this ordinary differential equation, yet this is not a finite-dimensional controller since the partial differential equation (9.41) must be integrated in real time.

4.1.2 Stationary Problem

We choose to just show a slight extension of the result of Sect. 2.2.2 applied to the investigation of the standard problem (9.10), for a more general system, and requiring less regularity, than in the previous subsection. Let therefore a system be defined as in (9.5) with $x(0) = x_0$, (9.6), and (9.7), but with f, g, and h time independent. Let also

$$L(x,u,w) = \|g(x,u)\|^2 - \gamma^2\|w\|^2 . \tag{9.42}$$

We want to find a nonanticipative control law $u(\cdot) = \hat{\phi}(y(\cdot))$ that would guarantee that there exists a number β such that the control $\hat{u}(\cdot)$ and the state trajectory $x(\cdot)$ generated always satisfy

$$\forall (x_0, w(\cdot)) \in \Omega , \quad J(\hat{u}(\cdot), w(\cdot)) = \int_0^\infty L(x(t), \hat{u}(t), w(t))\, \mathrm{d}t \le \beta^2 . \tag{9.43}$$

We leave it to the reader to specialize the result below to the case (9.42), and further to such system equations as used, e.g., in Ball et al. (1993) and Coutinho et al. (2002).

We need the following two HJCIB equations, using a control feedback $\hat{\varphi}(x)$ and the control $\hat{u}(t)$ actually used:

$$\forall x \in \mathbb{R}^n , \quad \inf_{w\in \mathsf{W}} [-V_x(x) f(x, \hat{\varphi}(x), w) - L(x, \hat{\varphi}(x), w)] = 0 , \quad V(0) = 0 , \tag{9.44}$$

and, $\forall x \in \mathbb{R}^n$,

$$\inf_{w|y(t)} [-W_t(t,x) - W_x(t,x) f(t, \hat{u}(t), w) - L(x, \hat{u}(t), w)] = 0 , \quad W(0,x) = 0 , \tag{9.45}$$

(by $\inf_{w|y(t)}$ we mean $\inf_{w|h(t,x,w)=y(t)}$). Let $X(t)$ be the set of reachable states at time t from any x_0 with $\hat{\varphi}$ and a minimizing w in (9.45), and $\hat{x}(t)$ defined as

$$V(\hat{x}(t)) - W(t, \hat{x}(t)) = \max_{x\in X(t)} [V(x) - W(t,x)] . \tag{9.46}$$

Theorem 8.

1. *The state feedback problem* (9.43) *has a solution if and only if there exists an admissible stabilizing state feedback $\hat{\varphi}(x)$ and a BUC viscosity supersolution*[1] *$V(x)$ of equation* (9.44). *Then, $u(t) = \hat{\varphi}(x(t))$ is a solution.*
2. *If, furthermore, there exists for every pair $(\hat{u}(\cdot), y(\cdot)) \in \mathcal{U} \times \mathcal{Y}$ a BUC viscosity solution $W(t,x)$ of equation* (9.45) *in $[0,\infty) \times \mathbb{R}^n$, and if, furthermore, there is for all $t \ge 0$ a unique $\hat{x}(t)$ satisfying equation* (9.46)*, then the controller $u(t) = \hat{\varphi}(\hat{x}(t))$ solves the output feedback problem* (9.43).

[1]See Barles (1994) and Bardi and Capuzzo-Dolcetta (1997) for a definition and fundamental properties.

Remark 9.

1. A supersolution of equation (9.44) may make the left-hand side strictly positive.
2. Equation (9.45), to be integrated in real time, behaves as an observer. More precisely, it is the counterpart for the conditional cost measure of Kushner's equation of nonlinear filtering (see Daum 2015) for the conditional probability measure.

If equation (9.44) has no finite solution, then the full state information $\mathcal{H}^\infty$ control problem has no solution, and, not surprisingly, nor does the standard problem with partial corrupted information considered here. (See van der Schaft 1996.) But in sufficiently extreme cases, we may also detect problems where the available information is the reason why there is no solution:

Theorem 9. *If equation* (9.44) *has no finite solution, or, if V is a viscosity solution of* (9.44) *and, for all $(\hat{u}(\cdot), y(\cdot))$, either* (9.45) *has no finite solution or $V - W$ is unbounded by above, then there is no causal controller that can keep the criterion bounded against all $w(\cdot)$.*

It is due to its special structure that in the linear quadratic problem, either of the above two theorems applies.

4.2 Option Pricing in Finance

As an example of the use of the ideas of Sect. 2.1.1 in a very different dynamic context, we sketch here an application to the emblematic problem of mathematical finance, the problem of option pricing. The present theory is therefore an alternative to Black and Scholes' famous one, using a set description of the disturbances instead of a stochastic one.[2] As compared to the latter, the former allows us to include in a natural fashion transaction costs and also discrete-time trading.

Many types of options are traded on various markets. As an example, we will emphasize here European "vanilla" buy options, or *Call*, and sell options, or *Put*, with closure in kind. Many other types are covered by several similar or related methods in Bernhard et al. (2013).

The treatment here follows (Bernhard et al. 2013, Part 3), but it is quite symptomatic that independent works using similar ideas emerged around the year 2000, although they generally appeared in print much later. See the introduction of Bernhard et al. (2013).

[2]See, however, in (Bernhard et al. 2013, Chap. 2) a probability-free derivation of Black and Scholes' formula.

4.2.1 The Problem

Portfolio

For the sake of simplicity, we assume an economy without interest rate nor inflation. The contract is signed at time 0 and ends at the *exercise time* T. It bears upon a single financial security whose price $u(t)$ varies with time in an unpredictable manner. We take as the disturbance in this problem its relative rate of growth $\dot{u}/u = \tau$. In the Black and Scholes theory, τ is modeled as a stochastic process, the sum of a deterministic drift and of a "white noise." In keeping with the topic of this chapter, we will not make it a stochastic process. Instead, we assume that two numbers are known, $\tau^- < 0$ and $\tau^+ > 0$, and all we assume is boundedness and measurability:

$$\forall t \in [0,T]\,,\ \tau(t) \in [\tau^-, \tau^+] \quad \text{and} \quad \tau(\cdot) \in \Omega = \mathcal{M}([0,T] \to [\tau^-, \tau^+])\,. \tag{9.47}$$

A *portfolio* is made of two components: shares of the security for a monetary amount $v(t)$ and an amount $y(t)$ of currency, for a total worth $v + y = w$. Both v and y may be positive, of course, but also negative through *futures* for v, a "short" portfolio, and borrowing for y. All variables may vary continuously.

A trader manages the portfolio in a *self-financed* fashion, meaning that he buys or sells shares of the security, withdrawing the money to buy from the currency part y of the portfolio, or adding to it the proceeds of the sales. In Merton's "continuous trading" fiction, the trader may trade at a continuous rate $\xi(t)$, taken positive for a buy and negative for a sale. But he can also trade a finite block of shares instantly, resulting in jumps in $v(\cdot)$, represented by impulses in $\xi(\cdot)$. We will therefore allow a finite sum

$$\xi(t) = \xi^c(t) + \sum_k \xi_k \delta(t - t_k) \quad \Leftrightarrow \quad \xi(\cdot) \in \Xi$$

with $\xi^c(\cdot)$ a measurable real function and $\{t_k\} \subset [0,T]$ and $\{\xi_k\} \subset \mathbb{R}$ two finite sequences, all chosen by the trader. We call Ξ the set of such distributions.

There are transaction costs incurred in any transaction that we will assume to be proportional to the amount of the transaction, with proportionality coefficients $C^- < 0$ for a sale of shares of the security and $C^+ > 0$ for a buy. We will write these transaction costs as $C^\varepsilon\langle\xi\rangle$ with the convention that this means that $\varepsilon = \mathrm{sign}(\xi)$. The same convention holds for such notation as $\tau^\varepsilon\langle X\rangle$, or later $q^\varepsilon\langle X\rangle$.

This results in the following control system, where the disturbance is τ and the control ξ:

$$\dot{u} = \tau u\,, \qquad \tau(\cdot) \in \Omega \tag{9.48}$$

$$\dot{v} = \tau v + \xi\,, \qquad \xi(\cdot) \in \Xi\,, \tag{9.49}$$

$$\dot{w} = \tau v - C^\varepsilon\langle\xi\rangle\,, \tag{9.50}$$

which the trader controls through a nonanticipative strategy $\xi(\cdot) = \phi(u(\cdot))$, which may in practice take the form of a state feedback $\xi(t) = \varphi(t, u(t), v(t))$ with an additional rule saying when to make impulses, i.e., jumps in v, and by what amount.

Hedging

A terminal payment by the trader is defined in the contract, in reference to an *exercise price* K. Adding to it the closure transactions costs with rates $c^- \in [C^-, 0]$ and $c^+ \in [0, C^+]$, the total terminal payment can be formulated with the help of two auxiliary functions $\hat{v}(T, u)$ and $\hat{w}(T, u)$ depending on the type of option considered according to the following table:

Closure in kind		$u \le \frac{K}{1+c^+}$	$\frac{K}{1+c^+} \le u \le \frac{K}{1+c^-}$	$\frac{K}{1+c^-} \le u$
Call	$\hat{v}(T, u)$	0	$\frac{(1+c^+)u-K}{c^+-c^-}$	u
	$\hat{w}(T, u)$	0	$-c^-\hat{v}(T, u)$	$u - K$
Put	$\hat{v}(T, u)$	$-u$	$\frac{(1+c^-)u-K}{c^+-c^-}$	0
	$\hat{w}(T, u)$	$K - u$	$-c^+\hat{v}(u)$	0

(9.51)

And the total terminal payment is $M(u(T), v(T))$, with

$$M(u, v) = \hat{w}(T, u) + c^{\varepsilon}\langle\hat{v}(T, u) - v\rangle \ . \tag{9.52}$$

Definition 2. An initial portfolio $(v(0), w(0))$ and a trading strategy ϕ constitute a *hedge at* $u(0)$ if they ensure

$$\forall \tau(\cdot) \in \Omega\,, \qquad w(T) \ge M(u(T), v(T))\,, \tag{9.53}$$

meaning that the final worth of the portfolio is enough to cover the payment owed by the trader according to the contract signed. It follows from (9.50) that (9.53) is equivalent to:

$$\forall \tau(\cdot) \in \Omega\,, \quad M(u(T), v(T)) + \int_0^T \Big(-\tau(t)v(t) + C^{\varepsilon}\langle\xi(t)\rangle\Big)\mathrm{d}t \le w(0)\,,$$

a typical guaranteed value according to Sect. 2.1.1. Moreover, the trader wishes to construct *the cheapest possible hedge* and hence solve the problem:

$$\min_{\phi \in \Phi} \sup_{\tau(\cdot) \in \Omega} \left[M(u(T), v(T)) + \int_0^T \Big(-\tau(t)v(t) + C^{\varepsilon}\langle\xi(t)\rangle\Big)\mathrm{d}t \right]. \tag{9.54}$$

Let $V(t, u, v)$ be the Value function associated with the differential game defined by (9.48), (9.49), and (9.54). The *premium* to be charged to the buyer of the contract, if $u(0) = u_0$, is

$$P(u_0) = V(0, u_0, 0)\,.$$

This problem defines the so-called robust control approach to option pricing. An extensive use of differential game theory yields the following results.

4.2.2 The Solution

Because of the impulses allowed in the control, Isaacs' equation is replaced by the following differential quasi-variational inequality (DQVI):

$$\left.\begin{aligned} &\forall (t,u,v) \in [0,T) \times \mathbb{R}_+ \times \mathbb{R}\,, \\ &\max\Big\{-V_t - \tau^\varepsilon \langle V_u u + (V_v - 1)v\rangle\ ,\, -(V_v + C^+)\ ,\ V_v + C^-\Big\} = 0\,, \\ &\forall (u,v) \in \mathbb{R}_+ \times \mathbb{R}\,, \quad V(T,u,v) = M(u,v)\,. \end{aligned}\right\} \tag{9.55}$$

Theorem 10. *The Value function associated with the differential game defined by equations* (9.48) (9.49)*, and* (9.54) *is the unique Lipshitz continuous viscosity solution of the differential variational inequality* (9.55)*.*

Solving the DQVI (9.55) may be done with the help of the following auxiliary functions. We define

$$q^-(t) = \max\{(1 + c^-)\exp[\tau^-(T - t)] - 1\,,\, C^-\}\,,$$
$$q^+(t) = \min\{(1 + c^+)\exp[\tau^+(T - t)] - 1\,,\, C^+\}\,.$$

Note that, for $\varepsilon \in \{-, +\}$, $q^\varepsilon = C^\varepsilon$ for $t \le t_\varepsilon$ and increases ($\varepsilon = +$) or decreases ($\varepsilon = -$) toward c^ε as $t \to T$, with:

$$t_\varepsilon = T - \frac{1}{\tau^\varepsilon} \ln\left(\frac{1 + C^\varepsilon}{1 + c^\varepsilon}\right)\,. \tag{9.56}$$

We also introduce the constant matrix $\mathcal{S}$ and the variable matrix $\mathcal{T}(t)$ defined by

$$\mathcal{S} = \begin{pmatrix} 1 & 0 \\ 1 & 0 \end{pmatrix}, \qquad \mathcal{T} = \frac{1}{q^+ - q^-}\begin{pmatrix} \tau^+ q^+ - \tau^- q^- & \tau^+ - \tau^- \\ -(\tau^+ - \tau^-)q^+ q^- & \tau^- q^+ - \tau^+ q^- \end{pmatrix}.$$

Finally, we name collectively two functions:

$$W(t,u) = \begin{pmatrix} \hat{v}(t,u) \\ \hat{w}(t,u) \end{pmatrix}$$

involved in the pair of coupled linear partial differential equations

$$W_t + \mathcal{T}(W_u u - \mathcal{S}W) = 0\,. \tag{9.57}$$

with the boundary conditions (9.51).

Theorem 11.

- *The partial differential equation* (9.57) *with boundary conditions* (9.51) *has a unique solution.*
- *The Value function of the game* (9.48), (9.49), *and* (9.54) *(i.e., the unique Lipshitz continuous viscosity solution of* (9.55)*) is given by*

$$V(t,u,v) = \hat{w}(t,u) + q^{\varepsilon}(t)\langle\hat{v}(t,u) - v\rangle\ . \tag{9.58}$$

- *The optimal hedging strategy, starting with an initial wealth* $w(0) = P(u(0))$*, is to make an initial jump to* $v = \hat{v}(0,u(0))$ *and keep* $v(t) = \hat{v}(t,u(t))$ *as long as* $t < t_{\varepsilon}$ *as given by* (9.56) *and do nothing for* $t \geq t_{\varepsilon}$*, with* $\varepsilon = \mathrm{sign}[\hat{v}(t,u(t)) - v(t)]$.

Remark 10.

- In practice, for classic options, $T - t_{\varepsilon}$ is very small (typically less than one day) so that a simplified trading strategy is obtained by choosing $q^{\varepsilon} = C^{\varepsilon}$, and if T is not extremely small, $P(u_0) = \hat{w}(0,u_0) + C^{\varepsilon}\langle\hat{v}(0,u_0)\rangle$.
- The curve $P(u)$ for realistic $[\tau^-,\tau^+]$ is qualitatively similar to that of the Black and Scholes theory, usually larger because of the trading costs not accounted for in the classic theory.
- The larger the interval $[\tau^-,\tau^+]$ chosen, the larger $P(u)$. Hence the choice of these bounds is a critical step in applying this theory. The hedge has been found to be very robust against occasional violations of the bounds in (9.47).

4.2.3 Discrete-Time Trading

One of the advantages to relinquish the traditional "geometric diffusion" stochastic model for the disturbance $\tau(\cdot)$ is to allow for a coherent theory of discrete-time trading. Let therefore $h = T/N$ be a time step, with $N \in \mathbb{N}$. Assume the trader is allowed to do some trading only at instants $t_k = kh$, $k \in \mathbb{N}$. This means that we keep only the impulsive part of $\xi(\cdot)$ and fix the impulse instants t_k. We therefore have restricted the available trader's choices, thus we will end up with a larger premium.

We need now the parameters

$$\tau_h^{\varepsilon} = \mathrm{e}^{h\tau^{\varepsilon}} - 1\,,\quad \varepsilon \in \{-,+\}\,.$$

We write u_k, v_k, w_k for $u(t_k), v(t_k), w(t_k)$. An exact discretization of our system is therefore as follows, with $\tau_k \in [\tau_h^-, \tau_h^+]$:

$$u_{k+1} = (1+\tau_k)u_k\,, \tag{9.59}$$

$$v_{k+1} = (1+\tau_k)(v_k+\xi_k)\,, \tag{9.60}$$

$$w_{k+1} = w_k + \tau_k(v_k+\xi_k) - C^\varepsilon\langle\xi_k\rangle\,. \tag{9.61}$$

The Value function of the restricted game is denoted by $V_k^h(u_k, v_k)$.

Theorem 12. *The Value function $\{V_k^h\}$ satisfies the Isaacs recurrence equation:*

$$\begin{aligned} &V_k^h(u,v) = \min_{\xi}\max_{\tau\in[\tau_h^-,\tau_h^+]}\left[V_{k+1}^h((1+\tau)u,(1+\tau)(v+\xi)) - \tau(v+\xi) + C^\varepsilon\langle\xi\rangle\right],\\ &\forall(u,v)\in\mathbb{R}_+\times\mathbb{R}\,,\quad V_N^h(u,v) = M(u,v)\,. \end{aligned} \tag{9.62}$$

Moreover, if one defines $V^h(t,u,v)$ as the Value of the game where the trader (maximizer) is allowed to make one jump at initial time, and then only at times t_k as above, we have:

Theorem 13. *The function V^h interpolates the sequence $\{V_k^h\}$ in the sense that, for all (u,v), $V^h(t_k,u,v) = V_k^h(u,v)$. As the step size is subdivided and goes to zero (e.g., $h = T/2^d$, $d\to\infty$), the function $V^h(t,u,v)$ decreases and converges to the function $V(t,u,v)$ uniformly on any compact in (u,v).*

Finally, one may extend the representation formula (9.58), just replacing $\hat{v}$ and $\hat{w}$ by $\hat{v}_k^h$ and $\hat{w}_k^h$ given collectively by a carefully chosen finite difference approximation of equation (9.57) (but the representation formula is then *exact*) as follows:

Let $q_k^\varepsilon = q^\varepsilon(t_k)$ be alternatively given by $q_N^\varepsilon = c^\varepsilon$ and the recursion

$$q_{k+\frac{1}{2}}^\varepsilon = (1+\tau_h^\varepsilon)q_{k+1}^\varepsilon + \tau_h^\varepsilon\,,$$

$$q_k^- = \max\{q_{k+\frac{1}{2}}^-, C^-\}\,,\qquad q_k^+ = \min\{q_{k+\frac{1}{2}}^+, C^+\}\,.$$

Also, let

$$Q_k^\varepsilon = (q_k^\varepsilon\;\;1)\,,\qquad W_k^h(u) = \begin{pmatrix}\hat{v}_k^h(u)\\ \hat{w}_k^h(u)\end{pmatrix}.$$

The following algorithm is derived from a detailed analysis of equation (9.62):

$$\begin{aligned} W_k^h(u) &= \frac{1}{q_{k+\frac{1}{2}}^+ - q_{k+\frac{1}{2}}^-}\begin{pmatrix}1 & -1\\ -q_{k+\frac{1}{2}}^- & q_{k+\frac{1}{2}}^+\end{pmatrix}\begin{pmatrix}Q_{k+1}^+W_{k+1}^h((1+\tau_h^+)u)\\ Q_{k+1}^-W_{k+1}^h((1+\tau_h^-)u)\end{pmatrix},\\ W_N^h(u) &= W(T,u)\,. \end{aligned}$$

And as previously,

$$V_k^h(u,v) = \hat{w}_k^h(u) + q_k^\varepsilon\big\langle \hat{v}_k^h(u) - v\big\rangle\,,$$

and for all practical purposes, $P(u_0) = \hat{w}_0^h(u_0) + C^\varepsilon\big\langle \hat{v}_0^h(u_0)\big\rangle$.

5 Conclusion

As stated in the introduction, game theoretic methods are only one part, may be a prominent one, of modern robust control. They typically cover a wide spectrum of potential applications, their limitations being in the difficulty to solve a nonlinear dynamic game, often with imperfect information. Any advances in this area would instantly translate into advances in robust control.

The powerful theory of linear quadratic games coupled with the min-max certainty equivalence principle makes it possible to efficiently solve the linear $\mathcal{H}^\infty$-optimal control problem, while the last example above shows that some very nonlinear problems may also receive a rather explicit solution via these game theoretic methods.

References

Akian M, Quadrat J-P, Viot M (1998) Duality between probability and optimization. In: Gunawardena J (ed) Indempotency. Cambridge University Press, Cambridge

Aubin J-P, Bayen AM, Saint-Pierre P (2011) Viability theory. New directions. Springer, Heidelberg/New York

Baccelli F, Cohen G, Olsder G-J, Quadrat J-P (1992) Synchronization and linearity. Wiley, Chichester/New York

Ball JA, Helton W, Walker ML (1993) H^∞ control for nonlinear systems with output feedback. IEEE Trans Autom Control 38:546–559

Bardi M, Capuzzo-Dolcetta I (1997) Optimal control and viscosity solutions of Hamilton-Jacobi-Bellman equations. Birkhaüser, Boston

Barles G (1994) Solutions de viscosité des équations de Hamilton-Jacobi. Springer, Paris/Berlin

Başar T, Bernhard P (1995) H^∞-optimal control and related minimax design problems: a differential games approach, 2nd edn. Birkhäuser, Boston

Başar T, Olsder G-J (1982) Dynamic noncooperative game theory. Academic Press, London/New York

Bensoussan A, Bernhard P (1993) On the standard problem of H_∞-optimal control for infinite dimensional systems. In: Identification and control of systems governed by patial differential equations (South Hadley, 1992). SIAM, Philadelphia, pp 117–140

Berkowitz L (1971) Lectures on differential games. In: Kuhn HW, Szegö GP (eds) Differential games and related topics. North Holland, Amsterdam, pp 3–45

Bernhard P (1994) A min-max certainty equivalence principle for nonlinear discrete time control problems. Syst Control Lett 24:229–234

Bernhard P, Rapaport A (1996) Min-max cetainty equivalence principle and differential games. Int J Robust Nonlinear Control 6:825–842

Bernhard P, Engwerda J, Roorda B, Schumacher H, Kolokoltsov V, Aubin J-P, Saint-Pierre P (2013) The interval market model in mathematical finance: a game theoretic approach. Birkhaüser, New York
Corless MJ, Leitman G (1981) Continuous state feedback guaranteeing uniform ultimate boundedness for uncertain dynamic systems. IEEE Trans Autom Control 26:1139–1144
Coutinho DF, Trofino A, Fu M (2002) Nonlinear H-infinity control: an LMI approach. In: 15th trienal IFAC world congress. IFAC, Barcelona, Spain
Daum FE (2015) Nonlinear filters. In: Samad T, Ballieul J (eds) Encyclopedia of systems and control. Springer, London, pp 870–876
Didinsky G, Başar T, Bernhard P (1993) Structural properties of minimax policies for a class of differential games arising in nonlinear H^{∞}-optimal control. Syst Control Lett 21:433–441
Doyle JC (1982) Analysis of feedback systems with structured uncertainties. IEEE Proc 129: 242–250
Doyle JC, Glover K, Kargonekar PP, Francis BA (1989) State-space solutions to the standard $\mathcal{H}_2$ and $\mathcal{H}_\infty$ control problems. IEEE Trans Autom Control 34:831–847
El Ghaoui L, Scorletti G (1996) Control of rational systems using linear-fractional representations and linear matrix inequalities. Automatica 32:1273–1284
Georgiou TT, MC Smith (1997) Robustness analysis of nonlinear feedback systems: an input-output approach. IEEE Trans Autom Control 42:1200–1221
Glover K (2015) H-infinity control. In: Bailleul J, Samad T (eds) Encyclopaedia of systems and control. Springer, London, pp 520–526
Gutman S (1979) Uncertain dynamical systems: a Lyapunov min-max approach. IEEE Trans Autom Control 24:437–443
Gutman S, Leitmann G (1976) Stabilizing feedback control for dynamical systems with bounded uncertainty. In: Proceedings of the conference on decision and control. IEEE, Clearwater, Florida
Kang-Zhi L (2015) Lmi approach to robust control. In: Bailleul J, Samad T (eds) Encyclopaedia of systems and control. Springer, London, pp 675–681
Kurzhanski AB, Varaiya P (2014) Dynamics and control of trajectory tubes. Birkhäuser, Cham
Qiu L (2015) Robust control in gap metric. In: Bailleul J, Samad T (eds) Encyclopaedia of systems and control. Springer, London, pp 1202–1207
Rapaport A, Gouzé J-L (2003) Parallelotopic and practical observers for nonlinear uncertain systems. Int J Control 76:237–251
Samad T, Ballieul J (2015) Encyclopedia of systems and control. Springer, London
Stoorvogel AA (1992) The H-infinity control problem. Prentice Hall, NeW York
van der Schaft A (1996) L_2-gain and passivity techniques in nonlinear control. Lecture notes in control and information sciences, vol 18. Springer, London
Vidyasagar M (1993) Nonlinear systems analysis. Prentice Hall, Englewood Cliff
Zames G (1981) Feedback and optimal sensitivity: model reference transformations, multiplicative seminorms, and approximate inverses. IEEE Trans Autom Control 26:301–320
Zames G, El-Sakkary AK (1980) Unstable ststems and feedback: the gap metric. In: Proceedings of the 16th Allerton conference, Monticello, pp 380–385
Zhou K, JC Doyle, Glover K (1996) Robust and optimal control. Prentice Hall, Upper Saddle River

Evolutionary Game Theory

10

Ross Cressman and Joe Apaloo

Contents

Abstract

Evolutionary game theory developed as a means to predict the expected distribution of individual behaviors in a biological system with a single species that evolves under natural selection. It has long since expanded beyond its biological roots and its initial emphasis on models based on symmetric games

R. Cressman (✉)
Department of Mathematics, Wilfrid Laurier University, Waterloo, ON, Canada
e-mail: rcressman@wlu.ca

J. Apaloo
Department of Mathematics, Statistics and Computer Science, St Francis Xavier University, Antigonish, NS, Canada
e-mail: japaloo@stfx.ca

T. Başar, G. Zaccour (eds.), *Handbook of Dynamic Game Theory*,
https://doi.org/10.1007/978-3-319-44374-4_6

with a finite set of pure strategies where payoffs result from random one-time interactions between pairs of individuals (i.e., on matrix games). The theory has been extended in many directions (including nonrandom, multiplayer, or asymmetric interactions and games with continuous strategy (or trait) spaces) and has become increasingly important for analyzing human and/or social behavior as well. This chapter initially summarizes features of matrix games before showing how the theory changes when the two-player game has a continuum of traits or interactions become asymmetric. Its focus is on the connection between static game-theoretic solution concepts (e.g., ESS, CSS, NIS) and stable evolutionary outcomes for deterministic evolutionary game dynamics (e.g., the replicator equation, adaptive dynamics).

Keywords
ESS · CSS · NIS · Neighborhood superiority · Evolutionary game dynamics · Replicator equation · Adaptive dynamics · Darwinian dynamics

1 Introduction

Evolutionary game theory developed as a means to predict the expected distribution of individual behaviors in a biological system with a single species that evolves under natural selection (Maynard Smith 1974; Maynard Smith and Price 1973). The theory's predictions of equilibrium behavior correspond to intuitive static solutions of the game formed through fitness (i.e., payoff) comparisons of different behaviors (i.e., strategies). A fundamental result is that, at a stable behavioral distribution, no individual in the population can increase its fitness by unilaterally changing strategy (see, e.g., condition (1) below). That is, a stable outcome for natural selection implies individuals will exhibit Nash equilibrium (NE) behavior (Nash 1950, 1951), a result that has come to be known as one aspect of the Folk Theorem of evolutionary game theory (Broom and Rychtar 2013; Cressman 2003; Hofbauer and Sigmund 1998, 2003; Sandholm 2010) given in Theorem 1 below.

However, as we will see, stability requires more than NE. The most common additional requirement, introduced already by Maynard Smith and Price (1973), is that of an *evolutionarily stable strategy* (ESS). According to John Maynard Smith (1982, page 10) in his influential book, Evolution and the Theory of Games, an ESS is "a strategy such that, if all members of the population adopt it, then no mutant strategy could invade the population under the influence of natural selection." He goes on to argue that his definition, which seems heuristically related to stability of a monomorphic resident strategy with respect to the invasion dynamics of mutants, is equivalent to the standard one (Definition 1 below) given through static payoff comparisons when the evolutionary game and invasion dynamics are modeled as in Sect. 2.1 by a symmetric normal form game and the replicator equation, respectively.

In fact, as illustrated throughout this article, there is a complex relationship between the static stability conditions (such as the ESS) and stability with respect to game dynamics (such as the replicator equation). It is this relationship that formed the initial basis of what is now known as evolutionary game theory.

Evolutionary game theory has long since expanded beyond its biological roots and become increasingly important for analyzing human and/or social behavior. Here, changes in strategy frequencies do not result from natural selection; rather, individuals (or societies) alter their behavior based on payoff consequences. The replicator equation then emerges from, for instance, individuals making rational decisions on how to imitate observed strategies that currently receive higher payoff (Schlag 1997). Depending on what information these decision-makers have (and how they use this information), a vast array of other game dynamics are possible (Hofbauer and Sigmund 2003; Sandholm 2010; Sigmund 2011). Evolutionary game theory has also become a standard method to choose among the many NE that often arise in these models of human interactions between players that can be individuals or other economic entities such as firms or even nations (e.g., Samuelson 1997; Sandholm 2010). Thus, the ESS can be viewed as an NE refinement or equilibrium selection technique.

It is in this latter capacity that evolutionary game theory initially gained prominence in the economic literature when applied to rational decision-making in classical noncooperative, symmetric games in either normal form or extensive form (see van Damme (1991) and Samuelson (1997) and the references therein). From this perspective, evolutionary games often consider other deterministic or stochastic evolutionary dynamics besides the replicator equation since these are thought to better represent decision-making applied to economic or learning models (Cressman 2003; Fudenberg and Levine 1998; Gintis 2000; Hofbauer and Sigmund 1998; Mesterton-Gibbons 2000; Nowak 2006; Sandholm 2010; Vega-Redondo 1996; Weibull 1995; Young 1998).

The biological perspective of evolutionary game theory has been summarized in several survey monographs and books (e.g., Bomze and Pötscher 1989; Broom and Rychtar 2013; Cressman 1992; Hines 1987; Hofbauer and Sigmund 1988, 1998; Mesterton-Gibbons 2000; Sigmund 1993; Vincent and Brown 2005).

Evolutionary game theory and its corresponding game dynamics have also expanded well beyond their initial emphasis on single-species games with a finite set of pure strategies where payoffs result from random one-time interactions between pairs of individuals (i.e., two-player symmetric normal form games or, more simply, matrix games). In this chapter, we highlight features of matrix games in the following section before investigating in Sect. 3 how the theory changes when the symmetric game has a continuum of pure strategies (or traits). Section 4 then generalizes the theory developed in Sects. 2 and 3 to asymmetric games. Specifically, two-player games with two roles are examined that either have finitely many pure strategies in each role (in normal or extensive form) or have a one-dimensional continuous trait space in each role.

2 Evolutionary Game Theory for Symmetric Normal Form Games

2.1 The ESS and Invasion Dynamics

The relationship between the ESS and stability of game dynamics is most clear when individuals in a single species can use only two possible strategies, denoted p^* and p to match notation used later in this article, and payoffs are linear. Suppose that $\pi(p, \hat{p})$ is the payoff to a strategy p used against strategy $\hat{p}$. In biological terms, $\pi(p, \hat{p})$ is the fitness of an individual using strategy p in a large population exhibiting behavior $\hat{p}$.[1] Then, an individual in a monomorphic population where everyone uses p^* cannot improve its fitness by switching to p if

$$\pi(p, p^*) \leq \pi(p^*, p^*) \quad \text{NE condition.} \tag{10.1}$$

If p playing against p^* does equally as well as p^* playing against p^* (i.e., if $\pi(p, p^*) = \pi(p^*, p^*)$), then stability requires the extra condition that p^* must do better than p in their rare contests against p. That is,

$$\pi(p, p) < \pi(p^*, p) \text{ if } \pi(p, p^*) = \pi(p^*, p^*) \quad \text{stability condition.} \tag{10.2}$$

For the game with strategies p^* and p, p^* is defined to be an ESS if it satisfies conditions (10.1) and (10.2).

To see why both these conditions are necessary for dynamic stability, under the assumption of Maynard Smith (1982) that "like begets like," the per capita change in the number of individuals using strategy p is its expected payoff. This leads to the following continuous-time invasion dynamics of a resident monomorphic population p^* by a small proportion ε of mutants using p.

$$\begin{aligned}\dot{\varepsilon} &= \varepsilon\left[\pi(p, \varepsilon p + (1-\varepsilon)p^*) - \pi(\varepsilon p + (1-\varepsilon)p^*, \varepsilon p + (1-\varepsilon)p^*)\right] \\ &= \varepsilon(1-\varepsilon)\left[\pi(p, \varepsilon p + (1-\varepsilon)p^*) - \pi(p^*, \varepsilon p + (1-\varepsilon)p^*)\right] \\ &= \varepsilon(1-\varepsilon)\left[(1-\varepsilon)(\pi(p, p^*) - \pi(p^*, p^*)) + \varepsilon(\pi(p, p) - \pi(p^*, p))\right]\end{aligned} \tag{10.3}$$

Here, we have used repeatedly that payoffs $\pi(p, \hat{p})$ are linear in both p and $\hat{p}$. In fact, this is the replicator equation of Sect. 2.2 for the matrix game with two pure strategies, p and p^*, and payoff matrix (10.5).

[1] In the basic biological model for evolutionary games, individuals are assumed to engage in random pairwise interactions. Moreover, the population is assumed to be large enough that an individual's fitness (i.e., reproductive success) $\pi(p, \hat{p})$ is the expected payoff of p if the average strategy in the population is $\hat{p}$. In these circumstances, it is often stated that the population is effectively infinite in that there are no effects due to finite population size. Such stochastic effects are discussed briefly in the final section.

If p^* is a strict NE (i.e., the inequality in (10.1) is strict), then $\dot{\varepsilon} < 0$ for all positive ε that are close to 0 since $(1-\varepsilon)(\pi(p, p^*) - \pi(p^*, p^*))$ is the dominant term corresponding to the common interactions against p^*. Furthermore, if this term is 0 (i.e., if p^* is not strict), we still have $\dot{\varepsilon} < 0$ from the stability condition (10.2) corresponding to the less common interactions against p. Conversely, if $\dot{\varepsilon} < 0$ for all positive ε that are close to 0, then p^* satisfies (10.1) and (10.2). Thus, the resident population p^* (i.e., $\varepsilon = 0$) is locally asymptotically stable[2] under (10.3) (i.e., there is dynamic stability at p^* under the replicator equation) if and only if p^* satisfies (10.1) and (10.2).

In fact, dynamic stability occurs at such a p^* in these two-strategy games for any game dynamics whereby the proportion (or frequency) of users of strategy $\hat{p}$ increases if and only if its expected payoff is higher than that of the alternative strategy. We then have the result that p^* is an ESS if and only if it satisfies (10.1) and (10.2) if and only if it is dynamically stable with respect to any such game dynamics.

These results assume that there is a resident strategy p^* and a single mutant strategy p. If there are other possible mutant strategies, an ESS p^* must be locally asymptotically stable under (10.3) for any such p in keeping with Maynard Smith's (1982) dictum that no mutant strategy can invade. That is, p^* is an ESS if and only if it satisfies (10.1) and (10.2) for all mutant strategies p (see also Definition 1 (b) of Sect. 2.2).

2.2 The ESS and the Replicator Equation for Matrix Games

In an evolutionary game with symmetric normal form, the population consists of individuals who must all use one of finitely many (say m) possible behaviors at any particular instant in time. These strategies are denoted e_i for $i = 1, \ldots, m$ and called pure strategies. Moreover, $S \equiv \{e_1, \ldots, e_m\}$ is called the pure-strategy set. An individual may also use a mixed strategy in $\Delta^m \equiv \{p = (p_1, \ldots, p_m) \mid \sum_{i=1}^m p_i = 1, p_i \geq 0\}$ where p_i is the proportion of the time this individual uses pure strategy e_i. If population size is large and the components of $\hat{p} \in \Delta^m$ are the current frequencies of strategies used in the population (i.e., $\hat{p}$ is the population state), then the payoff of an individual using p in a random pairwise interaction is given explicitly through the bilinear payoff function of the (two-player) symmetric normal form game, $\pi(p, \hat{p}) \equiv \sum_{i,j=1}^m p_i \pi(e_i, e_j) \hat{p}_j$, where, as before, $\pi(e_i, e_j)$ is the payoff to e_i against e_j.

[2]Clearly, the unit interval $[0, 1]$ is (forward) invariant under the dynamics (10.3) (i.e., if $\varepsilon(t)$ is the unique solution of (10.3) with initial value $\varepsilon(0) \in [0, 1]$, then $\varepsilon(t) \in [0, 1]$ for all $t \geq 0$). The rest point $\varepsilon = 0$ is (Lyapunov) *stable* if, for every neighborhood U of 0 relative to $[0, 1]$, there exists a neighborhood V of 0 such that $\varepsilon(t) \in U$ for all $t \geq 0$ if $\varepsilon(0) \in V \cap [0, 1]$. It is *attracting* if, for some neighborhood U of 0 relative to $[0, 1]$, $\varepsilon(t)$ converges to 0 whenever $\varepsilon(0) \in U$. It is *(locally) asymptotically stable* if it is both stable and attracting. Throughout the chapter, dynamic stability is equated to local asymptotic stability.

Based on this linearity, the following notation is commonly used for these games. Let e_i be represented by the ith unit column vector in R^m and $\pi(e_i, e_j)$ by entry A_{ij} in an $m \times m$ payoff matrix A. Then, with vectors in Δ^m thought of as column vectors, $\pi(p, \hat{p})$ is the inner product $p \cdot A\hat{p}$ of the two column vectors p and $A\hat{p}$. For this reason, symmetric normal form games are often called matrix games with payoffs given in this latter form.

To obtain the continuous-time, pure-strategy replicator equation (10.4) following the original fitness approach of Taylor and Jonker (1978), individuals are assumed to use pure strategies and the per capita growth rate in the number n_i of individuals using strategy e_i at time t is taken as the expected payoff of e_i from a single interaction with a random individual in the large population. That is, $\dot{n}_i = n_i \sum_{j=1}^{m} \pi(e_i, e_j) p_j \equiv n_i \pi(e_i, p)$ where p is the population state in the (mixed) strategy simplex Δ^m with components $p_i = n_i / \sum_{j=1}^{m} n_j$ the proportion of the population using strategy e_i at time t.[3] A straightforward calculus exercise[4] yields the replicator equation on Δ^m

$$\dot{p}_i = p_i \left(\pi(e_i, p) - \pi(p, p)\right) \quad \text{for} \quad i = 1, \ldots, m \tag{10.4}$$

where $\pi(p, p) = \sum_{j=1}^{m} p_j \pi(e_j, p)$ is the average payoff of an individual chosen at random (i.e., the population mean payoff). From the theory of dynamical systems, trajectories of (10.4) leave the interior of Δ^m forward invariant as well as each of its faces (Hofbauer and Sigmund 1998).

The replicator equation can be applied to the two-strategy game (on p^* and p) of Sect. 2.1 by taking these as the pure strategies with corresponding payoff matrix

[3]The approach of Taylor and Jonker (1978) also relies on the population being large enough (or effectively infinite) so that n_i and p_i are considered to be continuous variables.

[4]With $N \equiv \sum_{j=1}^{m} n_j$ the total population size,

$$\begin{aligned}
\dot{p}_i &= \frac{\dot{n}_i N - n_i \sum_{j=1}^{m} \dot{n}_j}{N^2} \\
&= \frac{n_i \pi(e_i, p) - p_i \sum_{j=1}^{m} n_j \pi(e_j, p)}{N} \\
&= p_i \pi(e_i, p) - p_i \sum_{j=1}^{m} p_j \pi(e_j, p) \\
&= p_i \left(\pi(e_i, p) - \pi(p, p)\right)
\end{aligned}$$

for $i = 1, \ldots, m$. This is the replicator equation (10.4) in the main text. Since $\dot{p}_i = 0$ when $p_i = 0$ and $\sum_{1=1}^{m} \dot{p}_i = \pi(p, p) - \pi(p, p) = 0$ when $p \in \Delta^m$, the interior of Δ^m is invariant as well as all its (sub)faces under (10.4). Since Δ^m is compact, there is a unique solution of (10.4) for all $t \geq 0$ for a given initial population state $p(0) \in \Delta^m$. That is, Δ^m is forward invariant under (10.4).

$$\begin{array}{c} \\ p^* \\ p \end{array} \begin{array}{c} \begin{array}{cc} p^* & p \end{array} \\ \left[\begin{array}{cc} \pi(p^*, p^*) & \pi(p^*, p) \\ \pi(p, p^*) & \pi(p, p) \end{array} \right] \end{array}. \tag{10.5}$$

With ε the proportion using strategy p (and $1 - \varepsilon$ using p^*), the one-dimensional replicator equation is given by (10.3). Then, from Sect. 2.1, p^* is an ESS of the matrix game on Δ^m if and only if it is locally asymptotically stable under (10.3) for all choices of mutant strategies $p \in \Delta^m$ with $p \neq p^*$ (see also Definition 1 (b) below).

The replicator equation (10.4) for matrix games is the first and most important game dynamics studied in connection with evolutionary game theory. It was defined by Taylor and Jonker (1978) (see also Hofbauer et al. 1979) and named by Schuster and Sigmund (1983). Important properties of the replicator equation are briefly summarized for this case in the Folk Theorem and Theorem 2 of the following section, including the convergence to and stability of the NE and ESS. The theory has been extended to other game dynamics for symmetric games (e.g., the best response dynamics and adaptive dynamics). The replicator equation has also been extended to many other types of symmetric games (e.g., multiplayer, population, and games with continuous strategy spaces) as well as to corresponding types of asymmetric games.

To summarize Sects. 2.1 and 2.2, we have the following definition.

Definition 1. Consider a matrix game on Δ^m.

(a) $p^* \in \Delta^m$ is a (symmetric) *NE* if it satisfies (10.1) for all $p \in \Delta^m$.
(b) $p^* \in \Delta^m$ is an *ESS* if it is an NE that satisfies (10.2) for all $p \in \Delta^m$ with $p \neq p^*$.
(c) The (pure strategy) *replicator equation* on Δ^m is

$$\dot{p}_i = p_i \left(\pi(e_i, p) - \pi(p, p) \right) \text{ for } i = 1, \ldots, m.$$

2.3 The Folk Theorem of Evolutionary Game Theory

Theorem 1. *The replicator equation for a matrix game on Δ^m satisfies:*

(a) *A stable rest point is an NE.*
(b) *A convergent trajectory in the interior of Δ^m evolves to an NE.*
(c) *A strict NE is locally asymptotically stable.*

Theorem 1 is the Folk Theorem of evolutionary game theory (Cressman 2003; Hofbauer and Sigmund 1998, 2003) applied to the replicator equation. The three conclusions are true for many matrix game dynamics (in either discrete or continuous time) and serve as a benchmark to test dynamical systems methods applied to general game dynamics and to non-matrix evolutionary games such as those considered in the remaining sections of this chapter.

The Folk Theorem means that biologists can predict the evolutionary outcome of their stable systems by examining NE behavior of the underlying game. It is as if individuals in these systems are rational decision-makers when in reality it is natural selection through reproductive fitness that drives the system to its stable outcome. This has produced a paradigm shift toward strategic reasoning in population biology. The profound influence it has had on the analysis of behavioral ecology is greater than earlier game-theoretic methods applied to biology such as Fisher's (1930) argument (see also Darwin 1871; Hamilton 1967; Broom and Krivan (▶ Chap. 23, "Biology and Evolutionary Games", this volume)) for the prevalence of the 50:50 sex ratio in diploid species and Hamilton's (1964) theory of kin selection.

The importance of strategic reasoning in population biology is further enhanced by the following result.

Theorem 2. *Consider a matrix game on* Δ^m.

(a) p^* *is an ESS of the game if and only if* $\pi(p^*, p) > \pi(p, p)$ *for all* $p \in \Delta^m$ *sufficiently close (but not equal) to* p^*.
(b) *An ESS* p^* *is a locally asymptotically stable rest point of the replicator equation* (10.4).
(c) *An ESS* p^* *in the interior of* Δ^m *is a globally asymptotically stable rest point of the replicator equation* (10.4).

The equivalent condition for an ESS contained in part (a) is the more useful characterization when generalizing the ESS concept to other evolutionary games.[5] It is called locally superior (Weibull 1995), neighborhood invader strategy (Apaloo 2006), or neighborhood superior (Cressman 2010). One reason for different names for this concept is that there are several ways to generalize local superiority to other evolutionary games and these have different stability consequences.

From parts (b) and (c), if p^* is an ESS with full support (i.e., the support $\{i \mid p_i^* > 0\}$ of p^* is $\{1, 2, \ldots, m\}$), then it is the only ESS. This result easily extends to the Bishop-Cannings theorem (Bishop and Cannings 1976) that the support of one ESS cannot be contained in the support of another, an extremely useful property when classifying the possible ESS structure of matrix games (Broom and Rychtar 2013). Haigh (1975) provides a procedure for finding ESSs in matrix games based on such results.

Parts (b) and (c) were an early success of evolutionary game theory since stability of the predicted evolutionary outcome under the replicator equation is assured at an ESS not only for the invasion of p^* by a subpopulation using a single

[5]The proof of this equivalence relies on the compactness of Δ^m and the bilinearity of the payoff function $\pi(p, q)$ as shown by Hofbauer and Sigmund (1998).

mutant strategy p but also by multiple pure strategies. In fact, if individuals use mixed strategies for which some distribution have average strategy p^*, then p^* is asymptotically stable under all corresponding mixed-strategy replicator equations if and only if p^* is an ESS (see the *strong stability* concept of Cressman 1992 and Hofbauer and Sigmund 1998). That is, stable evolutionary outcomes with respect to mixed-strategy replicator equations are equivalent to the ESS. Moreover, the converse of part (b) for the pure-strategy replicator equation (i.e., for (10.4)) is true when there are two pure strategies (i.e., $m = 2$). The three categories of such games (Hawk-Dove, Prisoner's Dilemma, and Coordination games) are classified and analyzed in Weibull (1995) (see also Broom and Krivan's (▶ Chap. 23, "Biology and Evolutionary Games") for the Hawk-Dove and Prisoner's Dilemma games).

However, there already exist non-ESS strategies p in three-strategy symmetric normal form games (i.e., for $m = 3$) that are asymptotically stable under (10.4) (such strategies p must be NE by the Folk Theorem). Broom and Krivan also provide a biologically relevant illustration of this phenomenon based on a generalized Rock-Scissors-Paper (RSP) game that exhibits cyclic dominance since P strictly dominates in the two-strategy $\{R, P\}$ game, S strictly dominates in the two-strategy $\{P, S\}$ game, and R strictly dominates in the two-strategy $\{R, S\}$ game. These games always have a unique NE p^* (that must be in the interior), but conditions on payoff matrix entries for p^* to be an ESS are different than those for stability with respect to (10.4).

The most elegant proof (Hofbauer et al. 1979) of the stability statements in parts (b) and (c) shows that $V(p) \equiv \prod p_i^{p_i^*}$ where the product is taken over $\{i : p_i^* > 0\}$ is a strict local Lyapunov function (i.e., $V(p^*) > V(p)$ and $\dot{V}(p) = V(p)(\pi(p^*, p) - \pi(p, p)) > 0$ for all $p \in \Delta^m$ sufficiently close but not equal to an ESS p^*).[6] It is tempting to add these stability statements to the Folk Theorem since they remain valid for many matrix game dynamics through the use of other Lyapunov functions. Besides the above differences between dynamic stability and ESS noted above for the RSP example, there are other reasons to avoid this temptation.

In particular, parts (b) and (c) of Theorem 2 are not true for discrete-time matrix game dynamics. One such dynamics is the discrete-time replicator equation of Maynard Smith (1982)

$$p_i' = p_i \frac{\pi(e_i, p)}{\pi(p, p)} \tag{10.6}$$

[6]Under the replicator equation, $\dot{V}(p) = \sum_{i=1}^m p_i^* p_i^{p_i^*-1} \dot{p}_i \prod_{\{j | j \neq i, p_j^* \neq 0\}} p_j^{p_j^*} = \sum_{i=1}^m p_i^* \prod_j p_j^{p_j^*} (\pi(e_i, p) - \pi(p, p)) = V(p)(\pi(p^*, p) - \pi(p, p)) > 0$ for all $p \in \Delta^m$ sufficiently close but not equal to an ESS p^*. Since $V(p)$ is a strict local Lyapunov function, p^* is locally asymptotically stable. Global stability (i.e., in addition to local asymptotic stability, all interior trajectories of (10.4) converge to p^*) in part (c) follows from global superiority (i.e., $\pi(p^*, p) > \pi(p, p)$ for all $p \neq p^*$) in this case.

where p'_i is the frequency of strategy e_i one generation later and $\pi(e_i, p)$ is the expected nonnegative number of offspring of each e_i-individual. When applied to matrix games, each entry in the payoff matrix is typically assumed to be positive (or at least nonnegative), corresponding to the contribution of this pairwise interaction to expected offspring. It is then straightforward to verify that (10.6) is a forward invariant dynamic on Δ^m and on each of its faces.

To see that an ESS may not be stable under (10.6), fix $|\varepsilon| < 1$ and consider the generalized RSP game with payoff matrix

$$A = \begin{array}{c} \\ R \\ S \\ P \end{array} \begin{array}{c} \begin{array}{ccc} R & S & P \end{array} \\ \begin{bmatrix} -\varepsilon & 1 & -1 \\ -1 & -\varepsilon & 1 \\ 1 & -1 & -\varepsilon \end{bmatrix} \end{array} \tag{10.7}$$

that has a unique NE $p^* = (1/3, 1/3, 1/3)$. For $\varepsilon = 0$, this is the standard zero-sum RSP game whose trajectories with respect to the replicator equation (10.4) form periodic orbits around p^* (Fig. 10.1a). For positive ε, p^* is an interior ESS and trajectories of (10.4) spiral inward as they cycle around p^* (Fig. 10.1b).

It is well known (Hofbauer and Sigmund 1998) that adding a constant c to every entry of A does not affect either the NE/ESS structure of the game or the trajectories of the continuous-time replicator equation. The constant c is a background fitness that is common to all individuals that changes the speed of continuous-time evolution but not the trajectory. If $c \geq 1$, all entries of this new payoff matrix are nonnegative, and so the discrete-time dynamics (10.6) applies. Now background fitness does change the discrete-time trajectory. In fact, for the matrix $1 + A$ (i.e., if $c = 1$) where A is the RSP game (10.7), p^* is unstable for all $|\varepsilon| < 1$ as can be shown through the linearization of this dynamics about the rest point p^* (specifically, the relevant eigenvalues of this linearization have modulus

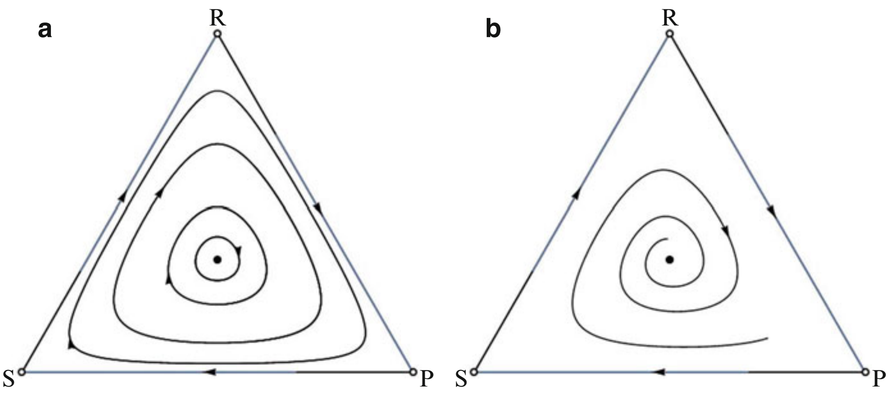


Fig. 10.1 Trajectories of the replicator equation (10.4) for the RSP game. (**a**) $\varepsilon = 0$. (**b**) $\varepsilon > 0$

greater than 1 (Cressman 2003)). The intuition here is that p' is far enough along the tangent at p in Fig. 10.1 that these points spiral outward from p^* under (10.6) instead of inward under (10.4).[7]

Cyclic behavior is common not only in biology (e.g., predator-prey systems) but also in human behavior (e.g., business cycles, the emergence and subsequent disappearance of fads, etc.). Thus, it is not surprising that evolutionary game dynamics include cycles as well. In fact, as the number of strategies increases, even more rich dynamical behavior such as chaotic trajectories can emerge (Hofbauer and Sigmund 1998).

What may be more surprising is the many classes of matrix games (Sandholm 2010) for which these complicated dynamics do not appear (e.g., potential, stable, supermodular, zero-sum, doubly symmetric games), and for these the evolutionary outcome is often predicted through rationality arguments underlying Theorems 1 and 2. Furthermore, these arguments are also relevant for other game dynamics examined in the following section.

Before doing so, it is important to mention that the replicator equation for doubly symmetric matrix games (i.e., a symmetric game whose payoff matrix is symmetric) is formally equivalent to the continuous-time model of natural selection at a single (diploid) locus with m alleles $A_1, \ldots A_m$ (Akin 1982; Cressman 1992; Hines 1987; Hofbauer and Sigmund 1998). Specifically, if a_{ij} is the fitness of genotype $A_i A_j$ and p_i is the frequency of allele A_i in the population, then (10.4) is the continuous-time selection equation of population genetics (Fisher 1930). It can then be shown that population mean fitness $\pi(p, p)$ is increasing (c.f. one part of the fundamental theorem of natural selection). Furthermore, the locally asymptotically stable rest points of (10.4) correspond precisely to the ESSs of the symmetric payoff matrix $A = \left(a_{ij}\right)_{i.j=1}^{m}$, and all trajectories in the interior of Δ^m converge to an NE of A (Cressman 1992, 2003). Analogous results hold for the classical discrete-time viability selection model with nonoverlapping generations and corresponding dynamics (10.6) (Nagylaki 1992).

2.4 Other Game Dynamics

A monotone selection dynamics (Samuelson and Zhang 1992) is of the form $\dot{p}_i = p_i g_i(p)$ where $g_i(p) > g_j(p)$ if and only if $\pi(e_i, p) > \pi(e_j, p)$ for $i, j = 1, \ldots, m$ and Δ^m is forward invariant (i.e., $\sum_{j=1}^m p_j g_j(p) = 0$). The replicator equation is the special case where $g_i(p) \equiv \pi(e_i, p) - \sum_{j=1}^m p_j \pi(e_j, p) = \pi(e_i, p) - \pi(p, p)$. For strategies e_i and e_j that are currently in use, monotone selection dynamics increase the relative frequency (p_i/p_j) of e_i compared to e_j if and only if e_i has higher expected payoff than e_j. For the RSP game (10.7) with

[7]This intuition is correct for small constants c greater than 1. However, for large c, the discrete-time trajectories approach the continuous-time ones and so $p^* = (1/3, 1/3, 1/3)$ will be asymptotically stable under (10.6) when $\varepsilon > 0$.

$0 < \varepsilon < 1$ fixed, the $g_i(p)$ can be chosen as continuously differentiable functions for which the interior ESS $p^* = (1/3, 1/3, 1/3)$ is not globally asymptotically stable under the corresponding monotone selection dynamics (c.f. Theorem 2(c)). In particular, Cressman (2003) shows there may be trajectories that spiral outward from initial points near p^* to a stable limit cycle in the interior of Δ^3 for these games.[8]

The best response dynamics (10.8) for matrix games was introduced by Gilboa and Matsui (1991) (see also Matsui 1992) as the continuous-time version of the fictitious play process, the first game dynamics introduced well before the advent of evolutionary game theory by Brown (1951) and Robinson (1951).

$$\dot{p} = BR(p) - p \tag{10.8}$$

In general, $BR(p)$ is the set of best responses to p and so may not be a single strategy. That is, (10.8) is a differential inclusion (Aubin and Cellina 1984). The stability properties of this game dynamics were analyzed by (Hofbauer 1995) (see also Hofbauer and Sigmund 2003) who first showed that there is a solution for all $t \geq 0$ given any initial condition.[9]

The best response dynamics (10.8) is a special case of a general dynamics of the form

$$\dot{p} = I(p)p - p \tag{10.9}$$

where $I_{ij}(p)$ is the probability an individual switches from strategy j to strategy i per unit time if the current state is p. Then the corresponding continuous-time game dynamics in vector form is then given by (10.9) where $I(p)$ is the $m \times m$ matrix with entries $I_{ij}(p)$. The transition matrix $I(p)$ can also be developed using the revision protocol approach promoted by Sandholm (2010).

The best response dynamics (10.8) results by always switching to the best strategy when a revision opportunity arises in that $I_{ij}(p)$ is given by

$$I_{ij}(p) = \begin{cases} 1 & \text{if } e_i = \arg\max \pi(e_j, p) \\ 0 & \text{otherwise} \end{cases} . \tag{10.10}$$

The Folk Theorem is valid when the best response dynamics replaces the replicator equation (Hofbauer and Sigmund 2003) as are parts (b) and (c) of Theorem 2.

[8] On the other hand, an ESS remains locally asymptotically stable for all selection dynamics that are uniformly monotone according to Cressman (2003) (see also Sandholm 2010).

[9] Since the best response dynamics is a differential inclusion, it is sometimes written as $\dot{p} \in BR(p) - p$, and there may be more than one solution to an initial value problem (Hofbauer and Sigmund 2003). Due to this, it is difficult to provide an explicit formula for the vector field corresponding to a particular solution of (10.8) when $BR(p)$ is multivalued. Since such complications are beyond the scope of this chapter, the vector field is only given when $BR(p)$ is a single point for the examples in this section (see, e.g., the formula in (10.10)).

In contrast to the replicator equation, convergence to the NE may occur in finite time (compare Fig. 10.2, panels (a) and (c)).

The replicator equation (10.4) can also be expressed in the form (10.9) using the proportional imitation rule (Schlag 1997) given by

$$I_{ij}(p) = \begin{cases} kp_i(\pi(e_i, p) - \pi(e_j, p)) & \text{if } \pi(e_i, p) \geq \pi(e_j, p) \\ 0 & \text{if } \pi(e_i, p) < \pi(e_j, p) \end{cases}$$

for $i \neq j$. Here k is a positive proportionality constant for which $\sum_{i \neq j} I_{ij}(p) \leq 1$ for all $1 \leq j \leq m$ and $p \in \Delta^m$. Then, since $I(p)$ is a transition matrix, $I_{jj}(p) = 1 - \sum_{i \neq j} I_{ij}(p)$. This models the situation where information is gained by sampling a random individual and then switching to the sampled individual's strategy with probability proportional to the payoff difference only if the sampled individual has higher fitness.

An interesting application of these dynamics is to the following single-species habitat selection game.

Example 1 (Habitat Selection Game and IFD). The foundation of the habitat selection game for a single species was laid by Fretwell and Lucas (1969) before evolutionary game theory appeared. They were interested in predicting how a species (specifically, a bird species) of fixed population size should distribute itself among several resource patches if individuals would move to patches with higher fitness. They argued the outcome will be an ideal free distribution (IFD) defined as a patch distribution whereby the fitness of all individuals in any occupied patch would be the same and at least as high as what would be their fitness in any unoccupied patch (otherwise some individuals would move to a different patch). If there are H patches (or habitats) and an individual's pure strategy e_i corresponds to being in patch i (for $i = 1, 2, \ldots, H$), we have a population game by equating the payoff of e_i to the fitness in this patch. The verbal description of an IFD in this "habitat selection game" is then none other than that of an NE. Although Fretwell and Lucas (1969) did not attach any dynamics to their model, movement among patches is discussed implicitly.

If patch fitness is decreasing in patch density (i.e., in the population size in the patch), Fretwell and Lucas proved that there is a unique IFD at each fixed total population size.[10] Moreover, the IFD is an ESS that is globally asymptotically stable under the replicator equation (Cressman and Krivan 2006; Cressman et al. 2004; Krivan et al. 2008). To see this, let $p \in \Delta^H$ be a distribution among the patches and $\pi(e_i, p)$ be the fitness in patch i. Then $\pi(e_i, p)$ depends only on the proportion

[10] Broom and Krivan (▶ Chap. 23, "Biology and Evolutionary Games", this volume) give more details of this result and use it to produce analytic expressions for the IFD in several important biological models. They also generalize the IFD concept when the assumptions underlying the analysis of Fretwell and Lucas (1969) are altered. Here, we concentrate on the dynamic stability properties of the IFD in its original setting.

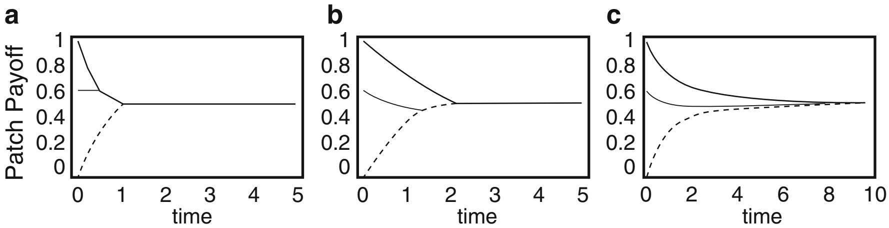


Fig. 10.2 Trajectories for payoffs of the habitat selection game when initially almost all individuals are in patch 2 and patch payoff functions are $\pi(e_1, p) = 1 - p_1, \pi(e_2, p) = 0.8(1 - \frac{10p_2}{9})$ and $\pi(e_3, p) = 0.6(1 - \frac{10p_3}{8})$. (**a**) Best response dynamics with migration matrices of the form $I^1(p)$; (**b**) dynamics for nonideal animals with migration matrices of the form $I^2(p)$; and (**c**) the replicator equation. In all panels, the *top curve* is the payoff in patch 1, the *middle curve* in patch 3, and the *bottom curve* in patch 2. The IFD (which is approximately $(p_1, p_2, p_3) = (0.51, 0.35, 0.14)$ with payoff 0.49) is reached at the smallest t where all three curves are the same (this takes infinite time in panel **c**)

p_i in this patch (i.e., has the form $\pi(e_i, p_i)$). To apply matrix game techniques, assume this is a linearly decreasing function of p_i.[11] Then, since the vector field $(\pi(e_1, p_1), \ldots, \pi(e_H, p_H))$ is the gradient of a real-valued function $F(p)$ defined on Δ^H, we have a potential game. Following Sandholm (2010), it is a strictly stable game and so has a unique ESS p^* which is globally asymptotically stable under the replicator equation. In fact, it is globally asymptotically stable under many other game dynamics as well that satisfy the intuitive conditions in the following Theorem.[12]

Theorem 3. *Suppose patch fitness is a decreasing function of patch density in a single-species habitat selection game. Then any migration dynamics* (10.9) *that satisfies the following two conditions evolves to the unique IFD.*

(a) *Individuals never move to a patch with lower fitness.*
(b) *If there is a patch with higher fitness than some occupied patch, some individuals move to a patch with highest fitness.*

[11]The results summarized in this example do not depend on linearity as shown in Krivan et al. (2008) (see also Cressman and Tao 2014).

[12]To see that the habitat selection game is a potential game, take $F(p) \equiv \sum_{i=1}^{H} \int_0^{p_i} \pi(e_i, u_i) du_i$. Then $\frac{\partial F(p)}{\partial p_i} = \pi(e_i, p_i)$. If patch payoff decreases as a function of patch density, the habitat selection game is a strictly stable game (i.e., $\sum (p_i - q_i)(\pi(e_i, p) - \pi(e_i, q)) < 0$ for all $p \neq q$ in Δ^H). This follows from the fact that $F(p)$ is strictly concave since $\frac{\partial^2 F(p)}{\partial p_i \partial p_j} = \begin{cases} \frac{\partial \pi(e_i, p_i)}{\partial p_i} & \text{if } i = j \\ 0 & \text{if } i \neq j \end{cases}$ and $\frac{\partial \pi(e_i, p_i)}{\partial p_i} < 0$. Global asymptotic stability of p^* for any dynamics (10.9) that satisfies the conditions of Theorem 3 follows from the fact that $W(p) \equiv \max_{1 \leq i \leq H} \pi(e_i, p)$ is a (decreasing) Lyapunov function (Krivan et al. 2008).

We illustrate Theorem 3 when there are three patches. Suppose that at p, patch fitnesses are ordered $\pi(e_1, p) > \pi(e_2, p) > \pi(e_3, p)$ and consider the two migration matrices

$$I^1(p) \equiv \begin{bmatrix} 1 & 1 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \qquad I^2(p) \equiv \begin{bmatrix} 1 & 1/3 & 1/3 \\ 0 & 2/3 & 1/3 \\ 0 & 0 & 1/3 \end{bmatrix}.$$

$I^1(p)$ corresponds to a situation where individuals who move go to patch 1 since they know it has highest fitness and so is associated with the best response dynamics (10.8).

On the other hand, $I^2(p)$ models individuals who only gain fitness information by sampling one patch at random, moving to this patch if it has higher fitness than its current patch (e.g., an individual in patch 2 moves if it samples patch 1 and otherwise stays in its own patch (with probabilities $1/3$ and $2/3$, respectively)). Trajectories for each of these two migration dynamics with the same initial conditions are illustrated in Fig. 10.2a, b, respectively. As can be seen, both converge to the IFD (as they must by Theorem 3) in finite time, even though their paths to this rational outcome are quite different. For comparison's sake, Fig. 10.2c provides the trajectory for the replicator equation.

Fretwell and Lucas (1969) briefly consider their IFD concept when patch fitness increases with patch density when density is low (the so-called Allee effect). Although Theorem 3 no longer applies, these habitat selection games are still potential games (but not strictly stable). Thus, each interior trajectory under many game dynamics (including the replicator equation and best response dynamics) converges to an NE (Sandholm 2010). Several NE are already possible for a given two-patch model (Morris 2002), some of which are locally asymptotically stable and some not (Cressman and Tran 2015; Krivan 2014). From these references, it is clear that there is a difference of opinion whether to define IFD as any of these NEs or restrict the concept to only those that are locally superior and/or asymptotically stable.

Habitat selection games also provide a natural setting for the effect of evolving population sizes, a topic of obvious importance in population biology that has so far received little attention in models of social behavior. A "population-migration" dynamics emerges if population size N evolves through fitness taken literally as reproductive success (Cressman and Krivan 2013; Broom and Krivan's (▸ Chap. 23, "Biology and Evolutionary Games", this volume)). As discussed there, if patch fitness is positive when unoccupied, decreases with patch density, and eventually becomes negative, then the system evolves to carrying capacity whenever the migration matrix $I(p; N)$ satisfies the two conditions in Theorem 3 for each population size N. In particular, the evolutionary outcome is independent of the time scale of migration compared to that of changing population size, a notable result since it is often not true when two dynamical processes are combined (Cressman and Krivan 2013).

3 Symmetric Games with a Continuous Trait Space

It was recognized early on that the relationship between evolutionary outcomes and the ESS concept is more complex when an individual's pure strategy (or trait) is taken from a continuum. As stated by Maynard Smith (1982, Appendix D), "Eshel (1983) and A. Grafen (personal communication) have independently noticed a criterion for the stability of an ESS which is different in kind ... when the strategy set is a continuous one." Although Eshel and Grafen (see also Eshel and Motro 1981) both came to this realization by analyzing examples of evolutionary outcomes for games from biology (e.g., stable sex ratios; investment in a competitive trait) or economics (e.g., adaptation of prices to demand), the issues that arise can already be illustrated by the following elementary mathematical example.

Consider the symmetric game whose pure-strategy set S is parameterized by real number v in an interval that contains 0 in its interior. Suppose the payoff $\pi(v,u)$ of strategy v against strategy u has the form

$$\pi(v,u) = av^2 + buv \qquad (10.11)$$

where a and b are fixed parameters (which are real numbers). It is straightforward to check that 0 is a strict NE if and only if $a < 0$.[13] Then, with the assumption that $a < 0$, $u^* = 0$ is an ESS according to Definition 1 (a) and (b) and so cannot be invaded.[14]

On the other hand, a strategy v against a monomorphic population using strategy u satisfies

$$\pi(v,u) > \pi(u,u) \text{ if and only if } (v-u)[a(v+u)+bu] > 0. \qquad (10.12)$$

For v close to $u \neq 0$, $a(v+u)+bu \cong (2a+b)u$. Thus, if $2a+b > 0$, then strategy v close to $u \neq 0$ can invade if v is farther from $u^* = 0$ than u (since $\pi(v,u) > \pi(u,u)$) but cannot invade if it is closer to u^* than u. Thus, if the system is slightly perturbed from $u^* = 0$ to $u \neq 0$, the monomorphic population will evolve away from u^*. That is, although the strict NE $u^* = 0$ when $a < 0$ cannot be invaded, it is not attainable as an outcome of evolution when $2a + b > 0$.

This result led Eshel (1983) to propose the continuously stable strategy (CSS) concept that requires more than u^* being a strict NE and others to develop (the canonical equation of) adaptive dynamics (see Definition 2 below and the literature cited there). Furthermore, although a strict NE is automatically locally superior for

[13] Specifically, $\pi(v,0) = av^2 < 0 = \pi(0,0)$ for all $v \neq 0$ if and only if $a < 0$.

[14] Much of the literature on evolutionary games for continuous trait space uses the term ESS to denote a strategy that is uninvadable in this sense. However, this usage is not universal. Since ESS has in fact several possible connotations for games with continuous trait space (Apaloo et al. 2009), we prefer to use the more neutral game-theoretic term of strict NE in these circumstances when the game has a continuous trait space.

matrix games (as in Theorem 2(a)), this is no longer true for games with continuous trait space. This discrepancy leads to the concept of a neighborhood invader strategy (NIS) in Sect. 3.2 below that is closely related to stability with respect to the replicator equation (see Theorem 2 there).

3.1 The CSS and Adaptive Dynamics

To avoid mathematical technicalities that arise in threshold cases, the following definition assumes that, if u^* is a pure-strategy NE in the interior of an interval S that is the pure-strategy set of the evolutionary game, then u^* is a strict NE.

Definition 2. (a) A strict NE u^* in the interior of a one-dimensional continuous strategy set is a CSS if, for some $\varepsilon > 0$ and any u with $\mid u - u^* \mid < \varepsilon$, there is a $\delta > 0$ such that, for $\mid v - u \mid < \delta$, $\pi(v,u) > \pi(u,u)$ if and only if $\mid v - u^* \mid < \mid u - u^* \mid$.[15]

(b) Up to a change in time scale, the canonical equation of adaptive dynamics is

$$\dot{u} = \frac{\partial \pi(v,u)}{\partial v} \mid_{v=u} \equiv \pi_1(u,u). \tag{10.13}$$

(c) An interior u^* is called convergence stable if it is locally asymptotically stable under (10.13).

To paraphrase Eshel (1983), the intuition behind Definition 2(a) is that, if a large majority of a population chooses a strategy close enough to a CSS, then only those mutant strategies which are even closer to the CSS will be selectively advantageous.

The canonical equation of adaptive dynamics (10.13) is the most elementary dynamics to model evolution in a one-dimensional continuous strategy set. It assumes that the population is always monomorphic at some $u \in S$ and that u evolves through trait substitution in the direction v of nearby mutants that can invade due to their higher payoff than u when playing against this monomorphism. Adaptive dynamics (10.13) was introduced by Hofbauer and Sigmund (1990) assuming monomorphic populations. It was given a more solid interpretation when populations are only approximately monomorphic by Dieckmann and Law (1996) (see also Dercole and Rinaldi 2008; Vincent and Brown 2005) where $\dot{u} = k(u)\pi_1(u,u)$ and $k(u)$ is a positive function that scales the rate of evolutionary change. Typically, adaptive dynamics is restricted to models for which $\pi(v,u)$ has

[15]Typically, $\delta > 0$ depends on u (e.g., $\delta < \mid u - u^* \mid$). Sometimes the assumption that u^* is a strict NE is relaxed to the condition of being a neighborhood (or local) strict NE (i.e., for some $\varepsilon > 0$, $\pi(v,u) < \pi(u,u)$ for all $0 < \mid v - u \mid < \varepsilon$).

continuous partial derivatives up to (at least) the second order.[16] Since invading strategies are assumed to be close to the current monomorphism, their success can then be determined through a local analysis.

Historically, convergence stability was introduced earlier than the canonical equation as a u^* that satisfies the second part of Definition 2 (a).[17] In particular, a convergence stable u^* may or may not be a strict NE. Furthermore, u^* is a CSS if and only if it is a convergence stable strict NE. These subtleties can be seen by applying Definition 2 to the game with quadratic payoff function (10.11) whose corresponding canonical equation is $\dot{u} = (2a + b)u$. The rest point (often called a singular point in the adaptive dynamics literature) $u^* = 0$ of (10.13) is a strict NE if and only if $a < 0$ and convergence stable if and only if $2a + b < 0$. From (10.12), it is clear that u^* is a CSS if and only if it is convergence stable and a strict NE.

That the characterization[18] of a CSS as a convergence stable strict NE extends to general $\pi(v, u)$ can be seen from the Taylor expansion of $\pi(u, v)$ about (u^*, u^*) up to the second order, namely,

$$\begin{aligned} \pi(u, v) = \pi(u^*, u^*) + \pi_1(u^*, u^*)(u - u^*) + \pi_2(u^*, u^*)(v - u^*) \\ + \frac{1}{2}\pi_{11}(u^*, u^*)(u - u^*)^2 + \pi_{12}(u^*, u^*)(u - u^*)(v - u^*) \\ + \frac{1}{2}\pi_{22}(u^*, u^*)(v - u^*)^2 + \text{ higher order terms.} \end{aligned} \quad (10.14)$$

That is, u^* is convergence stable if and only if $\pi_1(u^*, u^*) = 0$ and $\pi_{11}(u^*, u^*) + \pi_{12}(u^*, u^*) < 0$ since $\frac{d}{du}\left[\frac{\partial \pi(v, u)}{\partial v}|_{v=u}\right]|_{u=u^*} = \pi_{11}(u^*, u^*) + \pi_{12}(u^*, u^*)$. It is a CSS if and only if it is a neighborhood strict NE (i.e., $\pi_1(u^*, u^*) = 0$ and $\pi_{11}(u^*, u^*) < 0$) that is convergence stable. From now on, assume all partial derivatives are evaluated at (u^*, u^*) (e.g., $\pi_{11} = \pi_{11}(u^*, u^*)$).

[16]In particular, adaptive dynamics is not applied to examples such as the War of Attrition, the original example of a symmetric evolutionary game with a continuous trait space (Maynard Smith 1974, 1982; Broom and Krivan, (▶ Chap. 23, "Biology and Evolutionary Games", this volume), which have discontinuous payoff functions. In fact, by allowing invading strategies to be far away or individuals to play mixed strategies, it is shown in these references that the evolutionary outcome for the War of Attrition is a continuous distribution over the interval S. Distributions also play a central role in the following section. Note that, in Sect. 3, subscripts on π denote partial derivatives. For instance, the derivative of π with respect to the first argument is denoted by π_1 in (10.13). For the asymmetric games of Sect. 4, π_1 and π_2 denote the payoffs to player one and to player two, respectively.

[17]This concept was first called m-stability by Taylor (1989) and then convergence stability by Christiansen (1991), the latter becoming standard usage. It is straightforward to show that the original definition is equivalent to Definition 2 (c).

[18]This general characterization of a CSS ignores threshold cases where $\pi_{11}(u^*, u^*) = 0$ or $\pi_{11}(u^*, u^*) + \pi_{12}(u^*, u^*) = 0$. We assume throughout Sect. 3 that these degenerate situations do not arise for our payoff functions $\pi(v, u)$.

Since conditions for convergence stability are independent of those for strict NE, there is a diverse classification of singular points. Circumstances where a rest point u^* of (10.13) is convergence stable but not a strict NE (or vice versa) have received considerable attention in the literature. In particular, u^* can be a convergence stable rest point without being a neighborhood strict NE (i.e., $\pi_1 = 0$, $\pi_{11} + \pi_{12} < 0$ and $\pi_{11} > 0$). These have been called evolutionarily stable minima (Abrams et al. 1993)[19] and bifurcation points (Brown and Pavlovic 1992) that produce evolutionary branching (Geritz et al. 1998) via adaptive speciation (Cohen et al. 1999; Doebeli and Dieckmann 2000; Ripa et al. 2009). For (10.11), the evolutionary outcome is then a stable dimorphism supported on the endpoints of the interval S when (the canonical equation of) adaptive dynamics is generalized beyond the monomorphic model to either the replicator equation (see Remark 2 in Sect. 3.2) or to the Darwinian dynamics of Sect. 3.3.

Conversely, u^* can be a neighborhood strict NE without being a convergence stable rest point (i.e., $\pi_1 = 0$, $\pi_{11} + \pi_{12} > 0$ and $\pi_{11} < 0$). We now have bistability under (10.13) with the monomorphism evolving to one of the endpoints of the interval S.

3.2 The NIS and the Replicator Equation

The replicator equation (10.4) of Sect. 2 has been generalized to symmetric games with continuous trait space S by Bomze and Pötscher (1989) (see also Bomze 1991; Oechssler and Riedel 2001). When payoffs result from pairwise interactions between individuals and $\pi(v, u)$ is interpreted as the payoff to v against u, then the expected payoff to v in a random interaction is $\pi(v, P) \equiv \int_S \pi(v, u) P(du)$ where P is the probability measure on S corresponding to the current distribution of the population's strategies. With $\pi(P, P) \equiv \int_S \pi(v, P) P(dv)$ the mean payoff of the population and B a Borel subset of S, the replicator equation given by

$$\frac{dP_t}{dt}(B) = \int_B \left(\pi(u, P_t) - \pi(P_t, P_t)\right) P_t(du) \tag{10.15}$$

is well defined on the set of Borel probability measures $P \in \Delta(S)$ if the payoff function is continuous (Oechssler and Riedel 2001). The replicator equation describes how the population strategy distribution $P \in \Delta(S)$ evolves over time. From this perspective, the canonical equation (10.13) becomes a heuristic tool that approximates the evolution of the population mean by ignoring effects due to the diversity of strategies in the population (Cressman and Tao 2014).

For instance, if B is a subinterval of S, (10.15) describes how the proportion $P_t(B)$ of the population with strategy in this set evolves over time. In general, B can

[19]We particularly object to this phrase since it causes great confusion with the ESS concept. We prefer calling these evolutionary branching points.

be any Borel subset of S (i.e., any element of the smallest $\sigma-$algebra that contains all subintervals of S). In particular, if B is the finite set $\{u_1, \ldots, u_m\}$ and $P_0(B) = 1$ (i.e., the population initially consists of m strategy types), then $P_t(B) = 1$ for all $t \geq 1$ and (10.15) becomes the replicator equation (10.4) for the matrix game with $m \times m$ payoff matrix whose entries are $A_{ij} = \pi(u_i, u_j)$.

Unlike adaptive dynamics, a CSS may no longer be stable for the replicator equation (10.15). To see this, a topology on $\Delta(S)$ is needed. In the weak topology, $Q \in \Delta(S)$ is close to a $P \in \Delta(S)$ with finite support $\{u_1, \ldots, u_m\}$ if the $Q-$measure of a small neighborhood of each u_i is close to $P(\{u_i\})$ for all $i = 1, \ldots, m$. In particular, if the population P is monomorphic at a CSS u^* (i.e., P is the Dirac delta distribution δ_{u^*} with all of its weight on u^*), then any neighborhood of P will include all populations whose support is close enough to u^*. Thus, stability of (10.15) with respect to the weak topology requires that P_t evolves to δ_{u^*} whenever P_0 has support $\{u, u^*\}$ where u is near enough to u^*. That is, u^* must be globally asymptotically stable for the replicator equation (10.4) of Sect. 2 applied to the two-strategy matrix game with payoff matrix (c.f. (10.5))

$$\begin{array}{c} \\ u^* \\ u \end{array} \begin{array}{c} \begin{array}{cc} u^* & u \end{array} \\ \left[\begin{array}{cc} \pi(u^*, u^*) & \pi(u^*, u) \\ \pi(u, u^*) & \pi(u, u) \end{array} \right] \end{array}.$$

Ignoring threshold circumstances again, u^* must strictly dominate u in this game (i.e., $\pi(u^*, u^*) > \pi(u, u^*)$ and $\pi(u^*, u) > \pi(u, u)$).

When this dominance condition is applied to the game with payoff function (10.11), $u^* = 0$ satisfies $\pi(u^*, u^*) > \pi(u, u^*)$ (respectively, $\pi(u^*, u) > \pi(u, u)$) if and only if $a < 0$ (respectively, $a + b < 0$). Thus, if u^* is a strict NE (i.e., $a < 0$) and $2a + b < 0 < a + b$, then u^* is a CSS that is an unstable rest point of (10.15) with respect to the weak topology.

For general payoff functions, a monomorphic population δ_{u^*} is a stable rest point of (10.15) with respect to the weak topology if and only if $\pi(u^*, u^*) > \pi(u, u^*)$ and $\pi(u^*, u) > \pi(u, u)$ for all u sufficiently close but not equal to u^*.[20] This justifies the first part of the following definition.

Definition 3. Consider a symmetric game with continuous trait space S.

(a) $u^* \in S$ is a neighborhood invader strategy (NIS) if $\pi(u^*, u) > \pi(u, u)$ for all u sufficiently close but not equal to u^*. It is a neighborhood strict NE if $\pi(u^*, u^*) > \pi(u, u^*)$ for all u sufficiently close but not equal to u^*.

[20] Here, stability means that δ_{u^*} is neighborhood attracting (i.e., for any initial distribution P_0 with support sufficiently close to u^* and with $P_0(u^*) > 0$, P_t converges to δ_{u^*} in the weak topology). As explained in Cressman (2011) (see also Cressman et al. 2006), one cannot assert that δ_{u^*} is locally asymptotically stable under the replicator equation with respect to the weak topology or consider initial distributions with $P_0(u^*) = 0$. The support of P is the closed set given by $\{u \in S \mid P(\{y : \mid y - u \mid < \varepsilon\}) > 0 \text{ for all } \varepsilon > 0\}$.

(b) Suppose $0 \leq p^* < 1$ is fixed. Strategy $u^* \in S$ is *neighborhood p^*-superior* if $\pi(u^*, P) > \pi(P, P)$ for all $P \in \Delta(S)$ with $1 > P(\{u^*\}) \geq p^*$ and the support of P sufficiently close to u^*. It is *neighborhood superior* (respectively, *neighborhood half-superior*) if $p^* = 0$ (respectively, $p^* = \frac{1}{2}$). Strategy $u^* \in S$ is *globally p^*-superior* if $\pi(u^*, P) > \pi(P, P)$ for all $P \in \Delta(S)$ with $1 > P(\{u^*\}) \geq p^*$.

The NIS concept was introduced by Apaloo (1997) (see also McKelvey and Apaloo (1995), the "good invader" strategy of Kisdi and Meszéna (1995), and the "invading when rare" strategy of Courteau and Lessard (2000)). Cressman and Hofbauer (2005) developed the neighborhood superiority concept (they called it local superiority), showing its essential equivalence to stability under the replicator equation (10.15). It is neighborhood p^*-superiority that unifies the concepts of strict NE, CSS, and NIS as well as stability with respect to adaptive dynamics and with respect to the replicator equation for games with a continuous trait space. These results are summarized in the following theorem.

Theorem 2. *Suppose that S is one dimensional and $u^* \in int(S)$ is a rest point of adaptive dynamics* (10.13) *(i.e., $\pi_1(u^*, u^*) = 0$).*

(a) *u^* is an NIS and a neighborhood strict NE if and only if it is neighborhood superior.*
(b) *u^* is neighborhood attracting with respect to the replicator equation* (10.15) *if and only if it is neighborhood superior.*
(c) *u^* is a neighborhood strict NE if and only if it is neighborhood p^*-superior for some* $0 \leq p^* < 1$.
(d) *u^* is a CSS if and only if it is neighborhood half-superior if and only if it is a neighborhood strict NE that is locally asymptotically stable with respect to adaptive dynamics* (10.13).

The proof of Theorem 2 relies on results based on the Taylor expansion (10.14). For instance, along with the characterizations of a strict NE as $\pi_{11} < 0$ and convergence stability as $\pi_{11}+\pi_{12} < 0$ from Sect. 3.1, u^* is an NIS if and only if $\frac{1}{2}\pi_{11}+\pi_{12} < 0$. Thus, strict NE, CSS, and NIS are clearly distinct concepts for a game with a continuous trait space. On the other hand, it is also clear that a strict NE that is an NIS is automatically CSS.[21]

Remark 1. When Definition 3 (b) is applied to matrix games with the standard topology on the mixed-strategy space Δ^m, the bilinearity of the payoff function implies that p is neighborhood p^*-superior for some $0 \leq p^* < 1$ if and only if

[21] This result is often stated as ESS + NIS implies CSS (e.g., Apaloo 1997; Apaloo et al. 2009). Furthermore, an ESS + NIS is often denoted ESNIS in this literature.

$\pi(p, p') > \pi(p', p')$ for all p' sufficiently close but not equal to p (i.e., if and only if p is an ESS by Theorem 2 (a)). That is, neighborhood p^*-superiority is independent of the value of p^* for $0 \leq p^* < 1$. Consequently, the ESS, NIS, and CSS are identical for matrix games or, to rephrase, NIS and CSS of Sect. 3 are different generalizations of the matrix game ESS to games with continuous trait space.

Remark 2. It was shown above that a CSS u^* which is not an NIS is unstable with respect to the replicator equation by restricting the continuous trait space to finitely many nearby strategies. However, if the replicator equation (10.15) is only applied to distributions with interval support, Cressman and Hofbauer (2005) have shown, using an argument based on the iterated elimination of strictly dominated strategies, that a CSS u^* attracts all initial distributions whose support is a small interval containing u^*. This gives a measure-theoretic interpretation of Eshel's (1983) original idea that a population would move toward a CSS by successive invasion and trait substitution. The proof in Cressman and Hofbauer (2005) is most clear for the game with quadratic payoff function (10.11). In fact, for these games, Cressman and Hofbauer (2005) give a complete analysis of the evolutionary outcome under the replicator equation for initial distributions with interval support $[\alpha, \beta]$ containing $u^* = 0$. Of particular interest is the outcome when u^* is an evolutionary branching point (i.e., it is convergence stable ($2a + b < 0$) but not a strict NE ($a > 0$)). It can then be shown that a dimorphism P^* supported on the endpoints of the interval attracts all such initial distributions except the unstable δ_{u^*}.[22]

3.3 Darwinian Dynamics and the Maximum Principle

The processes of biological evolution are inherently dynamic. Of fundamental importance is the size of the population(s) and how this evolves in the ecological system. Thus any theory of evolutionary games is incomplete without methods to address both population dynamics and strategy evolution. Sections 3.1 and 3.2 discuss two dynamics for strategy evolution when the trait space is continuous, namely, adaptive dynamics and the replicator equation, respectively. Here we present the so-called Darwinian dynamics (Vincent and Brown 2005; Vincent et al. 1993) that also considers changing population size.[23] The development of this evolutionary and ecological dynamics is informed by Darwin's postulates. The

[22] In fact, $P^* = -\frac{(a+b)\alpha + a\beta}{b(\beta - \alpha)}\delta_\beta + \frac{(a+b)\beta + a\alpha}{b(\beta - \alpha)}\delta_\alpha$ since this dimorphism satisfies $\pi(P^*, Q) > \pi(Q, Q)$ for all distributions Q not equal to P^* (i.e., P^* is globally superior by the natural extension of Definition 3 (b) to non-monomorphic P^* as developed by Cressman and Hofbauer 2005).

[23] Although Darwinian dynamics can also be based solely on changing strategy frequency with population size fixed (Vincent and Brown 2005), the theory developed here considers changing population size combined with strategy evolution.

key component of the method is the fitness-generating function (or, for short, G-function) which is given as follows.

Suppose the incumbent population currently has individuals using traits $u_1, \ldots, u_r$ taken from a continuous valued trait v in an interval S. Let $n_i > 0$ be the number of individuals using trait u_i. Then the population state is given by $(\mathbf{u}, \mathbf{n})$ where $\mathbf{u} = (u_1, \ldots, u_r)$ and $\mathbf{n} = (n_1, \ldots, n_r)$. The fitness-generating function, $G(v, \mathbf{u}, \mathbf{n})$, gives the expected per capita growth rate of a focal individual using strategy $v \in S$ when the population is in state $(\mathbf{u}, \mathbf{n})$. Interpreting this rate as reproductive success (or fitness), n_i evolves according to the population (or ecological) dynamics, $\frac{dn_i}{dt} = n_i G(v, \mathbf{u}, \mathbf{n}) \mid_{v=u_i} = n_i G(u_i, \mathbf{u}, \mathbf{n})$. Strategy evolution follows the adaptive dynamics approach, namely, $\frac{du_i}{dt} = k_i \frac{\partial G(v, \mathbf{u}, \mathbf{n})}{\partial v} \mid_{v=u_i}$ for $i = 1, \cdots, r$, where k_i is positive and represents some measure of additive genetic variance. However we will assume for simplicity that the k_i's are all the same and have common value denoted by k. Darwinian dynamics is then modeled by combining these two processes to produce the following system of differential equations:

$$\frac{dn_i}{dt} = n_i G(u_i, \mathbf{u}, \mathbf{n}) \text{ for } i = 1, \ldots, r \qquad \textbf{(ecological dynamics)} \tag{10.16}$$

and

$$\frac{du_i}{dt} = k \frac{\partial G(v, \mathbf{u}, \mathbf{n})}{\partial v} \mid_{v=u_i} \text{ for } i = 1, \ldots, r \ \textbf{(evolutionary dynamics)} \tag{10.17}$$

The rest points $(\mathbf{u}^*, \mathbf{n}^*)$ of this resident system with all components of $\mathbf{u}^*$ different and with all components of $\mathbf{n}^*$ positive that are locally (globally) asymptotically stable are expected to be the outcomes in a corresponding local (or global) sense for this r strategy resident system.

As an elementary example, the payoff function (10.11) can be extended to include population size:

$$\begin{aligned} G(v, \mathbf{u}, \mathbf{n}) &= \pi\left(v, \frac{u_1 n_1 + u_2 n_2 + \ldots + u_r n_r}{n_1 + n_2 + \ldots + n_r}\right) + 1 - (n_1 + n_2 + \ldots + n_r) \\ &= a v^2 + b\left(\frac{u_1 n_1 + u_2 n_2 + \ldots + u_r n_r}{n_1 + n_2 + \ldots + n_r}\right) v + 1 - (n_1 + n_2 + \ldots + n_r). \end{aligned} \tag{10.18}$$

The story behind this mathematical example is that v plays one random contest per unit time and receives an expected payoff $\pi\left(v, \frac{u_1 n_1 + u_2 n_2 + \ldots + u_r n_r}{n_1 + n_2 + \ldots + n_r}\right)$ since the average strategy in the population is $\frac{u_1 n_1 + u_2 n_2 + \ldots + u_r n_r}{n_1 + n_2 + \ldots + n_r}$. The term $1-(n_1 + n_2 + \ldots + n_r)$ is a strategy-independent background fitness so that fitness decreases with total population size.

For $r = 1$, $G(v, u; n) = av^2 + buv + 1 - n$. From this, the Darwinian dynamics is

$$\frac{dn}{dt} = nG(v, u; n) \mid_{v=u} = n\left((a+b)u^2 + 1 - n\right)$$
$$\frac{du}{dt} = k\frac{\partial G(v, u; n)}{\partial v} \mid_{v=u} = k(2a+b)u. \tag{10.19}$$

The rest point of the evolutionary dynamics (i.e., $\frac{du}{dt} = 0$) is $u = 0$. With $u = 0$, the relevant rest point of the ecological dynamics (i.e., $\frac{dn}{dt} = 0$) is $n = 1$. The rest point$(u^*, n^*) = (0, 1)$ of (10.19) is globally asymptotically stable for this resident system if and only if u^* is convergence stable for adaptive dynamics (i.e., $2a+b<0$) when population size is fixed at $n^* = 1$.

However, to be a stable evolutionary outcome, $(u^*, n^*) = (0, 1)$ must resist invasion by any mutant strategy using strategy $v \neq u^* = 0$. Since the invasion fitness is $G(v, u^*, n^*) = av^2$, this requires that u^* is a strict NE (i.e., $a < 0$) when population size is fixed at n^*. That is, $(u^*, n^*) = (0, 1)$ is a stable evolutionary outcome for Darwinian dynamics with respect to the G-function (10.18) if and only if u^* is a CSS.[24]

Now suppose that $u^* = 0$ is convergence stable but not a strict NE (i.e., $a > 0$ and $2a + b < 0$) and so can be invaded by $v \neq 0$ since $G(v, u^*, n^*) > 0$. We then look for a dimorphism $(\mathbf{u}^*, \mathbf{n}^*) = (u_1^*, u_2^*, n_1^*, n_2^*)$ of the resident system (i.e., $r = 2$) for Darwinian dynamics with respect to the G-function (10.18). That is, we consider the four-dimensional dynamical system

$$\frac{dn_1}{dt} = n_1 G(v, u_1, u_2; n_1, n_2) \mid_{v=u_1} = n_1\left(au_1^2 + bu_1\frac{u_1n_1 + u_2n_2}{n_1 + n_2} + 1 - (n_1 + n_2)\right)$$
$$\frac{dn_2}{dt} = n_2 G(v, u_1, u_2; n_1, n_2) \mid_{v=u_2} = n_2\left(au_2^2 + bu_2\frac{u_1n_1 + u_2n_2}{n_1 + n_2} + 1 - (n_1 + n_2)\right)$$
$$\frac{du_1}{dt} = \frac{\partial G(v, u_1, u_2; n_1, n_2)}{\partial v} \mid_{v=u_1} = k\left(2au_1 + b\frac{u_1n_1 + u_2n_2}{n_1 + n_2}\right)$$
$$\frac{du_2}{dt} = \frac{\partial G(v, u_1, u_2; n_1, n_2)}{\partial v} \mid_{v=u_2} = k\left(2au_2 + b\frac{u_1n_1 + u_2n_2}{n_1 + n_2}\right).$$

From the evolutionary dynamics, a rest point must satisfy $2au_1 = 2au_2$ and so $u_1 = u_2$ (since we assume that $a \neq 0$). That is, this two-strategy resident system has no relevant stable rest points since this requires $u_1^* \neq u_2^*$. However, it also follows from this dynamics that $\frac{d(u_1 - u_2)}{dt} = 2ka\,(u_1 - u_2)$, suggesting that the dimorphic strategies are evolving as far as possible from each other since $ka > 0$. Thus, if the strategy space S is restricted to the bounded interval $[-\beta, \beta]$, we

[24] As in Sect. 3.1, we ignore threshold cases. Here, we assume that a and $2a + b$ are both nonzero.

might expect that u_1 and u_2 evolve to the endpoints β and $-\beta$, respectively. With $(u_1^*, u_2^*) = (\beta, -\beta)$, a positive equilibrium (n_1^*, n_2^*) of the ecological dynamics must satisfy $u_1 n_1 + u_2 n_2 = 0$, and so $n_1^* = n_2^* = \frac{1+a\beta^2}{2}$. That is, the rest point is $(u_1^*, u_2^*, n_1^*, n_2^*) = \left(\beta, -\beta; \frac{1+a\beta^2}{2}, \frac{1+a\beta^2}{2}\right)$ and it is locally asymptotically stable.[25] Furthermore, it resists invasion by mutant strategies since

$$\begin{aligned} G(v, u_1^*, u_2^*, n_1^*, n_2^*) &= av^2 + bv\frac{u_1^* n_1^* + u_2^* n_2^*}{n_1^* + n_2^*} + 1 - \left(n_1^* + n_2^*\right) \\ &= a\left(v^2 - \beta^2\right) < 0 \end{aligned}$$

for all $v \in S$ different from u_1^* and u_2^*.

To summarize the above discussion of Darwinian dynamics applied to G-function (10.18) on the interval $[\beta, -\beta]$, $(u^*, n^*) = (0, 1)$ is a stable evolutionary outcome if and only if u^* is a CSS (i.e., $a < 0$ and $2a + b < 0$). On the other hand, if $a > 0$ and $2a + b < 0$, then there is evolutionary branching and the dimorphism $(\mathbf{u}^*, \mathbf{n}^*) = \left(\beta, -\beta; \frac{1+a\beta^2}{2}, \frac{1+a\beta^2}{2}\right)$ becomes a stable evolutionary outcome. These two results are shown in Fig. 10.3 (see regions II and III there, respectively) along with the stable evolutionary outcomes in other regions of parameter space a and b. For instance, although we do not have a complete analysis of Darwinian dynamics with r traits initially present, our simulations suggest that, in region I which contains the first quadrant, a bistable situation arises whereby almost all trajectories converge to one of the monomorphisms supported at one end of the interval. Similarly, in the fourth quadrant (which comprises the evolutionary branching region III as well as region IV), we expect all trajectories to converge to the dimorphism.

In fact, the use of Darwinian dynamics to confirm the results of Fig. 10.3 can be generalized to find stable evolutionary outcomes when their analysis become theoretically intractable. That is, if

(i) the Darwinian dynamics for an r strategy system converges to a locally asymptotically stable equilibrium with all strategies present and

[25] Technically, at this rest point, $\frac{du_1}{dt} = 2ka\beta > 0$ and $\frac{du_2}{dt} = -2ka\beta < 0$ are not 0. However, their sign (positive and negative, respectively) means that the dimorphism strategies would evolve past the endpoints of S, which is impossible given the constraint on the strategy space.

These signs mean that local asymptotic stability follows from the linearization of the ecological dynamics at the rest point. It is straightforward to confirm this 2×2 Jacobian matrix has negative trace and positive determinant (since $a > 0$ and $b < 0$), implying both eigenvalues have negative real part.

The method can be generalized to show that, if $S = [\alpha, \beta]$ with $\alpha < 0 < \beta$, the stable evolutionary outcome predicted by Darwinian dynamics is now $u_1^* = \beta, u_2^* = \alpha$ with $n_1^* = (a\alpha\beta - 1)\frac{(a+b)\alpha + a\beta}{b(\beta-\alpha)}, n_2^* = (1 - a\alpha\beta)\frac{(a+b)\beta + a\alpha}{b(\beta-\alpha)}$ both positive under our assumption that $a > 0$ and $2a + b < 0$. In fact, this is the same stable dimorphism (up to the population size factor $1 - a\alpha\beta$) given by the replicator equation of Sect. 3.2 (see Remark 2).

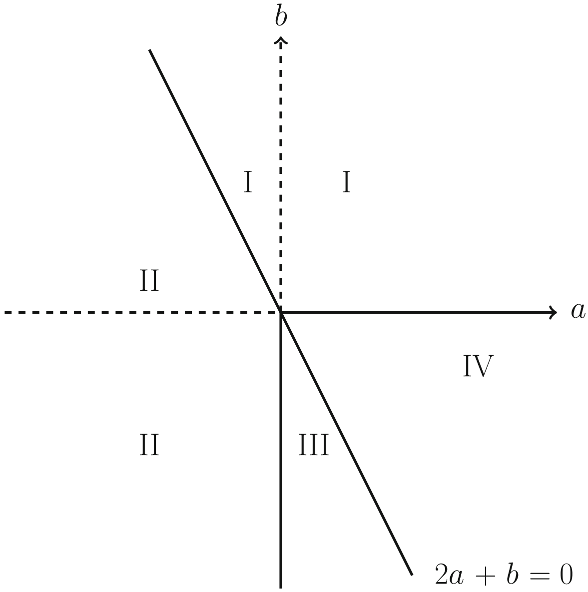


Fig. 10.3 Stable evolutionary outcomes for G-function (10.18) on the interval $[\beta, -\beta]$. From the theoretical analysis, there are four regions of parameter space (given by a and b) of interest. In region I, there are two stable evolutionary outcomes that are monomorphisms (u^*, n^*) given by $(\beta, 1 + (a + b)\beta^2)$ and $(-\beta, 1 + (a + b)\beta^2)$. In region II, the only stable evolutionary outcome is the CSS $(u^*, n^*) = (0, 1)$. In region III (evolutionary branching) and IV, the only stable evolutionary outcome is the dimorphism $(\mathbf{u}^*, \mathbf{n}^*) = \left(\beta, -\beta; \frac{1+a\beta^2}{2}, \frac{1+a\beta^2}{2}\right)$

(ii) this r strategy equilibrium remains stable when the system is increased to $r + 1$ strategies by introducing a new strategy (i.e., one strategy dies out), then we expect this equilibrium to be a stable evolutionary outcome.

On the other hand, the following Maximum Principle can often be used to find these stable evolutionary outcomes without the dynamics (or, conversely, to check that an equilibrium outcome found by Darwinian dynamics may in fact be a stable evolutionary outcome).

Theorem 5 (Maximum Principle). *Suppose that* $(\mathbf{u}^*, \mathbf{n}^*)$ *is an asymptotically stable rest point for Darwinian dynamics* (10.16) *and* (10.17) *applied to a resident system. If* $(\mathbf{u}^*, \mathbf{n}^*)$ *is a stable evolutionary outcome, then*

$$\max_{v \in S} G\left(v, \mathbf{u}^*, \mathbf{n}^*\right) = G\left(v, \mathbf{u}^*, \mathbf{n}^*\right)|_{v=u_i^*} = 0. \tag{10.20}$$

This fundamental result promoted by Vincent and Brown (see, for instance, their 2005 book) gives biologists the candidate solutions they should consider when looking for stable evolutionary outcomes to their biological systems. That is, by plotting the G-function as a function of v for a fixed candidate $(\mathbf{u}^*, \mathbf{n}^*)$, the maximum fitness must never be above 0 (otherwise, such a v could invade), and, furthermore, the fitness at each component strategy u_i^* in the $r-$strategy resident system $\mathbf{u}^*$ must be 0 (otherwise, u_i^* is not at a rest point of the ecological system). For many cases, $\max_{v \in S} G\left(v, \mathbf{u}^*, \mathbf{n}^*\right)$ occurs only at the component strategies u_i^* in

$\mathbf{u}^*$. In these circumstances, $\mathbf{u}^*$ is known as a quasi-strict NE in the game-theoretic literature (i.e., $G(v, \mathbf{u}^*, \mathbf{n}^*) \leq G(u_i, \mathbf{u}^*, \mathbf{n}^*)$ for all $i = 1, \ldots, r$ with equality if and only if $v = u_i^*$ for some i). If $r = 1$, $\mathbf{u}^*$ is a strict NE as remarked in Sect. 3.1.

When applied to the above example with G-function (10.18), $(u^*, n^*) = (0, 1)$ satisfies the Maximum Principle if and only if $a < 0$. Thus, an application of this principle is entirely consistent with the two cases examined above when $2a + b < 0$. However, one must be cautious in assuming there is an equivalence between $(\mathbf{u}^*, \mathbf{n}^*)$ being a stable evolutionary outcome and it satisfying the Maximum Principle. For instance, if $a < 0$ and $2a + b > 0$, then $(u^*, n^*) = (0, 1)$ satisfies the Maximum Principle, but it is not a stable evolutionary outcome. This was realized early on by Vincent and Brown who called a $(\mathbf{u}^*, \mathbf{n}^*)$ that satisfies the Maximum Principle a "candidate ESS" (e.g., Vincent and Brown 2005) which we would prefer to label as a "candidate stable evolutionary outcome."

As stated at the beginning of Sect. 3, the payoff function (10.11) (and its offshoot (10.18)) are used for mathematical convenience to illustrate the complex issues that arise for a game with continuous trait space. A more biologically relevant example is the so-called Lotka-Volterra (LV) competition model, whose basic $G-$function is of the form

$$G(v, \mathbf{u}, \mathbf{n}) = \frac{k}{K(v)} \left[K(v) - \sum_{j=1}^{r} a(v, u_j) n_j \right] \tag{10.21}$$

where $a(v, u_j)$ (the competition coefficient) and $K(v)$ (the carrying capacity) are given by

$$a(v, u_i) = \exp\left[-\frac{(v - u_i)^2}{2\sigma_a^2} \right] \text{ and } K(v) = K_m \exp\left[-\frac{v^2}{2\sigma_k^2} \right], \tag{10.22}$$

respectively, with trait space $\mathbf{R}$.

This particular $G-$function was already examined by Roughgarden (1979) from a non-game-theoretic approach.[26] who showed that the expected outcome of evolution for this model is a Gaussian distribution $P^*(u)$ of traits when the width of the niche axis, σ_k, and of the competition coefficient, σ_a, satisfy $\sigma_a < \sigma_k$.[27] Recently, Cressman et al. (2016) have reexamined the basic model as an

[26]Many others (e.g., Barabás and Meszéna 2009; Barabás et al. 2012, 2013; Bulmer 1974; D'Andrea et al. 2013; Gyllenberg and Meszéna 2005; Meszéna et al. 2006; Parvinen and Meszéna 2009; Sasaki 1997; Sasaki and Ellner 1995; Szabó and Meszéna 2006) have examined the general LV competition model.

[27]Specifically, the Gaussian distribution is given by

$$P^*(u) = \frac{K_m \sigma_k}{\sigma_a \sqrt{2\pi(\sigma_k^2 - \sigma_a^2)}} \exp(-u^2/(2(\sigma_k^2 - \sigma_a^2))).$$

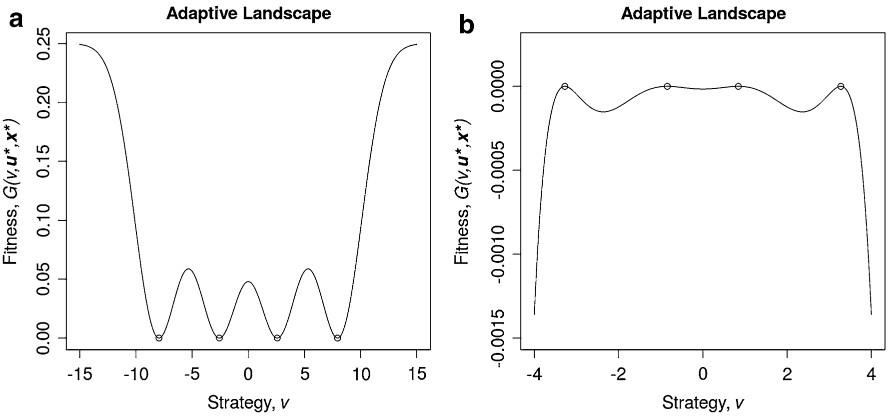


Fig. 10.4 The $G-$function $G(v, \mathbf{u}^*, \mathbf{n}^*)$ at a stable resident system $(\mathbf{u}^*, \mathbf{n}^*)$ with four traits where u_i^* for $i = 1, 2, 3, 4$ are the $v-$intercepts of the $G-$function (10.21) on the horizontal axis. (**a**) For (10.22), $(\mathbf{u}^*, \mathbf{n}^*)$ does not satisfy the Maximum Principle since $G(v, \mathbf{u}^*, \mathbf{n}^*)$ is at a minimum when $v = u_i^*$. (**b**) With carrying capacity adjusted so that it is only positive in the interval $(-\beta, \beta)$, $(\mathbf{u}^*, \mathbf{n}^*)$ does satisfy the Maximum Principle. Parameters: $\sigma_a^2 = 4, \sigma_k^2 = 200$, $K_m = 100, k = 0.1$ and for (**b**) $\beta = 6.17$

evolutionary game, using the Darwinian dynamics approach of this section. They show that, for each resident system with r traits, there is a stable equilibrium $(\mathbf{u}^*, \mathbf{n}^*)$ for Darwinian dynamics (10.16) and (10.17). However, $(\mathbf{u}^*, \mathbf{n}^*)$ does not satisfy the Maximum Principle (in fact, the components of $\mathbf{u}^*$ are minima of the G-function since $G(v, \mathbf{u}^*, \mathbf{n}^*)\ |_{v=u_i} = 0 < G(v, \mathbf{u}^*, \mathbf{n}^*)$ for all $v \neq u_i^*$ as in Fig. 10.4a). The resultant evolutionary branching leads eventually to $P^*(u)$ as the stable evolutionary outcome. Moreover, they also examined what happens when the trait space is effectively restricted to the compact interval $[-\beta, \beta]$ in place of $\mathbf{R}$ by adjusting the carrying capacity so that it is only positive between $-\beta$ and β. Now, the stable evolutionary outcome is supported on four strategies (Fig. 10.4b), satisfying the Maximum Principle (10.20).[28]

[28]Cressman et al. (2016) also examined what happens when there is a baseline competition between all individuals no matter how distant their trait values are. This leads to a stable evolutionary outcome supported on finitely many strategies as well. That is, modifications of the basic LV competition model tend to break up its game-theoretic solution $P^*(u)$ with full support to a stable evolutionary outcome supported on finitely many traits, a result consistent with the general theory developed by Barabás et al. (2012) (see also Gyllenberg and Meszéna 2005).

3.4 Symmetric Games with a Multidimensional Continuous Trait Space

The replicator equation (10.15), neighborhood strict NE, NIS, and neighborhood p^*-superiority developed in Sect. 3.2 have straightforward generalizations to multidimensional continuous trait spaces. In fact, the definitions there do not assume that S is a subset of $\mathbf{R}$ and Theorem 10.4(b) on stability of $\mathbf{u}^*$ under the replicator equation remains valid for general subsets S of $\mathbf{R}^n$ (see Theorem 6 (b) below). On the other hand, the CSS and canonical equation of adaptive dynamics (Definition 2) from Sect. 3.1 do depend on S being a subinterval of $\mathbf{R}$.

For this reason, our treatment of multidimensional continuous trait spaces will initially focus on generalizations of the CSS to multidimensional continuous trait spaces. Since these generalizations depend on the direction(s) in which mutants are more likely to appear, we assume that S is a compact convex subset of $\mathbf{R}^n$ with $u^* \in S$ in its interior. Following the static approach of Lessard (1990) (see also Meszéna et al. 2001), u^* is a neighborhood CSS if it is a neighborhood strict NE that satisfies Definition 2 (a) along each line through u^*. Furthermore, adaptive dynamics for the multidimensional trait spaces S has the form (Cressman 2009; Leimar 2009)

$$\frac{du}{dt} = C_1(u)\nabla_1 \pi(v,u) \mid_{v=u} \tag{10.23}$$

generalizing (10.13). Here $C_1(u)$ is an $n \times n$ covariance matrix modeling the mutation process (by scaling the rate of evolutionary change) in different directions (Leimar 2009).[29] We will assume that $C_1(u)$ for $u \in \text{int}(S)$ depends continuously on u. System (10.23) is called the canonical equation of adaptive dynamics (when S is multidimensional). u^* in the interior of S is called convergence stable with respect to $C_1(u)$ if it is a locally asymptotically stable rest point (also called a singular point) of (10.23).

The statement of the following theorem (and the proof of its various parts given in Cressman 2009 or Leimar 2009) relies on the Taylor expansion about (u^*, u^*) of the payoff function, namely,

$$\begin{aligned}\pi(u,v) &= \pi(u^*,u^*) + \nabla_1\pi(u^*,u^*)(u-u^*) + \nabla_2\pi(u^*,u^*)(v-u^*)\\ &\quad + \frac{1}{2}\left[(u-u^*)\cdot A(u-u^*) + 2(u-u^*)\cdot B(v-u^*) + (v-u^*)\cdot C(v-u^*)\right]\\ &\quad + \text{ higher-order terms.}\end{aligned}$$

[29] Covariance matrices C_1 are assumed to be positive definite (i.e., for all nonzero $u \in \mathbf{R}^n$, $u \cdot C_1 u > 0$) and symmetric. Similarly, a matrix A is negative definite if, for all nonzero $u \in \mathbf{R}^n$, $u \cdot Au < 0$.

Here, ∇_1 and ∇_2 are gradient vectors with respect to u and v, respectively (e.g., the ith component of $\nabla_1 \pi(u^*, u^*)$ is $\frac{\partial \pi(u', u^*)}{\partial u'_i} |_{u'=u^*}$), and A, B, C are the $n \times n$ matrices with ijth entries (all partial derivatives are evaluated at u^*):

$$A_{ij} \equiv \left[\frac{\partial^2}{\partial u'_j \partial u'_i} \pi(u', u^*) \right]; B_{ij} \equiv \left[\frac{\partial}{\partial u'_i} \frac{\partial}{\partial u_j} \pi(u', u) \right]; C_{ij} \equiv \left[\frac{\partial}{\partial u'_j} \frac{\partial}{\partial u'_i} \pi(u^*, u') \right].$$

Theorem 6. *Suppose $u^* \in int(S)$ is a rest point of (10.23) (i.e., $\nabla_1 \pi(u^*, u^*) = 0$).*

(a) *u^* is a neighborhood strict NE if and only if A is negative definite. It is convergence stable for all choices of $C_1(u)$ if and only if $A + B$ is negative definite. It is a CSS if and only if it is neighborhood half-superior if and only if it is a neighborhood strict NE that is convergence stable for all choices of $C_1(u)$.*
(b) *u^* is an NIS if and only if $\frac{1}{2}A + B$ is negative definite. It is neighborhood superior if and only if it is neighborhood attracting under the replicator equation (10.15) if and only if it is an NIS that is a neighborhood strict NE.*

Clearly, Theorem 6 generalizes the results on strict NE, CSS and NIS given in Theorem 2 of Sect. 3.2 to games with a multidimensional continuous trait space. As we have done throughout Sect. 3, these statements assume threshold cases (e.g., A or $A + B$ negative semi-definite) do not arise. Based on Theorem 6 (a), Leimar (2009) defines the concept of strong convergence stability as a u^* that is convergence stable for all choices of $C_1(u)$.[30] He goes on to show (see also Leimar 2001) that, in a more general canonical equation where $C_1(u)$ need not be symmetric but only positive definite, u^* is convergence stable for all such choices (called absolute convergence stability) if and only if $A + B$ is negative definite and symmetric.

In general, if there is no u^* that is a CSS (respectively, neighborhood superior), the evolutionary outcome under adaptive dynamics (respectively, the replicator equation) can be quite complex for a multidimensional trait space. This is already clear for multivariable quadratic payoff functions that generalize (10.11) as seen by the subtleties that arise for the two-dimensional trait space example analyzed by Cressman et al. (2006). These complications are beyond the scope of this chapter.

4 Asymmetric Games

Sections 2 and 3 introduced evolutionary game theory for two fundamental classes of symmetric games (normal form games and games with continuous trait space,

[30] A similar covariance approach was applied by Hines (1980) (see also Cressman and Hines 1984) for matrix games to show that $p^* \in \text{int}(\Delta^m)$ is an ESS if and only if it is locally asymptotically stable with respect to the replicator equation (10.4) adjusted to include an arbitrary mutation process.

respectively). Evolutionary theory also applies to non-symmetric games. An asymmetric game is a multiplayer game where the players are assigned one of N roles with a certain probability, and, to each role, there is a set of strategies. If it is a two-player game and there is only one role (i.e., $N = 1$), we then have a symmetric game as in the previous sections.

This section concentrates on two-player, two-role asymmetric games. These are also called two-species games (roles correspond to species) with intraspecific (respectively, interspecific) interactions among players in the same role (respectively, different roles). Sections 4.1 and 4.2 consider games when the players have finite pure-strategy sets $S = \{e_1, e_2, \ldots, e_m\}$ and $T = \{f_1, f_2, \ldots, f_n\}$ in roles one and two, respectively, whereas Sect. 4.3 has continuous trait spaces in each role.

4.1 Asymmetric Normal Form Games (Two-Player, Two-Role)

Following Selten (1980) (see also Cressman 2003, 2011; Cressman and Tao 2014; van Damme 1991), in a two-player asymmetric game with two roles (i.e., $N = 2$), the players interact in pairwise contests after they are assigned a pair of roles, k and ℓ, with probability $\rho_{\{k,\ell\}}$. In the two-role asymmetric normal form games, it is assumed that the expected payoffs $\pi_1(e_i; p, q)$ and $\pi_2(f_j; p, q)$ to e_i in role one (or species 1) and to f_j in role two (or species 2) are linear in the components of the population states $p \in \Delta^m$ and $q \in \Delta^n$. One interpretation of linearity is that each player engages in one intraspecific and one interspecific random pairwise interaction per unit time.

A particularly important special class, called truly asymmetric games (Selten 1980), has $\rho_{\{1,2\}} = \rho_{\{2,1\}} = \frac{1}{2}$ and $\rho_{\{1,1\}} = \rho_{\{2,2\}} = 0$. The only interactions in these games are between players in different roles (or equivalently, $\pi_1(e_i; p, q)$ and $\pi_2(f_j; p, q)$ are independent of p and q, respectively). Then, up to a possible factor of $\frac{1}{2}$ that is irrelevant in our analysis,

$$\pi_1(e_i; p, q) = \sum_{j=1}^{n} A_{ij} q_j = e_i \cdot Aq \text{ and } \pi_2(f_j; p, q) = \sum_{i=1}^{m} B_{ji} p_i = f_j \cdot Bp$$

where A and B are $m \times n$ and $n \times m$ (interspecific) payoff matrices. For this reason, these games are also called bimatrix games .

Evolutionary models based on bimatrix games have been developed to investigate such biological phenomena as male-female contribution to care of offspring in the Battle of the Sexes game of Dawkins (1976) and territorial control in the Owner-Intruder game (Maynard Smith 1982).[31] Unlike the biological interpretation of asymmetric games in most of Sect. 4 that identifies roles with separate species,

[31] These two games are described more fully in Broom and Krivan's (▶ Chap. 23, "Biology and Evolutionary Games", this volume).

the two players in both these examples are from the same species but in different roles. In general, asymmetric games can be used to model behavior when the same individual is in each role with a certain probability or when these probabilities depend on the players' strategy choices. These generalizations, which are beyond the scope of this chapter, can affect the expected evolutionary outcome (see, e.g., Broom and Rychtar 2013).

To extend the ESS definition developed in Sects. 2.1 and 2.2 to asymmetric games, the invasion dynamics of the resident monomorphic population $(p^*, q^*) \in \Delta^m \times \Delta^n$ by (p, q) generalizes (10.3) to become

$$\begin{aligned}\dot{\varepsilon}_1 &= \varepsilon_1(1-\varepsilon_1)(\pi_1(p;\varepsilon_1 p + (1-\varepsilon_1)p^*, \varepsilon_2 q + (1-\varepsilon_2)q^*) \\ &\quad -\pi_1(p^*;\varepsilon_1 p + (1-\varepsilon_1)p^*, \varepsilon_2 q + (1-\varepsilon_2)q^*)) \\ \dot{\varepsilon}_2 &= \varepsilon_2(1-\varepsilon_2)(\pi_2(q;\varepsilon_1 p + (1-\varepsilon_1)p^*, \varepsilon_2 q + (1-\varepsilon_2)q_1^*) \\ &\quad -\pi_2(q^*;\varepsilon_1 p + (1-\varepsilon_1)p^*, \varepsilon_2 q + (1-\varepsilon_2)q^*))\end{aligned} \tag{10.24}$$

where $\pi_1(p;\varepsilon_1 p + (1-\varepsilon_1)p^*, \varepsilon_2 q + (1-\varepsilon_2)q^*)$ is the payoff to p when the current states of the population in roles one and two are $\varepsilon_1 p + (1-\varepsilon_1)p^*$ and $\varepsilon_2 q + (1-\varepsilon_2)q^*$, respectively, etc. Here ε_1 (respectively, ε_2) is the frequency of the mutant strategy p in species 1 (respectively, q in species 2).

By Cressman (1992), (p^*, q^*) exhibits evolutionary stability under (10.24) (i.e., $(\varepsilon_1, \varepsilon_2) = (0, 0)$ is locally asymptotically stable under the above dynamics for all choices $p \neq p^*$ and $q \neq q^*$) if and only if

$$\text{either} \quad \pi_1(p; p, q) < \pi_1(p^*; p, q) \quad \text{or} \quad \pi_2(q; p, q) < \pi_2(q^*; p, q) \tag{10.25}$$

for all strategy pairs sufficiently close (but not equal) to (p^*, q^*). Condition (10.25) is the two-role analogue of local superiority for matrix games (see Theorem 2 (a)). If (10.25) holds for all $(p, q) \in \Delta^m \times \Delta^n$ sufficiently close (but not equal) to (p^*, q^*), then (p^*, q^*) is called a two-species ESS (Cressman 2003) or neighborhood superior (Cressman 2010).

The two-species ESS (p^*, q^*) enjoys similar evolutionary stability properties to the ESS of symmetric normal form games. It is locally asymptotically stable with respect to the replicator equation for asymmetric games given by

$$\begin{aligned}\dot{p}_i &= p_i \left[\pi_1(e_i; p, q) - \pi_1(p; p, q)\right] \text{ for } i = 1, \ldots, m \\ \dot{q}_j &= q_j \left[\pi_2(f_j; p, q) - \pi_2(q; p, q)\right] \text{ for } j = 1, \ldots, n\end{aligned} \tag{10.26}$$

and for all its mixed-strategy counterparts (i.e., (p^*, q^*) is strongly stable). Furthermore, if (p^*, q^*) is in the interior of $\Delta^m \times \Delta^n$, then it is globally asymptotically stable with respect to (10.26) and with respect to the best response dynamics that generalizes (10.8) to asymmetric games (Cressman 2003). Moreover, the Folk Theorem (Theorem 1) is valid for the replicator equation (10.26) where an NE is a strategy pair (p^*, q^*) such that $\pi_1(p; p^*, q^*) \leq \pi_1(p^*; p^*, q^*)$ for all $p \neq p^*$ and

$\pi_2(q; p^*, q^*) \leq \pi_2(q^*; p^*, q^*)$ for all $q \neq q^*$ (it is a strict NE if both inequalities are strict).

For bimatrix games, (p^*, q^*) is a two-species ESS if and only if it is a strict NE (i.e., $p \cdot Aq^* < p^* \cdot Aq^*$ for all $p \neq p^*$ and $q \cdot Bp^* < q^* \cdot Bp^*$ for all $q \neq q^*$).[32] Furthermore, for these games, (p^*, q^*) is locally asymptotically stable with respect to (10.26) if and only if it is a two-species ESS (i.e., a strict NE). Thus, in contrast to symmetric games, we have an equivalence between the static two-species ESS concept (10.25) and stable evolutionary outcomes. However, this is an unsatisfactory result in the sense that strict NE must be pure-strategy pairs and so the two-species ESS is a very restrictive concept for bimatrix games.

At the other extreme from bimatrix games, suppose that there are no interspecific interactions (e.g., $\rho_{\{1,2\}} = \rho_{\{2,1\}} = 0$ and $\rho_{\{1,1\}} = \rho_{\{2,2\}} = \frac{1}{2}$). These are also called completely symmetric two-role asymmetric games. Then, (p^*, q^*) is a two-species ESS if and only if p^* is a single-species ESS for species one and q^* is a single-species ESS for species two. For example, when $q = q^*$, we need $\pi_1(p^*; p, q^*) \equiv \pi_1(p^*, p) > \pi_1(p, p) \equiv \pi_1(p; p, q^*)$ for all p that are sufficiently close (but not equal) to p^*. From Theorem 2 (a), this last inequality characterizes the single-species ESS (of species one). From this result, there can be two-species ESSs that are not strict NE (see also Example 2 below). In particular, there can be completely mixed ESSs.

From these two extremes, we see that the concept of a two-species ESS combines and generalizes the concepts of single-species ESS of matrix games and the strict NE of bimatrix games.

A more biologically relevant example of two-species interactions analyzed by evolutionary game theory (where there are both interspecific and intraspecific interactions) is the following two-habitat selection model of Cressman et al. (2004). Specifically, this model is a Lotka-Volterra competitive two-species system in each patch where it is assumed that each species' migration is always toward the patch with the highest payoff for this species (see Example 1). An ESS always exists in this model, and, depending on parameters, the ESS is mixed (i.e., both species coexist in each patch) in some cases, while, in others, one of the species resides only in one patch at the ESS.

Example 2 (Two-species habitat selection game). Suppose that there are two species competing in two different habitats (or patches) and that the overall population size (i.e., density) of each species is fixed. Also assume that the fitness of an individual depends only on its species, the patch it is in, and the density of both species in this patch. Then strategies of species one and two can be parameterized by the proportions p_1 and q_1, respectively, of these species that are in patch one. If individual fitness (i.e., payoff) is positive when a patch is unoccupied and linearly

[32]To see this result first proven by Selten (1980), take $(p, q) = (p, q^*)$. Then (10.25) implies $p \cdot Aq^* < p^* \cdot Aq^*$ or $q^* \cdot Bp^* < q^* \cdot Bp^*$ for all $p \neq p^*$. Thus, $p \cdot Aq^* < p^* \cdot Aq^*$ for all $p \neq p^*$. The same method can now be applied to $(p, q) = (p^*, q)$.

decreasing in patch densities, then payoff functions have the form

$$\pi_1(e_i; p, q) = r_i \left(1 - \frac{p_i M}{K_i} - \frac{\alpha_i q_i N}{K_i}\right)$$
$$\pi_2(f_i; p, q) = s_i \left(1 - \frac{q_i N}{L_i} - \frac{\beta_i p_i M}{L_i}\right).$$

Here, $\pi_1(e_i; p, q)$ (respectively, $\pi_2(f_i; p, q)$) is the fitness of a species one individual (respectively, species two individual) in patch i, $p_2 = 1 - p_1$, and $q_2 = 1 - q_1$. All other parameters are fixed and positive.[33]

By linearity, these payoffs can be represented by a two-species asymmetric game with payoff matrices

$$A = \begin{bmatrix} r_1 - \frac{r_1 M}{K_1} & r_1 \\ r_2 & r_2 - \frac{r_2 M}{K_2} \end{bmatrix} \qquad B = \begin{bmatrix} -\frac{\alpha_1 r_1 N}{K_1} & 0 \\ 0 & -\frac{\alpha_2 r_2 N}{K_2} \end{bmatrix}$$

$$C = \begin{bmatrix} -\frac{\beta_1 s_1 M}{L_1} & 0 \\ 0 & -\frac{\beta_2 s_2 M}{L_2} \end{bmatrix} \qquad D = \begin{bmatrix} s_1 - \frac{s_1 N}{L_1} & s_1 \\ s_2 & s_2 - \frac{s_2 N}{L_2} \end{bmatrix}.$$

For example, $\pi_1(e_i; p, q) = e_i \cdot (Ap + Bq)$. At a rest point (p, q) of the replicator equation (10.26), all individuals present in species one must have the same fitness as do all individuals present in species two.

Suppose that both patches are occupied by each species at the rest point (p, q). Then (p, q) is an NE and (p_1, q_1) is a point in the interior of the unit square that satisfies

$$r_1 \left(1 - \frac{p_1 M}{K_1} - \frac{\alpha_1 q_1 N}{K_1}\right) = r_2 \left(1 - \frac{(1 - p_1) M}{K_2} - \frac{\alpha_2 (1 - q_1) N}{K_2}\right)$$
$$s_1 \left(1 - \frac{q_1 N}{L_1} - \frac{\beta_1 p_1 M}{L_1}\right) = s_2 \left(1 - \frac{(1 - q_1) N}{L_2} - \frac{\beta_2 (1 - p_1) M}{L_2}\right).$$

That is, these two “equal fitness" lines (which have negative slopes) intersect at (p_1, q_1) as in Fig. 10.5.

The interior NE (p, q) is a two-species ESS if and only if the equal fitness line of species one is steeper than that of species two. That is, (p, q) is an interior two-

[33]This game is also considered briefly by Broom and Krivan (▶ Chap. 23, “Biology (Application of Evolutionary Game Theory)”, this volume). There the model parameters are given biological interpretations (e.g., M is the fixed total population size of species one and K_1 is its carrying capacity in patch one, etc.). Linearity then corresponds to Lotka-Volterra type interactions. As in Example 1 of Sect. 2.4, our analysis again concentrates on the dynamic stability of the evolutionary outcomes.

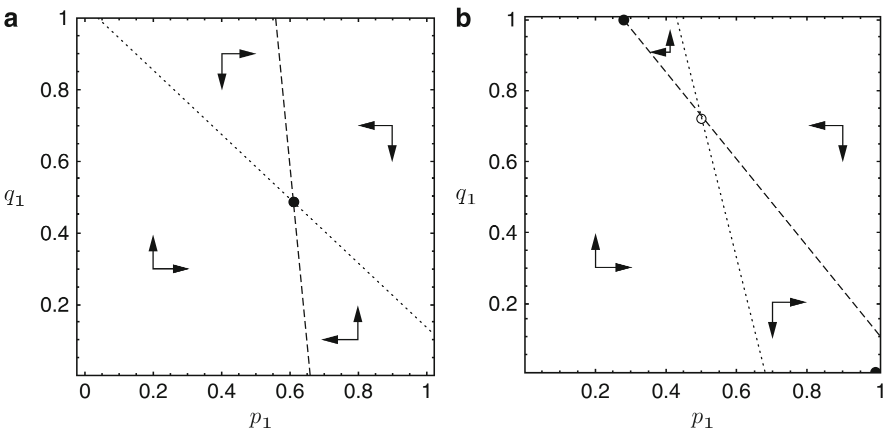


Fig. 10.5 The ESS structure of the two-species habitat selection game. The *arrows* indicate the direction of best response. The equal fitness lines of species one (*dashed line*) and species two (*dotted line*) intersect in the unit square. *Solid dots* are two-species ESSs. (**a**) A unique ESS in the interior. (**b**) Two ESSs on the boundary

species ESS in Fig. 10.5a but not in Fig. 10.5b. The interior two-species ESS in Fig. 10.5a is globally asymptotically stable under the replicator equation.

Figure 10.5b has two two-species ESSs, both on the boundary of the unit square. One is a pure-strategy pair strict NE with species one and two occupying separate patches ($p_1 = 1, q_1 = 0$), whereas the other has species two in patch one and species one split between the two patches ($0 < p_1 < 1, q_1 = 1$). Both are locally asymptotically stable under the replicator equation with basins of attraction whose interior boundaries form a common invariant separatrix. Only for initial conditions on this separatrix that joins the two vertices corresponding to both species in the same patch do trajectories evolve to the interior NE.

If the equal fitness lines do not intersect in the interior of the unit square, then there is exactly one two-species ESS. This is on the boundary (either a vertex or on an edge) and is globally asymptotically stable under the replicator equation (Krivan et al. 2008).

For these two species models, some authors consider an interior NE to be a (two-species) IFD (see Example 1 for the intuition of a single-species IFD). Example 2 shows such NE may be unstable (Fig. 10.5b) and so justifies the perspective of others who restrict the IFD concept to two-species ESSs (Krivan et al. 2008).

Remark 3. The generalization of the two-species ESS concept (10.25) to three (or more) species is a difficult problem (Cressman et al. 2001). It is shown there that it is possible to characterize a monomorphic three-species ESS as one where, at all nearby strategy distributions, at least one species does better using its ESS strategy. However, such an ESS concept does not always imply stability of the three-species replicator equation that is based on the entire set of pure strategies for each species.

4.2 Perfect Information Games

Two-player extensive form games whose decision trees describe finite series of interactions between the same two players (with the set of actions available at later interactions possibly depending on what choices were made previously) were introduced alongside normal form games by von Neumann and Morgenstern (1944). Although (finite, two-player) extensive form games are most helpful when used to represent a game with long (but finite) series of interactions between the same two players, differences with normal form intuition already emerge for short games (Cressman 2003; Cressman and Tao 2014). In fact, from an evolutionary game perspective, these differences with normal form intuition are apparent for games of perfect information with short decision trees as illustrated in the remainder of this section that follows the approach of Cressman (2011).

A (finite, two-player) *perfect information game* is given by a rooted game tree Γ where each nonterminal node is a decision point of one of the players or of nature. In this latter case, the probabilities of following each of the edges that start at the decision point and lead away from the root are fixed (by nature). A path to a node x is a sequence of edges and nodes connecting the root to x. The edges leading away from the root at each player decision node are this player's choices (or actions) at this node. There must be at least two choices at each player decision node. A pure (behavior) strategy for a player specifies a choice at all of his decision nodes. A mixed behavior strategy for a player specifies a probability distribution over the set of actions at each of his decision nodes. Payoffs to both players are specified at each terminal node $z \in Z$. A probability distribution over Z is called an outcome.

Example 3 (Chain Store game). Figure 10.6a is an elementary perfect information game with no moves by nature. At each terminal node, payoffs to both players are indicated with the payoff of player 1 above that of player 2. Player 1 has one decision node u where he chooses between the actions L and R. If he takes action L, player 1 gets payoff 1 and player 2 gets 4. If he takes action R, then we reach the decision point v of player 2 who then chooses between ℓ and r leading to both players receiving payoff 0 or both payoff 2, respectively.

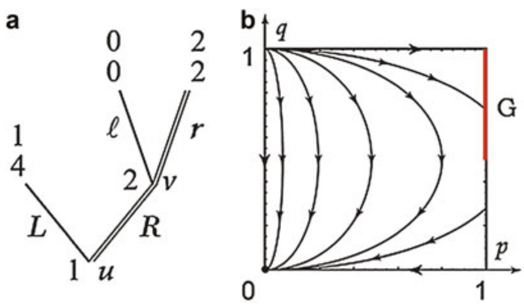


Fig. 10.6 The Chain Store game. (**a**) The extensive form. (**b**) Trajectories of the replicator equation with respect to the game's normal form and the NE structure given by the NE component G (shown as a *red line* segment) and the *solid dot* at the origin corresponding to the SPNE

What are the Nash equilibria (NEs) for this example? If players 1 and 2 choose R and r, respectively, with payoff 2 for both, then

1. player 2 does worse through unilaterally changing his strategy by playing r with probability $1-q$ less than 1 (since $0q + 2(1-q) < 2$) and
2. player 1 does worse through unilaterally changing his strategy by playing L with positive probability p (since $1p + 2(1-p) < 2$).

Thus, the strategy pair (R, r) is a strict NE corresponding to the outcome $(2, 2)$.[34]

In fact, if player 1 plays R with positive probability at an NE, then player 2 must play r. From this it follows that player 1 must play R with certainty (i.e., $p = 0$) (since his payoff of 2 is better than 1 obtained by switching to L). Thus any NE with $p < 1$ must be (R, r). On the other hand, if $p = 1$ (i.e., player 1 chooses L), then player 2 is indifferent to what strategy he uses since his payoff is 4 for any (mixed) behavior. Furthermore, player 1 is no better off by playing R with positive probability if and only if player 2 plays ℓ at least half the time (i.e., $\frac{1}{2} \leq q \leq 1$). Thus

$$G \equiv \{(L, q\ell + (1-q)r \mid \frac{1}{2} \leq q \leq 1\}$$

is a set of NE, all corresponding to the outcome $(1, 4)$. G is called an NE component since it is a connected set of NE that is not contained in any larger connected set of NE.

The NE structure of Example 3 consists of the single strategy pair $G^* = \{(R, r)\}$ which is a strict NE and the set G. These are indicated as a solid point and line segment, respectively, in Fig. 10.6b where $G^* = \{(p, q) \mid p = 0, q = 0\} = \{(0, 0)\}$.

Remark 4. Example 3 is a famous game known as the Entry Deterrence game or the Chain Store game introduced by the Nobel laureate Reinhard Selten (Selten 1978; see also van Damme 1991 and Weibull 1995). Player 2 is a monopolist who wants to keep the potential entrant (player 1) from entering the market that has a total value of 4. He does this by threatening to ruin the market (play ℓ giving payoff 0 to both players) if player 1 enters (plays R), rather than accepting the entrant (play r and split the total value of 4 to yield payoff 2 for each player). However, this is often viewed as an incredible (or unbelievable) threat since the monopolist should accept the entrant if his decision point is reached (i.e., if player 1 enters) since this gives the higher payoff to him (i.e., $2 > 0$).

[34]When the outcome is a single node, this is understood by saying the outcome is the payoff pair at this node.

Some game theorists argue that a generic perfect information game (see Remark 6 below for the definition of generic) has only one rational NE equilibrium outcome and this can be found by backward induction. This procedure starts at a final player decision point (i.e., a player decision point that has no player decision points following it) and decides which unique action this player chooses there to maximize his payoff in the subgame with this as its root. The original game tree is then truncated at this node by creating a terminal node there with payoffs to the two players given by this action. The process is continued until the game tree has no player decision nodes left and yields the subgame perfect NE (SPNE). That is, the strategy constructed by backward induction produces an NE in each subgame Γ_u corresponding to the subtree with root at the decision node u (Kuhn 1953). For generic perfect information games (see Remark 6), the SPNE is a unique pure-strategy pair and is indicated by the double lines in the game tree of Fig. 10.6a. The SPNE of Example 3 is G^*. If an NE is not subgame perfect, then this perspective argues that there is at least one player decision node where an incredible threat would be used.

Example 4 (Continued). Can evolutionary dynamics be used to select one of the two NE outcomes of the Chain Store game? Suppose players 1 and 2 use mixed strategies p and q, respectively. The payoffs of pure strategies L and R for player 1 (denoted $\pi_1(L,q)$ and $\pi_1(R,q)$, respectively) are 1 and $0q+2(1-q)$, respectively. Similarly, the payoffs of pure strategies ℓ and r for player 2 are $\pi_2(p,\ell) = 4p + (1-p)0$ and $\pi_2(p,r) = 4p+(1-p)2$, respectively. Thus, the expected payoffs are $\pi_1(p,q) = p+(1-p)2(1-q)$ and $\pi_2(p,q) = q4p+(1-q)(4p+(1-p)2)$ for players 1 and 2, respectively. Under the replicator equation, the probability of using a pure strategy increases if its payoff is higher than these expected payoffs. For this example, the replicator equation is (Weibull 1995, see also Remark 5 below)

$$\begin{aligned} \dot{p} &= p(1-(p+(1-p)2(1-q))) = p(1-p)(2q-1) \qquad (10.27)\\ \dot{q} &= q(4p-[q4p+(1-q)(4p+(1-p)2)]) = -2q(1-q)(1-p). \end{aligned}$$

The rest points are the two vertices $\{(0,0),(0,1)\}$ and the edge $\{(1,q) \mid 0 \le q \le 1\}$ joining the other two vertices. Notice that, for any interior trajectory, q is strictly decreasing and that p is strictly increasing (decreasing) if and only if $q > \frac{1}{2}$ $(q < \frac{1}{2})$.

Trajectories of (10.27) are shown in Fig. 10.6b. The SPNE of the Chain Store game G^* is the only asymptotically stable NE.[35] That is, asymptotic stability of the evolutionary dynamics selects a unique outcome for this example whereby player 1 enters the market and the monopolist is forced to accept this.

[35]This is clear for the replicator equation (10.27). For this example with two strategies for each player, it continues to hold for all other game dynamics that satisfy the basic assumption that the frequency of one strategy increases if and only if its payoff is higher than that of the player's other strategy.

Remark 5. Extensive form games can always be represented in normal form. The bimatrix normal form of Example 3 is

$$\begin{array}{r} \\ \text{Not Enter } (L) \\ \text{Enter } (R) \end{array} \begin{array}{c} \text{Ruin } (\ell) \quad \text{Accept } (r) \\ \left[\begin{array}{cc} 1,4 & 1,4 \\ 0,0 & 2,2 \end{array} \right] \end{array}.$$

By convention, player 1 is the row player and player 2 the column player. Each bimatrix entry specifies payoffs received (with player 1's given first) when the two players use their corresponding pure-strategy pair. That is, the bimatrix normal form, also denoted $\left[A, B^T\right]$, is given by

$$A = \begin{bmatrix} 1 & 1 \\ 0 & 2 \end{bmatrix} \text{ and } B = \begin{bmatrix} 4 & 4 \\ 0 & 2 \end{bmatrix}^T = \begin{bmatrix} 4 & 0 \\ 4 & 2 \end{bmatrix}$$

where A and B are the payoff matrices for player 1 and 2, respectively. With these payoff matrices, the replicator equation (10.26) becomes (10.27).

This elementary example already shows a common feature of the normal form approach for such games, namely, that some payoff entries are repeated in the bimatrix. As a normal form, this means the game is nongeneric (in the sense that at least two payoff entries in A (or in B) are the same) even though it arose from a generic perfect information game. For this reason, most normal form games cannot be represented as perfect information games.

To generalize the evolutionary analysis of Example 3 to other perfect information games, the following results for Example 3 are straightforward to prove. By Theorem 7 below, these results continue to hold for most perfect information games.

1. Every NE outcome is a single terminal node.[36]
2. Every NE component G includes a pure-strategy pair.
3. The outcomes of all elements of G are the same.
4. Every interior trajectory of the replicator equation converges to an NE.
5. Every pure-strategy NE is stable but not necessarily asymptotically stable.
6. Every NE that has a neighborhood whose only rest points are NE is stable.
7. If an NE component is interior attracting, it includes the SPNE.
8. Suppose (p, q) is an NE. It is asymptotically stable if and only if it is strict. Furthermore, (p, q) is asymptotically stable if and only if playing this strategy pair reaches every player decision point with positive probability (i.e., (p, q) is pervasive).

[36]For Example 3, this is either $(2, 2)$ or $(1, 4)$.

Theorem 7 (Cressman 2003). *Results 2 to 8 are true for all generic perfect information games. Result 1 holds for generic perfect information games without moves by nature.*

Remark 6. By definition, an extensive form game Γ is *generic* if no two pure-strategy pairs that yield different outcomes have the same payoff for one of the players. For a perfect information game Γ with no moves by nature, Γ is generic if and only if no two terminal nodes have the same payoff for one of the players. If Γ is not generic, the SPNE outcome may not be unique since several choices may arise at some player decision point in the backward induction process when there are payoff ties. Some of the results of Theorem 7 are true for general perfect information games and some are not. For instance, Result 1 is not true for some nongeneric games or for generic games with moves by nature. Result 4, which provides the basis to connect dynamics with NE in Results 5 to 8, remains an open problem for nongeneric perfect information games. On the other hand, Result 4 has recently been extended to other game dynamics. Specifically, every trajectory of the best response dynamics[37] converges to an NE component for all generic perfect information games (Xu 2016).

Theorem 7 applies to all generic perfect information games such as that given in Fig. 10.7. Since no pure-strategy pair in Fig. 10.7 can reach both the left-side subgame and the right-side subgame, none are pervasive. Thus, no NE can be asymptotically stable by Theorem 7 (Results 1 and 8), and so no single strategy pair can be selected on dynamic grounds by the replicator equation.

However, it is still possible that an NE outcome is selected on the basis of its NE component being locally asymptotically stable as a set. By Result 7, the NE component containing the SPNE is the only one that can be selected in this manner. In this regard, Fig. 10.7 is probably the easiest example (Cressman 2003) of a perfect information game where the NE component G^* of the SPNE outcome $(2, 3)$ is not interior attracting (i.e., there are interior initial points arbitrarily close to G^*

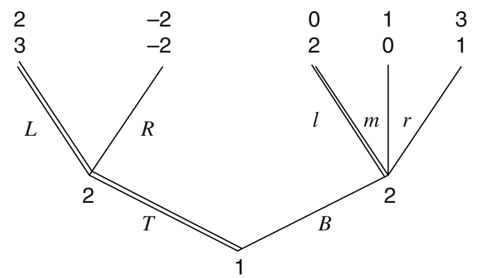


Fig. 10.7 The extensive form of a perfect information game with unstable SPNE component

[37]This is the obvious extension to bimatrix games of the best response dynamics (10.8) for symmetric (matrix) games.

whose interior trajectory under the replicator equation does not converge to this NE component). That is, Fig. 10.7 illustrates that the converse of Result 7 (Theorem 7) is not true and so the SPNE outcome is not always selected on dynamic grounds.

To see this, some notation is needed. The (mixed) strategy space of player 1 is the one-dimensional strategy simplex $\Delta(\{T, B\}) = \{(p_T, p_B) \mid p_T + p_B = 1, 0 \leq p_T, p_B \leq 1\}$. This is also denoted $\Delta^2 \equiv \{(p_1, p_2) \mid p_1 + p_2 = 1, 0 \leq p_i \leq 1\}$. Similarly, the strategy simplex for player 2 is the five-dimensional set $\Delta(\{L\ell, Lm, Lr, R\ell, Rm, Rr\}) = \{(q_{L\ell}, q_{Lm}, q_{Lr}, q_{R\ell}, q_{Rm}, q_{Rr}) \in \Delta^6\}$. The replicator equation is then a dynamics on the 6−dimensional space $\Delta(\{T, B\}) \times \Delta(\{L\ell, Lm, Lr, R\ell, Rm, Rr\})$. The SPNE component (i.e., the NE component containing the SPNE $(T, L\ell)$) is

$$G^* = \{(T, q) \mid q_{L\ell} + q_{Lm} + q_{Lr} = 1, q_{Lm} + 3q_{Lr} \leq 2\}$$

corresponding to the set of strategy pairs with outcome $(2, 3)$ where neither player can improve his payoff by unilaterally changing his strategy (Cressman 2003). For example, if player 1 switches to B, his payoff of 2 changes to $0q_{L\ell} + 1q_{Lm} + 3q_{Lr} \leq 2$. The only other pure-strategy NE is $\{B, R\ell\}$ with outcome $(0, 2)$ and corresponding NE component $G = \{(B, q) \mid q_{L\ell} + q_{R\ell} = 1, \frac{1}{2} \leq q_{R\ell} \leq 1\}$. In particular, $(T, \frac{1}{2}q_{Lm} + \frac{1}{2}q_{Lr}) \in G^*$ and $(B, R\ell) \in G$.

Using the fact that the face $\Delta(\{T, B\}) \times \Delta(\{Lm, Lr\})$ has the same structure as the Chain Store game of Example 1 (where p corresponds to the probability player 1 uses T and q the probability player 2 uses Lr), points in the interior of this face with $q_{Lr} > \frac{1}{2}$ that start close to $(T, \frac{1}{2}q_{Lm} + \frac{1}{2}q_{Lr})$ converge to (B, Lr). From this, Cressman (2011) shows that there are trajectories in the interior of the full game that start arbitrarily close to G^* that converge to a point in the NE component G. In particular, G^* is not interior attracting.

Remark 7. The partial dynamic analysis of Fig. 10.7 given in the preceding two paragraphs illustrates nicely how the extensive form structure (i.e., the game tree for this perfect information game) helps with properties of NE and the replicator equation. Similar considerations become even more important for extensive form games that are not of perfect information. For instance, all matrix games can be represented in extensive form (c.f. Remark 5), but these never have perfect information.[38] Thus, for these symmetric extensive form games (Selten 1983), the eight Results of Theorem 7 are no longer true, as we know from Sect. 2. However, the backward induction procedure can be generalized to the subgame structure of a symmetric extensive form game Γ to produce an SPNE (Selten 1983). When the

[38] An extensive form game that is not of perfect information has at least one player "information set" containing more than one decision point of this player. This player must take the same action at all these decision points. Matrix games then correspond to symmetric extensive form games (Selten 1983) where there is a bijection from the information sets of player 1 to those of player 2. Bimatrix games can also be represented in extensive form.

process is applied to NE of the (symmetric) subgames that are locally asymptotically stable under the replicator equation (10.4), a locally asymptotically stable SPNE p^* of Γ emerges (Cressman 2003) when (p^*, p^*) is pervasive (c.f. Result 8, Theorem 7). As Selten (1988) showed, this result is no longer true when local asymptotic stability is replaced by the ESS structure of the subgames. A description of the issues that arise in these games is beyond the scope of this chapter. The interested reader is directed to van Damme (1991) and Cressman (2003) as well as Selten (1983, 1988) for further details.

4.3 Asymmetric Games with One-Dimensional Continuous Trait Spaces

In this section, we will assume that the trait spaces S and T for the two roles are both one-dimensional compact intervals and that payoff functions have continuous partial derivatives up to the second order so that we avoid technical and/or notational complications. For $(u, v) \in S \times T$, let $\pi_1(u'; u, v)$ (respectively, $\pi_2(v'; u, v)$) be the payoff to a player in role 1 (respectively, in role 2) using strategy $u' \in S$ (respectively, $v' \in T$) when the population is monomorphic at (u, v). Note that π_1 has a different meaning here than in Sect. 3 where it was used to denote a partial derivative (e.g., equation (10.13)). Here, we extend the concepts of Sect. 3 (CSS, adaptive dynamics, NIS, replicator equation, neighborhood superior, Darwinian dynamics) to asymmetric games.

To start, the canonical equation of adaptive dynamics (c.f. (10.13)) becomes

$$\begin{aligned} \dot{u} &= k_1(u, v)\frac{\partial}{\partial u'}\pi_1(u'; u, v) \mid_{u'=u} \\ \dot{v} &= k_2(u, v)\frac{\partial}{\partial v'}\pi_2(v'; u, v) \mid_{v'=v} \end{aligned} \tag{10.28}$$

where $k_i(u, v)$ for $i = 1, 2$ are positive continuous functions of (u, v) . At an interior rest point (u^*, v^*) of (10.28),

$$\frac{\partial \pi_1}{\partial u'} = \frac{\partial \pi_2}{\partial v'} = 0.$$

Following Cressman (2009), (u^*, v^*) is called convergence stable if it is locally asymptotically stable under (10.28) for any choice of k_1 and k_2. Furthermore, (u^*, v^*) is a neighborhood strict NE if $\pi_1(u'; u^*, v^*) < \pi_1(u^*; u^*, v^*)$ and $\pi_2(v'; u^*, v^*) < \pi_2(v^*; u^*, v^*)$ for all u' and v' sufficiently close but not equal to u^* and v^*, respectively. Clearly, a neighborhood strict NE (u^*, v^*) in the interior of $S \times T$ is a rest point of (10.28).

The characterizations of convergence stability and strict NE in the following theorem are given in terms of the linearization of (10.28) about (u^*, v^*), namely,

$$\begin{bmatrix} \dot{u} \\ \dot{v} \end{bmatrix} = \begin{bmatrix} k_1(u^*, v^*) & 0 \\ 0 & k_2(u^*, v^*) \end{bmatrix} \begin{bmatrix} A+B & C \\ D & E+F \end{bmatrix} \begin{bmatrix} u-u^* \\ v-v^* \end{bmatrix} \tag{10.29}$$

where

$$A \equiv \frac{\partial^2}{\partial u' \partial u'} \pi_1(u'; u^*, v^*); B \equiv \frac{\partial}{\partial u'} \frac{\partial}{\partial u} \pi_1(u'; u, v^*); C \equiv \frac{\partial}{\partial u'} \frac{\partial}{\partial v} \pi_1(u'; u^*, v)$$

$$D \equiv \frac{\partial}{\partial v'} \frac{\partial}{\partial u} \pi_2(v'; u, v^*); E \equiv \frac{\partial}{\partial v'} \frac{\partial}{\partial v} \pi_2(v'; u^*, v); F \equiv \frac{\partial^2}{\partial v' \partial v'} \pi_2(v'; u^*, v^*)$$

and all partial derivatives are evaluated at the equilibrium. If threshold values involving these six second-order partial derivatives are ignored throughout this section, the following result is proved by Cressman (2010, 2011) using the Taylor series expansions of $\pi_1(u'; u, v)$ and $\pi_2(v'; u, v)$ about (u^*, v^*) that generalize (10.14) to three variable functions.

Theorem 6. *Suppose (u^*, v^*) is a rest point of (10.28) in the interior of $S \times T$.*

(a) *(u^*, v^*) is a neighborhood strict NE if and only if A and F are negative.*
(b) *(u^*, v^*) is convergence stable if and only if, for all nonzero $(u, v) \in \mathbf{R}^2$, either $u((A+B)u + Cv) < 0$ or $v(Du + (E+F)v) < 0$ if and only if $A+B < 0$, $E+F < 0$ and $(A+B)(E+F) > CD$.*[39]

In Sects. 3.1 and 3.4, it was shown that a CSS for symmetric games is a neighborhood strict NE that is convergence stable under all adaptive dynamics (e.g., Theorem 6 (a)). For asymmetric games, we define a CSS as a neighborhood strict NE that is convergence stable. That is, (u^*, v^*) is a CSS if it satisfies both parts of Theorem 6. Although the inequalities in the latter part of (b) are the easiest to use to confirm convergence stability in practical examples, it is the first set of inequalities that is most directly tied to the theory of CSS, NIS, and neighborhood p^*-superiority (as well as stability under evolutionary dynamics), especially as the trait spaces become multidimensional. It is again neighborhood superiority according to the following definition that unifies this theory (see Theorem 9 below).

Definition 4. Suppose (u^*, v^*) is in the interior of $S \times T$.

(a) Fix $0 \leq p^* < 1$. Strategy pair (u^*, v^*) is *neighborhood p^*-superior* if

$$\text{either } \pi_1(u^*; P, Q) > \pi_1(P; P, Q) \text{ or } \pi_2(v^*; P, Q) > \pi_2(Q; P, Q) \tag{10.30}$$

[39] These equivalences are also shown by Leimar (2009) who called the concept strong convergence stability.

for all $(P, Q) \in \Delta(S)\times\Delta(T)$ with $1 \geq P(\{u^*\}) \geq p^*$, $1 \geq Q(\{v^*\}) \geq p^*$ and the support of (P, Q) sufficiently close (but not equal) to (u^*, v^*). (u^*, v^*) is *neighborhood half-superior* if $p^* = \frac{1}{2}$.[40] (u^*, v^*) is *neighborhood superior* if $p^* = 0$. (u^*, v^*) is *(globally) p^*−superior* if the support of (P, Q) in (10.30) is an arbitrary subset of $S \times T$ (other than $\{(u^*, v^*)\}$).

(b) Strategy pair (u^*, v^*) is a *neighborhood invader strategy (NIS)* if, for all (u, v) sufficiently close (but not equal) to (u^*, v^*), either $\pi_1(u^*; u, v) > \pi_1(u; u, v)$ or $\pi_2(v^*; u, v) > \pi_2(v; u, v)$.

Definition 4 from Cressman (2010, 2011) is the generalization to asymmetric games of Definition 3 in Sect. 3.2. It is also clear that the concept of neighborhood p^*−superior in (10.30) is close to that of two-species ESS given in (10.25). In fact, for asymmetric normal form games (i.e., with S and T finite strategy spaces and payoff linearity), a strategy pair is a two-species ESS if and only if it is neighborhood p^*−superior according to Definition 4 for some $0 \leq p^* < 1$ (c.f. Remark 1 in Sect. 3.2). The following result then generalizes Theorem 2 in Sect. 3.2 (see also Theorem 6 in Sect. 3.4) to asymmetric games (Cressman 2010, 2011).

Theorem 9. *Suppose that (u^*, v^*) is in the interior of $S \times T$.*

(a) (u^*, v^*) *is a neighborhood CSS if and only if it is neighborhood half-superior.*
(b) (u^*, v^*) *is an NIS if and only if, for all nonzero $(u, v) \in \mathbf{R}^2$, either $u((A + 2B)u + 2Cv) < 0$ or $v(2Du + (2E + F)v) < 0$.*
(c) (u^*, v^*) *is a neighborhood strict NE and NIS if and only if it is neighborhood superior.*
(d) *Consider evolution under the replicator equation* (10.31) *that generalizes* (10.15) *to asymmetric games.* (u^*, v^*) *is neighborhood attracting if and only if it is neighborhood superior.*[41]

The replicator equation for an asymmetric game with continuous trait spaces is given by

$$\begin{aligned} \frac{dP_t}{dt}(U) &= \int_U \left(\pi_1(u'; P_t, Q_t) - \pi_1(P_t; P_t, Q_t)\right) P_t(du') \\ \frac{d, Q_t}{dt}(V) &= \int_V \left(\pi_2(v'; P_t, Q_t) - \pi_2(Q_t; P_t, Q_t)\right) Q_t(dv') \end{aligned} \tag{10.31}$$

where U and V are Borel subsets of S and T, respectively.

[40]In (10.30), we assume payoff linearity in the distributions P and Q. For example, the expected payoff to u' in a random interaction is $\pi(u'; P, Q) \equiv \int_S \int_T \pi_1(u'; u, v) Q(dv) P(du)$ where P (Q) is the probability measure on S (T) corresponding to the current distribution of the population one's (two's) strategies. Furthermore, $\pi(P; P, Q) \equiv \int_S \pi(u'; P, Q) P(du')$, etc.

[41]Note that (u^*, v^*) is neighborhood attracting if (P_t, Q_t) converges to $(\delta_{u^*}, \delta_{v^*})$ in the weak topology whenever the support of (P_0, Q_0) is sufficiently close to (u^*, v^*) and $(P_0, Q_0) \in \Delta(S)\times\Delta(T)$ satisfies $P_0(\{u^*\})Q_0(\{v^*\}) > 0$.

Remark 8. The above theory for asymmetric games with one-dimensional continuous trait spaces has been extended to multidimensional trait spaces (Cressman 2009, 2010). Essentially, the results from Sect. 3.4 for symmetric games with multidimensional trait space carry over with the understanding that CSS, NIS, and neighborhood p^*-superiority are now given in terms of Definition 4 and Theorem 9.

Darwinian dynamics for asymmetric games have also been studied (Abrams and Matsuda 1997; Brown and Vincent 1987, 1992; Marrow et al. 1992; Pintor et al. 2011). For instance, in predator-prey systems, the G-function for predators will most likely be different from that of the prey (Brown and Vincent 1987, 1992). Darwinian dynamics, which combines ecological and evolutionary dynamics (c.f. Sect. 3.3), will now model strategy and population size evolution in both species. The advantage to this approach to evolutionary games is that, as in Sect. 3.3, stable evolutionary outcomes can be found that do not correspond to monomorphic populations (Brown and Vincent 1992; Pintor et al. 2011).

5 Conclusion

This chapter has summarized evolutionary game theory for two-player symmetric and asymmetric games based on random pairwise interactions. In particular, it has focused on the connection between static game-theoretic solution concepts (e.g., ESS, CSS, NIS) and stable evolutionary outcomes for deterministic evolutionary game dynamics (e.g., the replicator equation, adaptive dynamics).[42] As we have seen, the unifying principle of local superiority (or neighborhood p^*-superiority) has emerged in the process. These game-theoretic solutions then provide a definition of stability that does not rely on an explicit dynamical model of behavioral evolution. When such a solution corresponds to a stable evolutionary outcome, the detailed analysis of the underlying dynamical system can be ignored. Instead, it is the heuristic static conditions of evolutionary stability that become central to understanding behavioral evolution when complications such as genetic, spatial, and population size effects are added to the evolutionary dynamics.

In fact, stable evolutionary outcomes are of much current interest for other, often non-deterministic, game-dynamic models that incorporate stochastic effects due to finite populations or models with assortative (i.e., nonrandom) interactions (e.g., games on graphs). These additional features, summarized ably by Nowak (2006), are beyond the scope of this chapter. So too are models investigating the evolution of human cooperation whose underlying games are either the two-player Prisoner's Dilemma game or the multiplayer Public Goods game (Binmore 2007). This is

[42] These deterministic dynamics all rely on the assumption that the population size is large enough (sometimes stated as "effectively infinite") so that changes in strategy frequency can be given through the payoff function (i.e., through the strategy's expected payoff in a random interaction).

another area where there is a great deal of interest, both theoretically and through game experiments.

As the evolutionary theory behind these models is a rapidly expanding area of current research, it is impossible to know in what guise the conditions for stable evolutionary outcomes will emerge in future applications. On the other hand, it is certain that Maynard Smith's original idea underlying evolutionary game theory will continue to play a central role.

6 Cross-References

▶ Biology and Evolutionary Games

Acknowledgements The authors thank Abdel Halloway for his assistance with Fig. 10.4. This project has received funding from the European Union's Horizon 2020 research and innovation program under the Marie Skłodowska-Curie grant agreement No 690817.

References

Abrams PA, Matsuda H (1997) Fitness minimization and dynamic instability as a consequence of predator–prey coevolution. Evol Ecol 11:1–10

Abrams PA, Matsuda H, Harada Y (1993) Evolutionary unstable fitness maxima and stable fitness minima of continuous traits. Evol Ecol 7:465–487

Akin E (1982) Exponential families and game dynamics. Can J Math 34:374–405

Apaloo J (1997) Revisiting strategic models of evolution: the concept of neighborhood invader strategies. Theor Popul Biol 52:71–77

Apaloo J (2006) Revisiting matrix games: the concept of neighborhood invader strategies. Theor Popul Biol 69:235–242

Apaloo J, Brown JS, Vincent TL (2009) Evolutionary game theory: ESS, convergence stability, and NIS. Evol Ecol Res 11:489–515

Aubin J-P, Cellina A (1984) Differential inclusions. Springer, Berlin

Barabás G, Meszéna G (2009) When the exception becomes the rule: the disappearance of limiting similarity in the Lotka–Volterra model. J Theor Biol 258:89–94

Barabás G, Pigolotti S, Gyllenberg M, Dieckmann U, Meszéna G (2012) Continuous coexistence or discrete species? A new review of an old question. Evol Ecol Res 14:523–554

Barabás G, D'Andrea R, Ostling A (2013) Species packing in nonsmooth competition models. Theor Ecol 6:1–19

Binmore K (2007) Does game theory work: the bargaining challenge. MIT Press, Cambridge, MA

Bishop DT, Cannings C (1976) Models of animal conflict. Adv Appl Probab 8:616–621

Bomze IM (1991) Cross entropy minimization in uninvadable states of complex populations. J Math Biol 30:73–87

Bomze IM, Pötscher BM (1989) Game theoretical foundations of evolutionary stability. Lecture notes in economics and mathematical systems, vol 324. Springer-Verlag, Berlin

Broom M, Rychtar J (2013) Game-theoretic models in biology. CRC Press, Boca Raton

Brown GW (1951) Iterative solutions of games by fictitious play. In: Koopmans TC (ed) Activity analysis of production and allocation. Wiley, New York, pp 374–376

Brown JS, Pavlovic NB (1992) Evolution in heterogeneous environments: effects of migration on habitat specialization. Theor Popul Biol 31:140–166

Brown JS, Vincent TL (1987) Predator-prey coevolution as an evolutionary game. In: Applications of control theory in ecology. Lecture notes in biomathematics, vol 73. Springer, Berlin, pp 83–101

Brown JS, Vincent TL (1992) Organization of predator-prey communities as an evolutionary game. Evolution 46:1269–1283

Bulmer M (1974) Density-dependent selection and character displacement. Am Nat 108:45–58

Christiansen FB (1991) On conditions for evolutionary stability for a continuously varying character. Am Nat 138:37–50

Cohen Y, Vincent TL, Brown JS (1999) A g-function approach to fitness minima, fitness maxima, evolutionary stable strategies and adaptive landscapes. Evol Ecol Res 1:923–942

Courteau J, Lessard S (2000) Optimal sex ratios in structured populations. J Theor Biol 207:159–175

Cressman R (1992) The stability concept of evolutionary games (a dynamic approach). Lecture notes in biomathematics, vol 94. Springer-Verlag, Berlin

Cressman R (2003) Evolutionary dynamics and extensive form games. MIT Press, Cambridge, MA

Cressman R (2009) Continuously stable strategies, neighborhood superiority and two-player games with continuous strategy spaces. Int J Game Theory 38:221–247

Cressman R (2010) CSS, NIS and dynamic stability for two-species behavioral models with continuous trait spaces. J Theor Biol 262:80–89

Cressman R (2011) Beyond the symmetric normal form: extensive form games, asymmetric games and games with continuous strategy spaces. In: Sigmund K (ed) Evolutionary game dynamics. Proceedings of symposia in applied mathematics, vol 69. American Mathematical Society, Providence, pp 27–59

Cressman R, Hines WGS (1984) Correction to the appendix of 'three characterizations of population strategy stability'. J Appl Probab 21:213–214

Cressman R, Hofbauer J (2005) Measure dynamics on a one-dimensional continuous strategy space: theoretical foundations for adaptive dynamics. Theor Popul Biol 67:47–59

Cressman R, Krivan V (2006) Migration dynamics for the ideal free distribution. Am Nat 168:384–397

Cressman R, Krivan V (2013) Two-patch population podels with adaptive dispersal: the effects of varying dispersal speed. J Math Biol 67:329–358

Cressman R, Tao Y (2014) The replicator equation and other game dynamics. Proc Natl Acad Sci USA 111:10781–10784

Cressman R, Tran T (2015) The ideal free distribution and evolutionary stability in habitat selection games with linear fitness and Allee effect. In: Cojocaru MG et al (eds) Interdisciplinary topics in applied mathematics, modeling and computational sciences. Springer proceedings in mathematics and statistics, vol 117. Springer, New York, pp 457–464

Cressman R, Garay J, Hofbauer J (2001) Evolutionary stability concepts for N-species frequency-dependent interactions. J Theor Biol 211:1–10

Cressman R, Halloway A, McNickle GG, Apaloo J, Brown J, Vincent TL (2016, preprint) Infinite niche packing in a Lotka-Volterra competition game

Cressman R, Krivan V, Garay J (2004) Ideal free distributions, evolutionary games, and population dynamics in mutiple-species environments. Am Nat 164:473–489

Cressman R, Hofbauer J, Riedel F (2006) Stability of the replicator equation for a single-species with a multi-dimensional continuous trait space. J Theor Biol 239:273–288

D'Andrea R, Barabás G, Ostling A (2013) Revising the tolerance-fecundity trade-off; or, on the consequences of discontinuous resource use for limiting similarity, species diversity, and trait dispersion. Am Nat 181:E91–E101

Darwin C (1871) The descent of man and selection in relation to sex. John Murray, London

Dawkins R (1976) The selfish gene. Oxford University Press, Oxford

Dercole F, Rinaldi S (2008) Analysis of evolutionary processes. The adaptive dynamics approach and its applications. Princeton University Press, Princeton

Dieckmann U, Law R (1996) The dynamical theory of coevolution: a derivation from stochastic ecological processes. J Math Biol 34:579–612

Doebeli M, Dieckmann U (2000) Evolutionary branching and sympatric speciation caused by different types of ecological interactions. Am Nat 156:S77–S101
Eshel I (1983) Evolutionary and continuous stability. J Theor Biol 103:99–111
Eshel I, Motro U (1981) Kin selection and strong evolutionary stability of mutual help. Theor Popul Biol 19:420–433
Fisher RA (1930) The genetical theory of natural selection. Clarendon Press, Oxford
Fretwell DS, Lucas HL (1969) On territorial behavior and other factors influencing habitat distribution in birds. Acta Biotheor 19:16–32
Fudenberg D, Levine DK (1998) The theory of learning in games. MIT Press, Cambridge, MA
Geritz SAH, Kisdi É, Meszéna G, Metz JAJ (1998) Evolutionarily singular strategies and the adaptive growth and branching of the evolutionary tree. Evol Ecol 12:35–57
Gilboa I, Matsui A (1991) Social stability and equilibrium. Econometrica 59:185–198
Gintis H (2000) Game theory evolving. Princeton University Press, Princeton
Gyllenberg M, Meszéna G (2005) On the impossibility of coexistence of infinitely many strategies. J Math Biol 50:133–160
Haigh J (1975) Game theory and evolution. Adv Appl Probab 7:8–11
Hamilton WD (1964) The genetical evolution of social behavior I and II. J Theor Biol 7:1–52
Hamilton WD (1967) Extraordinary sex ratios. Science 156:477–488
Hines WGS (1980) Three characterizations of population strategy stability. J Appl Probab 17: 333–340
Hines WGS (1987) Evolutionarily stable strategies: a review of basic theory. Theor Popul Biol 31:195–272
Hofbauer J (1995, preprint) Stability for the best response dynamics
Hofbauer J, Sigmund K (1988) The theory of evolution and dynamical systems. Cambridge University Press, Cambridge
Hofbauer J, Sigmund K (1990) Adaptive dynamics and evolutionary stability. Appl Math Lett 3:75–79
Hofbauer J, Sigmund K (1998) Evolutionary games and population dynamics. Cambridge University Press, Cambridge
Hofbauer J, Sigmund K (2003) Evolutionary game dynamics. Bull Am Math Soc 40:479–519
Hofbauer J, Schuster P, Sigmund K (1979) A note on evolutionarily stable strategies and game dynamics. J Theor Biol 81:609–612
Kisdi É, Meszéna G (1995) Life histories with lottery competition in a stochastic environment: ESSs which do not prevail. Theor Popul Biol 47:191–211
Krivan V (2014) The Allee-type ideal free distribution. J Math Biol 69:1497–1513
Krivan V, Cressman R, Schneider C (2008) The ideal free distribution: a review and synthesis of the game theoretic perspective. Theor Popul Biol 73:403–425
Kuhn H (1953) Extensive games and the problem of information. In: Kuhn H, Tucker A (eds) Contributions to the thoery of games II. Annals of mathematics, vol 28. Princeton University Press, Princeton, pp 193–216
Leimar O (2001) Evolutionary change and Darwinian demons. Selection 2:65–72
Leimar O (2009) Multidimensional convergence stability. Evol Ecol Res 11:191–208
Lessard S (1990) Evolutionary stability: one concept, several meanings. Theor Popul Biol 37: 159–170
Marrow P, Law R, Cannings C (1992) The coevolution of predator–prey interactions: ESSs and Red Queen dynamics. Proc R Soc Lond B 250:133–141
Matsui A (1992) Best response dynamics and socially stable strategies. J Econ Theory 57:343–362
Maynard Smith J (1974) The theory of games and the evolution of animal conflicts. J Theor Biol 47:209–221
Maynard Smith J (1982) Evolution and the theory of games. Cambridge University Press, Cambridge
Maynard Smith J, Price G(1973) The logic of animal conflicts. Nature 246:15–18
McKelvey R, Apaloo J (1995) The structure and evolution of competition-organized ecological communities. Rocky Mt J Math 25:417–436

Mesterton-Gibbons M (2000) An introduction to game-theoretic modelling, 2nd edn. American Mathematical Society, Providence

Meszéna G, Kisdi E, Dieckmann U, Geritz SAH, Metz JAJ (2001) Evolutionary optimisation models and matrix games in the unified perspective of adaptive dynamics. Selection 2:193–210

Meszéna G, Gyllenberg M, Pásztor L, Metz JAJ (2006) Competitive exclusion and limiting similarity: a unified theory. Theor Popul Biol 69:69–87

Morris D (2002) Measure the Allee effect: positive density effect in small mammals. Ecology 83:14–20

Nagylaki T (1992) Introduction to theoretical population genetics. Biomathematics, vol 21. Springer-Verlag, Berlin

Nash J (1950) Equilibrium points in n-person games. Proc Natl Acad Sci USA 36:48–49

Nash J (1951) Non-cooperative games. Ann Math 54:286–295

Nowak MA (2006) Evolutionary dynamics. Harvard University Press, Cambridge, MA

Oechssler J, Riedel F (2001) Evolutionary dynamics on infinite strategy spaces. Econ Theory 17:141–162

Parvinen K, Meszéna G (2009) Disturbance-generated niche-segregation in a structured metapopulation model. Evol Ecol Res 11:651–666

Pintor LM, Brown JS, Vincent TL (2011) Evolutionary game theory as a framework for studying biological invasions. Am Nat 177:410–423

Ripa J, Storling L, Lundberg P, Brown JS (2009) Niche co-evolution in consumer-resource communities. Evol Ecol Res 11:305–323

Robinson J (1951) An iterative method of solving a game. Ann Math 54:296–301, 333–340

Roughgarden J (1979) Theory of population genetics and evolutionary ecology: an introduction. Macmillan, New York

Samuelson L (1997) Evolutionary games and equilibrium selection. MIT Press, Cambridge, MA

Samuelson L, Zhang J (1992) Evolutionary stability in asymmetric games. J Econ Theory 57: 363–391

Sandholm WH (2010) Population games and evolutionary dynamics. MIT Press, Cambridge, MA

Sasaki A (1997) Clumped distribution by neighborhood competition. J Theor Biol 186: 415–430

Sasaki A, Ellner S (1995) The evolutionarily stable phenotype distribution in a random environment. Evolution 49:337–350

Schlag KH (1997) Why imitate, and if so, how? A boundedly rational approach to multi-armed bandits. J Econ Theory 78:133–156

Schuster P, Sigmund K (1983) Replicator dynamics. J Theor Biol 100:533–538

Selten R (1978) The chain-store paradox. Theory Decis 9:127–159

Selten R (1980) A note on evolutionarily stable strategies in asymmetrical animal contests. J Theor Biol 84:93–101

Selten R (1983) Evolutionary stability in extensive two-person games. Math Soc Sci 5:269–363

Selten R (1988) Evolutionary stability in extensive two-person games: corrections and further development. Math Soc Sci 16:223–266

Sigmund K (1993) Games of life. Oxford University Press, Oxford

Sigmund K (ed) (2011) Evolutionary game dynamics. Proceedings of symposia in applied mathematics, vol 69. American Mathematical Society, Providence

Szabó P, Meszéna G (2006) Limiting similarity revisited. Oikos 112:612–619

Taylor PD (1989) Evolutionary stability in one-parameter models under weak selection. Theor Popul Biol 36:125–143

Taylor PD, Jonker L (1978) Evolutionarily stable strategies and game dynamics. Math Biosci 40:145–156

van Damme E (1991) Stability and perfection of Nash equilibria, 2nd edn. Springer-Verlag, Berlin

Vega-Redondo F (1996) Evolution, games, and economic behavior. Oxford University Press, Oxford

Vincent TL, Brown JS (2005) Evolutionary game theory, natural selection, and Darwinian dynamics. Cambridge University Press, Cambridge

Vincent TL, Cohen Y, Brown JS (1993) Evolution via strategy dynamics. Theor Popul Biol 44: 149–176
von Neumann J, Morgenstern O (1944) Theory of games and economic behavior. Princeton University Press, Princeton
Weibull J (1995) Evolutionary game theory. MIT Press, Cambridge, MA
Xu Z (2016) Convergence of best-response dynamics in extensive-form games. J Econ Theory 162:21–54
Young P (1998) Individual strategy and social structure. Princeton University Press, Princeton

Game-Theoretic Learning in Distributed Control

11

Jason R. Marden and Jeff S. Shamma

Contents

This work was supported by ONR Grant #N00014-17-1-2060 and NSF Grant #ECCS-1638214 and by funding from King Abdullah University of Science and Technology (KAUST).

J. R. Marden (✉)
Department of Electrical and Computer Engineering, University of California, Santa Barbara, CA, USA
e-mail: jrmarden@ece.ucsb.edu

J. S. Shamma
Computer, Electrical and Mathematical Science and Engineering Division (CEMSE), King Abdullah University of Science and Technology (KAUST), Thuwal, Saudi Arabia
e-mail: jeff.shamma@kaust.edu.sa

T. Başar, G. Zaccour (eds.), *Handbook of Dynamic Game Theory*,
https://doi.org/10.1007/978-3-319-44374-4_9

Abstract

In distributed architecture control problems, there is a collection of interconnected decision-making components that seek to realize desirable collective behaviors through local interactions and by processing local information. Applications range from autonomous vehicles to energy to transportation. One approach to control of such distributed architectures is to view the components as players in a game. In this approach, two design considerations are the components' incentives and the rules that dictate how components react to the decisions of other components. In game-theoretic language, the incentives are defined through utility functions, and the reaction rules are online learning dynamics. This chapter presents an overview of this approach, covering basic concepts in game theory, special game classes, measures of distributed efficiency, utility design, and online learning rules, all with the interpretation of using game theory as a prescriptive paradigm for distributed control design.

Keywords

Learning in games · Evolutionary games · Multiagent systems · Distributed decision systems

1 Introduction

There is growing interest in distributed architecture or networked control systems, with emergent applications ranging from smart grid to autonomous vehicle networks to mobile sensor platforms. As opposed to a traditional control system architecture, there is no single decision-making entity with full information and full authority that acts as an overall system controller. Rather, decisions are made by a collective of interacting entities with local information and limited communication capabilities. The challenge is to derive distributed controllers to induce desirable collective behaviors.

One approach to distributed architecture systems is to view the decision-making components as individual players in a game and to formulate the distributed control problem in terms of game theory. The basic elements of what constitutes a game are (i) a set of players or agents; (ii) for each player, a set of choices; and (iii) for each player, preferences over the *collective* choices of agents, typically expressed in the form of a utility function. In traditional game theory (e.g., Fudenberg and Tirole 1991), these elements are a *model* of a collection of decision-makers, typically in a societal context (e.g., competing firms, voters, bidders in an auction, etc.). In the context of distributed control, these elements are *design considerations* in that one has the degree of freedom on how to decompose a distributed control problem and how to design the preferences/utility functions to properly incentivize agents. Stated differently, game theory in this context is being used as a *prescriptive* paradigm, rather than a *descriptive* paradigm (Marden and Shamma 2015; Shoham et al. 2007).

Formulating a distributed control problem in game-theoretic terms implicitly suggests that the outcome – or, more appropriately, the solution concept – of the resulting game is a desirable collective configuration. The most well-known solution concept is Nash equilibrium, in which each player's choice is optimal with respect to the choices of other agents. Other solution concepts, which are generalizations of Nash equilibrium, are correlated and coarse correlated equilibrium (Young 2004). Typically, a solution concept does not uniquely specify the outcome of a game (e.g., a game can have multiple Nash equilibria), and so there is the issue that some outcomes are better than others.

A remaining concern is how a solution concept emerges at all. Given the complete description of a game, an outside party can proceed to compute (modulo computational complexity considerations Daskalakis et al. 2009) a proposed solution concept realization. In actuality, the data of a game (e.g., specific utility functions) is distributed among the players and not necessarily shared or communicated. Rather, over time players might make observations of the choices of the other players and eventually the collective play converges to some limiting structure. This latter scenario is the topic of game-theoretic learning, for which there are multiple survey articles and monographs (e.g., Fudenberg and Levine 1998; Hart 2005; Shamma 2014; Young 2004). Under the descriptive paradigm, the learning in games discussion provides a plausibility argument of how players may arrive at a specified solution concept realization. Under the prescriptive paradigm, the learning in games discussion suggests an *online algorithm* that can lead agents to a desirable solution concept realization.

This article will provide an overview of approaching distributed control from the perspective of game theory. The presentation will touch on each of the aforementioned aspects of problem formulation, game design, and game-theoretic learning.

2 Game-Theoretic Distributed Resource Utilization

2.1 Setup

Various problems of interest take the form of allocating a collection of assets to utilize a set of resources to a desired effect. In sensor coverage problems (e.g., Cortes et al. 2002), the "assets" are mobile sensors, and the "resources" are the regions to be covered by the sensors. For any given allocation, there is an overall score reflecting the quality of the coverage. In traffic routing problems (e.g., Roughgarden 2005), the "assets" are vehicles (or packets in a communication setting), and the "resources" are roads (or channels). The objective is to route traffic from origins to destinations in order to minimize a global cost such as congestion.

It will be instructive to interpret the forthcoming discussion on distributed control in the framework of such distributed resource utilization problems. As previously mentioned, the framework captures a variety of applications of interest. Furthermore, focusing on this specific setting will enhance the clarity of exposition.

More formally, the problem is to allocate a collection of assets $N = \{1, 2, \ldots, n\}$ over a collection of resources $\mathcal{R} = \{1, 2, \ldots, m\}$ in order to optimize a given system-level objective. The set $\mathcal{A}_i \subseteq 2^{\mathcal{R}}$ is the allowable resource selections by asset i. In terms of the previous examples, an allowable resource selection is an area covered by a sensor or a set of roads used by vehicle. The system-level objective is a mapping $W : \mathcal{A} \to \mathbb{R}$ where $\mathcal{A} = \mathcal{A}_1 \times \cdots \times \mathcal{A}_n$ denotes the set of joint resource selections. We denote a collective configuration by the tuple $a = (a_1, a_2, \ldots, a_n)$ where $a_i \in \mathcal{A}_i$ is the choice, or *action*, of agent i.

Moving toward a game-theoretic model, we will identify the set of assets as the set of agents or players. Likewise, we will identify $\mathcal{A}_i$ as the choice set of agent i. We defer for now specifying a utility function for agent i.

Looking forward to the application of game-theoretic learning, we will consider agents selecting actions iteratively over an infinite time horizon $t \in \{1, 2, \ldots\}$. Depending on the update rules of the agents, the outcome is a sequence of joint actions $a(1), a(2), a(3), \ldots$. The action of agent i at time t is chosen according to some update policy, $\pi_i(\cdot)$, i.e.,

$$a_i(t) = \pi_i\,(\text{information available to agent } i \text{ at time } t)\,. \tag{11.1}$$

The update policy $\pi_i(\cdot)$ specifies how agent i processes available information to formulate a decision. We will be more explicit about the argument of the $\pi_i(\cdot)$'s in the forthcoming discussion. For now, the information available to an agent can include both knowledge regarding previous action choices of other agents and certain system-level information that is propagated throughout the system.

The main goal is to design both the agents' utility functions and the agents' local policies $\{\pi_i\}_{i \in N}$ to ensure that the emergent collective behavior optimizes the global objective W in terms of the asymptotic properties of $W(a(t))$ as $t \to \infty$.

2.2 Prescriptive Paradigm

Once the players and their choices have been set, the remaining elements in the prescriptive paradigm that are yet to be designed are (i) the agent utility functions and (ii) the update policies, $\{\pi_i\}_{i \in N}$. One can view this specification in terms of the following two-step design procedure:

Step #1: Game Design. The first step of the design involves defining the underlying interaction structure in a game-theoretic environment. In particular, this choice involves defining a utility function for each agent $i \in N$ of the form $U_i : \mathcal{A} \to \mathbb{R}$. The utility of agent i for an action profile $a = (a_1, a_2, \ldots, a_n)$ is expressed as $U_i(a)$ or alternatively $U_i(a_i, a_{-i})$ where a_{-i} denotes the collection of actions other than player i in the joint action a, i.e., $a_{-i} = (a_1, \ldots, a_{i-1}, a_{i+1}, \ldots, a_n)$. A key feature of this design choice is the coupling of the agents' utility functions where the utility, or payoff, of one agent is affected by the actions of other agents.

Step #2: Learning Design. The second step involves defining the decision-making rules for the agents. That is, how does each agent process available information to formulate a decision. A typical assumption in the framework of learning in games is that each agent uses historical information from previous actions of itself and other players. Accordingly, at each time t the decision of each agent $i \in N$ is made independently through a learning rule of the form

$$a_i(t) = \pi_i \left(\{a(\tau)\}_{\tau=1,\ldots,t-1} \, ; U_i(\cdot) \right). \tag{11.2}$$

There are two important considerations in the above formulation. First, we stated for simplicity that agents can observe the actions of all other agents. In games with a graphical structure (Kearns et al. 2001), one only requires historical information from a *subset* of other players. Other reductions are also possible, such as aggregate information of other players of even just measurements of one's own utility (Fudenberg and Levine 1998; Hart 2005; Shamma 2014; Young 2004).[1] Second, implicit in the above construction is that the learning rule is defined independently of the utility function, and an agent's utility function then enters as a parameter of a specified learning rule.

This second consideration offers a distinction between conventional distributed control and game-theoretic distributed control in the role of the utility function for the individual agents $\{U_i\}_{i\in N}$. An agent may be using a specific learning rule, but the realized behavior depends on the specified utility function. In a more conventional approach, there need not be such a decomposition. Rather, one might directly specify the agents' control policies $\{\pi_i\}_{i\in N}$ and perform an analysis regarding the emergent properties of the given design, e.g., as is done in models of flocking or bio-inspired controls (Olfati-Saber 2006). An advantage of the decomposition is that one can analyze learning rules for classes of games and separately examine whether or not specified utility functions conform to such an assumed game class.

The following example demonstrates how a given distributed control policy $\{\pi\}_{i\in N}$ can be reinterpreted as a game-theoretic control approach with appropriately defined agent utility functions $\{U_i\}_{i\in N}$.

Example 1 (Consensus). Consider the well-studied consensus/rendezvous problem (Blondel et al. 2005b; Jadbabaie et al. 2003; Olfati-Saber and Murray 2003; Touri and Nedic 2011) where the goal is to drive the agents to agreement on a state $x^* \in \mathbb{R}$ when each agent has limited information regarding the state of other agents in the

[1]Alternative agent control policies where the policy of agent i also depends on previous actions of agent i or auxiliary "side information" could also be replicated by introducing an underlying state in the game-theoretic environment. The framework of state-based games, introduced in Marden (2012), represents one such framework that could accomplish this goal.

systems. Specifically, we will say that the set of admissible states (or actions) of each agent $i \in N$ is $\mathcal{A}_i = \mathbb{R}$ and agent i at stage t can observe the previous state choices at stage $t-1$ of a set of neighboring agents denoted by $\mathcal{N}_i(t) \subseteq N \setminus \{i\}$. Consider the following localized averaging dynamics where the decision of an agent $i \in N$ at time t is of the form

$$a_i(t) = \frac{1}{\mathcal{N}_i(t)} \sum_{j \in \mathcal{N}_i(t)} a_j(t-1). \tag{11.3}$$

Given an initial state profile $a(0)$, the dynamics in (11.3) produces a sequence of state profiles $a(1), a(2), \ldots$. Whether or not the state profiles converge to consensus under the above dynamics (or variants thereof) has been extensively studied in the existing literature (Blondel et al. 2005a; Olfati-Saber and Murray 2003; Tsitsiklis et al. 1986).

Now we will present a game-theoretic design that leads to the same collective behavior. More formally, consider a game-theoretic model where each agent $i \in N$ is assigned an action set $\mathcal{A}_i = \mathbb{R}$ and a utility function of the form

$$U_i(a_i, a_{-i}) = -\frac{1}{2\,|\mathcal{N}_i(t)|} \sum_{j \in \mathcal{N}_i(t)} (a_i - a_j)^2, \tag{11.4}$$

where $|\mathcal{N}_i(t)|$ denotes the cardinality of the set $\mathcal{N}_i(t)$. Now, suppose each agent follows the well-known best-response learning rule of the form

$$a_i(t) \in B_i(a_{-i}(t)) = \arg\max_{a_i \in \mathcal{A}_i} U_i(a_i, a_{-i}(t-1)),$$

where $B_i(a_{-i}(t))$ is referred to as the best-response set of agent i to the action profile $a_{-i}(t)$. Given an initial state profile $a(0)$, it is straightforward to show that the ensuing action or state profiles $a(1), a(2), \ldots,$ will be equivalent for both design choices.

The above example illustrates the separation between the learning rule and the utility function. The learning rule is best-response dynamics. When the utility function is the above quadratic form, then the combination leads to the usual distributed averaging algorithm. If the utility function is changed (e.g., weighted, non-quadratic, etc.), then the realization of best-response learning is altered, as well as the structure of the game defined by the collection of the utility functions, but the learning rule remains best-response dynamics.

An important property of best-response dynamics and other learning rules of interest is that the actions of agent i can depend explicitly on the utility function of agent i but not (explicitly) on the utility functions of other agents. This property of learning rules in the learning in games literature is called being *uncoupled* (Babichenko 2012; Hart and Mansour 2010; Hart and Mas-Colell 2003; Young

2004). Of course, the action stream of agent i, i.e., $a_i(0), a_i(1), \ldots$, does depend on the *actions* of other agents, but not the utility functions behind those actions.

It turns out that there are many instances in which control policies not derived from a game-theoretic perspective can be reinterpreted as the realization of an uncoupled learning rule from a game-theoretic perspective. These include control policies that have been widely studied in the cooperative control literature with application domains such as consensus and flocking (Olfati-Saber et al. 2007; Tsitsiklis 1987), sensor coverage (Martinez et al. 2007; Murphey 1999), and routing information over networks (Roughgarden 2005), among many others.

While the design of such control policies can be approached in either a traditional perspective or a game-theoretic perspective, there are potential advantages associated with viewing control design from a game-theoretic perspective. In particular, a game-theoretic perspective allows for a modularized design architecture, i.e., the separation of game design and learning design, that can be exploited in a plug-and-play fashion to provide control algorithms with automatic performance guarantees:

Game Design Methodologies. There are several established methodologies for the design of agent objective functions, e.g., Shapley value and marginal contribution (Marden and Wierman 2013). The methodologies, which will be briefly reviewed in Sect. 3.6, are systematic procedures for deriving the agent objective functions $\{U_i\}_{i \in N}$ from a given system-level objective function G. These methodologies often provide structural guarantees on the resulting game, e.g., existence of a pure Nash equilibrium or a potential game structure, that can be exploited in distributed learning.

Learning Design Methodologies. The field of learning in games has sought out to establish decision-making rules that lead to Nash equilibrium or other solution concepts in strategic form games. In general, it has been shown (see Hart and Mas-Colell 2003) that there are no "natural" dynamics that converge to Nash equilibria for all games, where natural refers to dynamics that do not rely on some form of centralized coordination, e.g., exhaustive search of the joint action profiles. For example, there are no rules of the form (11.2) that provide convergence to a Nash equilibrium in any game. However, the same limitations do not hold when we transition from "all games" to "all games of a given structure." In particular, there are several positive results in the context of learning in games for special classes of games (e.g., potential games and variants thereof). These results, which will be discussed in Sect. 4, identify learning dynamics that yield desirable performance guarantees when applied to the realm of potential games.

Performance Guarantees. Merging a game design methodology with an appropriate learning design methodology can often result in agent control policies with automatic performance guarantees. For example, employing a game design where agent utility functions constitute a potential game coupled with a learning algorithm that ensures convergence to a pure Nash equilibrium in potential games provides agent control policies that converge to the Nash equilibrium of the derived game. Furthermore, additional structure on the agents' utility functions can often be exploited to provide efficiency bounds on the Nash equilibria, cf., price of anarchy (Nisan et al. 2007), as well as approximations for the underlying convergence rates (Borowski et al. 2013; Montanari and Saberi 2009; Shah and Shin 2010).

Human-Agent Collaborative Systems. Game theory constitutes a design choice for control policies in distributed systems comprised purely of engineering components. However, when a networked system consists of both engineering and human decision-making entities, e.g., the smart grid, game theory transitions from a design choice to a necessity. The involvement of human decision-making entities in a system requires that the system operator utilizes game theory for the purpose of modeling and influencing the human decision-making entities to optimize system performance.

3 Solution Concepts, Game Structures, and Efficiency

Recall that an important metric in the game-theoretic approach to distributed control is the asymptotic properties of a system-level objective function, i.e., $W(a(t))$ as $t \to \infty$. These asymptotic properties depend on both aspects of the prescriptive paradigm, i.e., the utility functions and learning rule. The specification of utility functions in itself defines an underlying game that is repeatedly played over stages. In this section, we review properties related to this underlying game in terms of solution concepts, game structures, and measures of efficiency.

In this section we will temporarily distance ourselves from the design objectives set forth in this manuscript with the purpose of identifying properties of games that are relevant to our mission. To that end, we will consider a finite strategic form game G with agent set $N = \{1, 2, \ldots, n\}$ where each agent $i \in N$ has an action set $\mathcal{A}_i$ and a utility function $U_i : \mathcal{A} \to \mathbb{R}$. Further, there exists a system-level objective $W : \mathcal{A} \to \mathbb{R}$ that a system designer is interested in maximizing. We will often denote such a game by the tuple $G = \{N, \{\mathcal{A}_i\}, \{U_i\}, W\}$ where we use the shorthand notation $\{\cdot\}$ instead of $\{\cdot\}_{i \in N}$ to denote the agents' action sets or utility functions.

3.1 Solution Concepts

The most widely known solution concept in game theory is a pure Nash equilibrium, defined as follows.

Definition 1. An action profile $a^{\text{ne}} \in \mathcal{A}$ is a pure Nash equilibrium if for any agent $i \in N$

$$U_i(a_i^{\text{ne}}, a_{-i}^{\text{ne}}) \geq U_i(a_i, a_{-i}^{\text{ne}}), \forall a_i \in \mathcal{A}_i. \tag{11.5}$$

A pure Nash equilibrium represents an action profile where no agent has a unilateral incentive to alter its action provided that the behavior of the remaining agents is unchanged. A pure Nash equilibrium need not exist for any game G.

The definition of Nash equilibrium also extends to scenarios where the agents can probabilistically choose their actions. Define a strategy of agent i as $p_i \in \Delta(\mathcal{A}_i)$ where $\Delta(\mathcal{A}_i)$ denotes the simplex over the finite action set $\mathcal{A}_i$. We will express a strategy p_i by the tuple $\{p_i^{a_i}\}_{a_i \in \mathcal{A}_i}$ where $p_i^{a_i} \geq 0$ for any $a_i \in \mathcal{A}_i$ and $\sum_{a_i \in \mathcal{A}_i} p_i^{a_i} = 1$. We will evaluate the utility of an agent $i \in N$ for a strategy profile $p = (p_1, \ldots, p_n)$ as

$$U_i(p_i, p_{-i}) = \sum_{a \in \mathcal{A}} U_i(a) \times p_1^{a_1} \times \cdots \times p_n^{a_n}. \tag{11.6}$$

which has the usual interpretation of the expected utility under independent randomized actions.

We can now state the definition of Nash equilibrium when extended to mixed (or probabilistic) strategies.

Definition 2. A strategy profile $p^{\text{ne}} \in \Delta(\mathcal{A}_1) \times \cdots \times \Delta(\mathcal{A}_n)$ is a mixed Nash equilibrium if for any agent $i \in N$

$$U_i(p_i^{\text{ne}}, p_{-i}^{\text{ne}}) \geq U_i(p_i, p_{-i}^{\text{ne}}), \ \forall p_i \in \Delta(\mathcal{A}_i). \tag{11.7}$$

Unlike pure Nash equilibria, a mixed Nash equilibrium is guaranteed to exist in any[2] game G.

A common critique regarding the viability of pure or mixed Nash equilibria as a characterization of achievable behavior in multiagent systems is that the complexity associated with computing such equilibria is often prohibitive (Daskalakis et al. 2009). We now introduce a weaker solution concept, which is defined relative to a joint distribution $z \in \Delta(\mathcal{A})$, that does not suffer from such issues.

Definition 3. A joint distribution $z = \{z^a\}_{a \in \mathcal{A}} \in \Delta(\mathcal{A})$ is a coarse correlated equilibrium if for any agent $i \in N$

$$\sum_{a \in \mathcal{A}} U_i(a_i, a_{-i}) z^{(a_i, a_{-i})} \geq \sum_{a \in \mathcal{A}} U_i(a_i', a_{-i}) z^{(a_i, a_{-i})}, \ \forall a_i' \in \mathcal{A}_i. \tag{11.8}$$

A coarse correlated equilibrium is a joint distribution z such that each agent's expected utility according to that distribution is at least as high as the agent's expected utility for committing to any fixed action $a_i' \in \mathcal{A}$, while all the other agents play according to their marginal distribution of z. It is straightforward to verify that

[2]Recall that we are assuming a finite set of players, each with a finite set of actions.

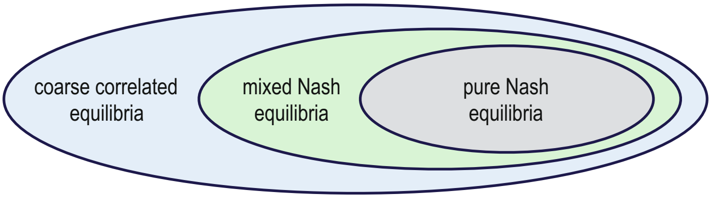


Fig. 11.1 The relationship between the three solution concepts: pure Nash equilibrium, mixed Nash equilibrium, and coarse correlated equilibrium

any mixed Nash equilibrium is a coarse correlated equilibrium; hence, the set of coarse correlated equilibria is nonempty for any game, G. Furthermore, as we will see in Sect. 4.4, there are simple learning algorithms that ensure that the empirical frequency of play will approach the set of coarse correlated equilibria in a reasonable period of time. We will discuss techniques for characterizing the efficiency of this type of collective behavior in Sect. 3.3.[3]

Figure 11.1 highlights the relationship between the three solution concepts discussed above.

3.2 Measures of Efficiency

It is important to highlight that the above equilibrium definitions have no dependence on the system-level objective function. The goal here is to understand how the efficiency associated with such equilibria compares to the optimal behavior with respect to a system-level objective function. Here, we investigate two common worst-case measures, termed *price of anarchy* and *price of stability* (Nisan et al. 2007), for characterizing the inefficiency associated with equilibria in games.

The first measure that we consider is the price of anarchy, which is defined as the worst-case ratio between the performance of the worst equilibrium and the optimal system behavior. We use the terminology worst equilibrium as the price of anarchy could be defined by restricting attention to any of the aforementioned equilibrium sets. Focusing on pure Nash equilibria for simplicity, the price of anarchy associated with a game G is defined as

$$\mathrm{PoA}(G) = \min_{a^{\mathrm{ne}} \in \mathrm{PNE}(G)} \left\{ \frac{W(a^{\mathrm{ne}})}{W(a^{\mathrm{opt}})} \right\} \leq 1, \tag{11.9}$$

[3]Another common equilibrium set, termed correlated equilibrium, is similar to coarse correlated equilibrium where the difference lies in the consideration of conditional deviations as opposed to the unconditional deviations considered in (11.8). A formal definition of correlated equilibrium can be found in Young (2004).

where $a^{\text{opt}} \in \arg\max_{a \in \mathcal{A}} W(a)$ and $\text{PNE}(G)$ denotes the set of pure Nash equilibria in the game G. Note that the price of anarchy given in (11.9) provides a lower bound on the performance associated with any pure Nash equilibrium in the game G.

The second measure that we consider is the price of stability, which is defined as the best-case ratio between the performance of the *best* equilibrium and the optimal system behavior. Focusing on pure Nash equilibria for simplicity, the price of stability associated with a game G is defined as

$$\text{PoS}(G) = \max_{a^{\text{ne}} \in \text{PNE}(G)} \left\{ \frac{W(a^{\text{ne}})}{W(a^{\text{opt}})} \right\} \leq 1. \tag{11.10}$$

By definition, $\text{PoS}(G) \geq \text{PoA}(G)$. The price of stability is a more optimistic measure of the efficiency loss associated with pure Nash equilibrium. When analyzing dynamics that converge to specific types of equilibrium, e.g., the best Nash equilibrium, the price of stability may be a more reasonable characterization of the efficiency associated with the limiting behavior.

The above definition of price of anarchy and price of stability also extend to situations where there is uncertainty regarding the structure of the specific game. To that end, let $\mathcal{G}$ denote a family of possible games. The price of anarchy and price of stability associated with the family of games is then defined as the worst-case performance of all games within that family, i.e.,

$$\text{PoA}(\mathcal{G}) = \min_{G \in \mathcal{G}} \left\{ \text{PoA}(G) \right\}, \tag{11.11}$$

$$\text{PoS}(\mathcal{G}) = \min_{G \in \mathcal{G}} \left\{ \text{PoS}(G) \right\}. \tag{11.12}$$

Clearly, $1 \geq \text{PoS}(\mathcal{G}) \geq \text{PoA}(\mathcal{G})$. For clarity, a $\text{PoA}(\mathcal{G}) = 0.5$ implies that regardless of the underlying game $G \in \mathcal{G}$, any pure Nash equilibrium is at least 50% efficient when compared to the performance of the optimal allocation for that game.

The definitions of price of anarchy and price of stability given in (11.9) and (11.10) can be extended to broader classes of equilibria, i.e., mixed Nash equilibria or coarse correlated equilibria, in the logical manner. To perform the above analysis for broader equilibrium sets, we extend the definition of the welfare function to a distribution $z \in \Delta(\mathcal{A})$ as $W(z) = \sum_{a \in \mathcal{A}} W(a) z^a$. Note that for a given family of games $\mathcal{G}$, the price of anarchy associated with pure Nash equilibria would be better (closer to 1) than the price of anarchy associated with coarse correlated equilibrium. Since coarse correlated equilibria contain Nash equilibria, one would naturally expect that the efficiency associated with equilibria could be far worse than the efficiency associated with Nash equilibria. Surprisingly, it often turns out that this is not the case as we will see below.

3.3 Smoothness

Charactering the inefficiency of equilibria often is challenging and often involves a nontrivial domain specific analysis. One attempt at providing a universal approach to characterizing efficiency loss in distributed systems, termed smoothness (Roughgarden 2015), is given in the following theorem.

Theorem 1. *Consider any game G where the agents' utility functions satisfy $\sum_{i\in N} U_i(a) \leq W(a)$ for any $a \in \mathcal{A}$. If there exists parameters $\lambda > 0$ and $\mu > -1$ such that for any two action profiles $a, a^* \in \mathcal{A}$*

$$\sum_i U_i(a_i^*, a_{-i}) \geq \lambda \cdot W(a^*) - \mu \cdot W(a), \tag{11.13}$$

then the efficiency associated with any coarse correlated equilibrium $z^{\text{cce}} \in \Delta(\mathcal{A})$ of G must satisfy

$$\frac{W(z^{\text{cce}})}{W(a^{\text{opt}})} \geq \frac{\lambda}{1+\mu}. \tag{11.14}$$

We will refer to a game G as $(\lambda, \mu)-$smooth if the game satisfies (11.13).

Theorem 1 demonstrates that the problem of evaluating the price of anarchy in a given game can effectively be recast as a problem of solving for the appropriate coefficients (λ, μ) that satisfy (11.13) and maximize $\frac{\lambda}{1+\mu}$. This analysis naturally extends to guarantees over a family of games $\mathcal{G}$,

$$\text{PoA}(\mathcal{G}) \geq \inf_{\lambda>0, \mu>-1} \left\{ \frac{\lambda}{1+\mu} : G \text{ is } (\lambda, \mu) \text{ smooth for all } G \in \mathcal{G} \right\}, \tag{11.15}$$

where the above expression is referred to as the *robust price of anarchy* (Roughgarden 2015). In line with the forthcoming discussion (cf., Sect. 4.4), implementing a learning rule that leads to the set of coarse correlated equilibria provides performance guarantees that conform to this robust price of anarchy.

One example of an entire class of games with known price of anarchy bounds is congestion games with affine congestion functions (Roughgarden 2005) (see also Example 2). Another class is valid utility games, introduced in Vetta (2002), which is very relevant to distributed resource utilization problems. A critical property of valid utility games is a system-level objective that is *submodular*. Submodularity corresponds to a notion of decreasing marginal returns that is a common feature of many objective function in engineering systems. A set-based function $f : 2^N \to \mathbb{R}$ is submodular if for any $S \subseteq T \subseteq N \setminus \{i\}$, we have

$$f(S \cup \{i\}) - f(S) \geq f(T \cup \{i\}) - f(T). \tag{11.16}$$

In each of these settings, Roughgarden (2015) has derived the appropriate smoothness parameters hence providing the price of anarchy guarantees. Accordingly, the resulting price of anarchy holds for coarse correlated equilibrium as well as Nash equilibrium.

Theorem 2 (Roughgarden 2015; Vetta 2002). *Consider any game* $G = (N, \{\mathcal{A}_i\}, \{U_i\}, W)$ *that satisfies the following three properties:*

(i) *The objective function* W *is submodular;*
(ii) *For any agent* $i \in N$ *and any action profile* $a \in \mathcal{A}$,

$$U_i(a) \geq W(a) - W(a_i = \emptyset, a_{-i}),$$

where $a_i = \emptyset$ *is when player* i *is removed from the game;*
(iii) *For any action profile* $a \in \mathcal{A}$, *the sum of the agents' utilities satisfies*

$$\sum_{i \in N} U_i(a) \leq W(a).$$

We will refer to such a game as a valid utility game. Any valid utility game G *is smooth with parameters* $\lambda = 1$ *and* $\mu = 1$*; hence, the robust price of anarchy is* $1/2$ *for the class of valid utility games. Accordingly, the efficiency guarantees associated with any coarse correlated equilibrium* $z^{\text{cce}} \in \Delta(\mathcal{A})$ *in a valid utility game satisfies*

$$W(z^{\text{cce}}) \geq \left(\frac{1}{2}\right) W(a^{\text{opt}}).$$

One example of a valid utility game is the vehicle-target assignment problem which will be presented in Example 4. Here, the system-level objective function is submodular and Condition (ii) in Theorem 2 is satisfied by the given design. Further, it is straightforward to verify that Condition (iii) is also satisfied. Accordingly, all coarse correlated equilibria in the designed game for the vehicle-target assignment problem are at least 50% efficient. Consequently, the application of learning rules that lead to coarse correlated equilibria (cf., Sect. 4.4) will lead to a collective behavior in line with these efficiency guarantees.

3.4 Game Structures

The two components associated with a game-theoretic design are the agent utility functions, which define an underlying game, and the learning rule. Both components impact various performance objectives associated with the distributed control design. The specification of the agent utility functions directly impacts the price of anarchy, which can be viewed as the efficiency associated with the asymptotic collective behavior. On the other hand, the specification of the learning algorithm dictates the transient behavior in its attempt to drive the collective behavior to the solution concept of interest.

At first glance it appears that the objectives associated with these two components are unrelated to one another. For example, one could employ a design where (i) the agents' utility functions are chosen to optimize the price of anarchy of pure Nash equilibria and (ii) a learning algorithm is employed that drives the collective behavior to a pure Nash equilibrium. Unfortunately, such decoupling is not necessarily possible due to limitations associated with (ii). As previously discussed, there are no "natural dynamics" of the form

$$a_i(t) = \Pi_i\left(a(0), a(1), \ldots, a(t-1); U_i\right) \tag{11.17}$$

that lead to a (pure or mixed) Nash equilibrium in every game (Hart and Mas-Colell 2003), where "natural" refers to uncoupled dynamics (i.e., agents are uninformed of the utility functions of other agents) and rules out behaviors such as exhaustive search or centralized coordination.

Given such impossibility results, it is imperative that the game design component addresses objectives beyond just price of anarchy. In particular, it is of paramount importance that the resulting game has properties that can be exploited in distributed learning. In this section we will review such game structures. Each of these game structures provides a degree of alignment between the agents' utility functions $\{U_i\}$ and a system-level potential function $\phi : \mathcal{A} \to \mathbb{R}$.

The first class of games we introduce, termed potential games (Monderer and Shapley 1996), exhibits perfect alignment between the agents' utility functions and the potential function ϕ.

Definition 4 (Potential Game). A game G is an (exact) potential game if there exists a potential function $\phi : \mathcal{A} \to \mathbb{R}$ such that for any action profile $a \in \mathcal{A}$, agent $i \in N$, and action choice $a_i' \in \mathcal{A}_i$, we have

$$U_i(a_i', a_{-i}) - U_i(a_i, a_{-i}) = \phi(a_i', a_{-i}) - \phi(a_i, a_{-i}). \tag{11.18}$$

Note that any maximizing action profile $a \in \arg\max_{a \in \mathcal{A}} \phi(a)$ is a pure Nash equilibrium; hence, a pure Nash equilibrium is guaranteed to exist in any potential game. Further, as we will see in the forthcoming Sect. 4, the structure inherent to potential games can be exploited to bypass the impossibility result highlighted above. In other words, there are natural dynamics that lead to a Nash equilibrium in any potential game. We will survey some of these dynamics in Sect. 4.

There are several variants of potential games that seek to relax the equality given in (11.18) while preserving the exploitability of the game structure for distributed learning. One of the properties that is commonly exploited in distributed learning is the monotonicity of the potential function along a *better reply path*, which is defined as follows:

Definition 5 (Better Reply Path). A better reply path is a sequence of joint actions $a^1, a^2, \ldots, a^m$ such that for each $k \in \{1, \ldots, m-1\}$ (i) $a^{k+1} = (a_i, a_{-i}^k)$ for some agent $i \in N$ with action $a_i \in \mathcal{A}_i$, $a_i \neq a_i^k$, and (ii) $U_i(a^{k+1}) > U_i(a^k)$.

Informally, a better reply path is a sequence of joint actions where each subsequent joint action is the result of an advantageous unilateral deviation. In a potential game, the potential function is monotonically increasing along a better reply path. Since the joint action set $\mathcal{A}$ is finite, any better reply will lead to a pure Nash equilibrium in a finite number of iterations. This property is known as the *finite improvement property* (Monderer and Shapley 1996).[4]

We now introduce the class of weakly acyclic games which relaxes the finite improvement property condition.

Definition 6 (Weakly Acyclic Game). A game G is weakly acyclic under better replies if for any joint action $a \in \mathcal{A}$ there exists a better reply path from a to a pure Nash equilibrium of G.

As with potential games, a pure Nash equilibrium is guaranteed to exist in any weakly acyclic game. One advantage of considering broader game classes as a mediating layer for game-theoretic control designs is the expansion of available game design methodologies for designing agent utility functions within that class.

3.5 Illustrative Examples

At first glance it may appear that the framework of potential games (or weakly acyclic games) is overly restrictive as a framework for the design of networked control systems. Here, we provide three examples of potential games, which illustrates the breadth of the problem domains that can be modeled and analyzed within this framework.

The first example focuses on distributed routing and highlights how a reasonable model of user behavior, i.e., users seeking to minimize their experienced congestion, constitutes a potential game.

Example 2 (Distributed Routing). A routing problem consists of a collection of self-interested agents that need to utilize a common network to satisfy their individual demands. The network is characterized by a collection of edges $E = \{e_1, \ldots, e_m\}$ where each edge $e \in E$ is associated with an anonymous congestion function $c_e : \{1, 2, \ldots\} \to \mathbb{R}$ that defines the congestion associated with that edge as a function of the number of agents using that edge. That is, $c_e(k)$ is the congestion on edge e when there are $k \geq 1$ agents using that edge. Each agent $i \in N$ is associated

[4] Commonly studied variants of exact potential games, e.g., ordinal or weighted potential games, also possess the finite improvement property.

with an action set $\mathcal{A}_i \subseteq 2^E$, which satisfies the agent's underlying demands, as well as a local cost function $J_i : \mathcal{A} \to \mathbb{R}$ of the form

$$J_i(a_i, a_{-i}) = \sum_{e \in a_i} c_e(|a|_e),$$

where $|a|_e = |\{i \in N : e \in a_i\}|$ denotes the number of agents using edge e in the allocation a.[5] In general, a system designer would like to allocate the agents over the network to minimize the aggregate congestion given by

$$C(a) = \sum_{e \in E} |a|_e \cdot c_e\left(|a|_e\right).$$

It is well known that any routing game of the above form, which is commonly referred to as an anonymous congestion game, is a potential game with a potential function $\phi : \mathcal{A} \to \mathbb{R}$ of the form

$$\phi(a) = \sum_{e \in E} \sum_{k=1}^{|a|_e} c_e\left(k\right).$$

This implies that a pure Nash equilibrium is guaranteed to exist in any anonymous congestion, namely, any action profile that minimizes $\phi(a)$. Furthermore, it is often the case that this is unique pure Nash equilibrium with regard to aggregate behavior, i.e., $a^{\text{ne}} \in \arg\min_{a \in \mathcal{A}} \phi(a)$. The fact that the potential function and the system cost are not equivalent, i.e., $\phi(\cdot) \neq C(\cdot)$, can lead to inefficiencies of the resulting Nash equilibria.

The second example focuses on coordination games over graphs. A coordination game is typically posed between two agents where each agent's utility function favors agreement on an action choice over disagreement. However, the agents may have different preferences over which action is agreed upon. Graphical coordination games, or coordination games over graphs, extend such two agent scenarios to n agent scenarios where the underlying graph depicts the population that each agent is seeking to coordinate with.

Example 3 (Graphical Coordination Games). Graphical coordination games characterize a class of strategic interactions where the agents' utility functions are derived from local interactions with neighboring agents. In a graphical coordination game, each agent $i \in N$ is associated with a common action set $\mathcal{A}_i = \bar{\mathcal{A}}$, a neighbor set $\mathcal{N}_i \subseteq N$, and a utility function of the form

[5]Here, we use cost functions $J_i(\cdot)$ instead of utility functions $U_i(\cdot)$ in situation where the agents are minimizers instead of maximizers.

$$U_i(a) = \sum_{j \in \mathcal{N}_i} \mathcal{U}(a_i, a_j) \tag{11.19}$$

where $\mathcal{U} : \bar{\mathcal{A}} \times \bar{\mathcal{A}} \to \mathbb{R}$ captures the (symmetric) utility associated with a pairwise interaction. As an example, $\mathcal{U}(a_i, a_j)$ designates the payoff for agent i selecting action a_i that results from the interaction with agent j selecting action a_j. Throughout, we adopt the convention that the payoff $\mathcal{U}(a_i, a_j)$ is associated with the player i whose action a_i is the first in the tuple (a_i, a_j).

In the case where the common action set has two actions, i.e., $\bar{\mathcal{A}} = \{x, y\}$, and the interaction graph is undirected, i.e., $j \in \mathcal{N}_i \Leftrightarrow i \in \mathcal{N}_j$, it is straightforward to show that this utility structure gives rise to a potential game with a potential function of the form

$$\phi(a) = \frac{1}{2} \sum_{(i,j) \in E} \phi_{\text{pw}}(a_i, a_j) \tag{11.20}$$

where $\phi_{\text{pw}} : \bar{\mathcal{A}} \times \bar{\mathcal{A}} \to \mathbb{R}$ is a local potential function. One choice for this local potential function is the following:

$$\begin{aligned}
\phi_{\text{pw}}(x, x) &= 0, \\
\phi_{\text{pw}}(y, x) &= \mathcal{U}(y, x) - \mathcal{U}(x, x), \\
\phi_{\text{pw}}(x, y) &= \mathcal{U}(y, x) - \mathcal{U}(x, x), \\
\phi_{\text{pw}}(y, y) &= (\mathcal{U}(y, y) - \mathcal{U}(x, y)) - (\mathcal{U}(y, x) - \mathcal{U}(x, x)).
\end{aligned}$$

Observe that any potential function $\phi'_{\text{pw}} = \phi_{\text{pw}} + \alpha$ where $\alpha \in \mathbb{R}$ also leads to a potential function for the given graphical coordination game.

The first two examples show how potential games could naturally emerge in two different types of strategic scenarios. The last example we present focuses on an engineering-inspired resource allocation problem, termed the vehicle-target assignment problem (Murphey 1999), where the vehicles' utility functions are engineered so that the resulting game is a potential game.

Example 4 (Vehicle-Target Assignment Problem). In the well-studied vehicle-target assignment problem, there is a finite set of targets $\mathcal{T}$, and each target $t \in \mathcal{T}$ has a relative value of importance $v_t \geq 0$. Further, there are a set of vehicles $N = \{1, 2, \ldots, n\}$ where each vehicle $i \in N$ has an invariant success/destroy probability satisfying $0 \leq p_i \leq 1$ and a set of possible assignment $\mathcal{A}_i \subseteq 2^{\mathcal{T}}$. The goal of vehicle-target assignment problem is to find an allocation of vehicles to targets $a \in \mathcal{A}$ to optimize a global objective $W : \mathcal{A} \to \mathbb{R}$ of the form

$$W(a) = \sum_{t \in \mathcal{T}(a)} v_t \cdot \left(1 - \prod_{j:t \in a_j} (1 - p_j)\right)$$

where $\mathcal{T}(a) \subseteq \mathcal{T}$ denotes the collection of targets that are assigned to at least one agent, i.e., $\mathcal{T}(a) = \cup_{i \in N} a_i$.

Note that in this engineering-based application, there is no appropriate model of utility functions of the engineered vehicles. Rather, vehicle utility functions are designed with the goal of engineering desirable system-wide behavior. Consider one such design where the utility functions of the vehicles are set as the marginal contribution of the vehicles to the system-level objective, i.e., for each vehicle $i \in N$ and allocation $a \in \mathcal{A}$ we have

$$U_i(a) = \sum_{t \in a_i} v_t \cdot \left(1 - \prod_{j:t \in a_j} (1 - p_j)\right) - v_t \cdot \left(1 - \prod_{j \neq i:t \in a_j} (1 - p_j)\right),$$

$$= \sum_{t \in a_i} v_t \cdot \left(p_i \prod_{j \neq i:t \in a_j} (1 - p_j)\right).$$

Given this design of utility functions, it is straightforward to verify that the resulting game is a potential game with potential function $\phi(a) = W(a)$. This immediately implies that any optimal allocation, $a^{\text{opt}} \in \arg\max_{a \in \mathcal{A}} W(a)$, is a pure Nash equilibrium. However, other inefficient Nash equilibria may also exist due to the lack of uniqueness of Nash equilibrium for such scenarios.

3.6 A Brief Review of Game Design Methodologies

The examples in the previous section illustrate various settings that happen to fall under the special category of potential games. Given that utility function specification is a design degree of freedom in the prescriptive paradigm, it is possible to exploit this degree of freedom to design utility functions to induce desirable structural properties.

There are several objectives that a system designer needs to consider when designing the game that defines the interaction framework of the agents in a multiagent system (Marden and Wierman 2013). These goals could include (i) ensuring the existence of a pure Nash equilibrium, (ii) ensuring that the agents' utility functions fit into the realm of potential games, or (iii) ensuring that the agents' utility functions optimize the price of anarchy/price of stability over an admissible class of agent utility functions, e.g., local utility functions. While recent research has identified the full space of methodologies that guarantee (i) and (ii) (Gopalakrishnan et al. 2014), the existing research has yet to provide mechanisms for optimizing the price of anarchy.

The following theorem provides one methodology for the design of agent utility functions with guarantees on the resulting game structure (Marden and Wierman 2013; Wolpert and Tumor 1999).

Theorem 3. *Consider the class of resource utilization problems defined in Sect. 2.1 with agent set N, action sets $\{\mathcal{A}_i\}$, and a global objective $W : \mathcal{A} \to \mathbb{R}$. Define the marginal contribution utility function for each agent $i \in N$ and allocation $a \in \mathcal{A}$ as*

$$U_i(a) = \phi(a) - \phi(a_i^b, a_{-i}), \tag{11.21}$$

where $\phi : \mathcal{A} \to \mathbb{R}$ is any system-level function and $a_i^b \in \mathcal{A}_i$ is the fixed baseline action for agent i. Then the resulting game $G = \{N, \{\mathcal{A}_i\}, \{U_i\}, W\}$ is an exact potential game where the potential function is ϕ.

A few notes are in order regarding Theorem 3. First, the assignment of the agents' utility functions is a byproduct of the chosen system-level design function ϕ and the transformation of ϕ into the agents' utility functions, which is given by (11.21) and the choice of the baseline action a_i^b for each agent $i \in N$. Observe that the utility design presented in Example 4 is precisely the design detailed in Theorem 3 where $\phi = W$ and $a_i^b = \emptyset$ for each agent $i \in N$. While a system designer could clearly set $\phi = W$, judging whether this design choice is effective centers on a detailed analysis regarding the properties of the resulting game, e.g., price of anarchy. In fact, recent research has demonstrated that setting $\phi = W$ does not optimize the price of anarchy for a large class of objective functions W. Furthermore, there are also alternative mechanisms for transforming the system-level function ϕ to agent utility functions $\{U_i\}$, as opposed to (11.21), that provide similar guarantees on the structure of the resulting game, e.g., Shapley and weighted Shapley values (Gopalakrishnan et al. 2014). It remains an open question as to what combination, i.e., the transformation and system-level design function that the transformation operates on, gives rise to the optimal utility design.

4 Distributed Learning Rules

We now turn our attention toward distributed learning rules. We can categorize the learning algorithms into the following four areas:

Model-Based Learning. In model-based learning, each agent observes the past behavior of the other agents and uses this information to develop a model for the action choice of the other agents at the ensuing period. Equipped with this model, each agent can then optimally select its actions based on its expected utility at the ensuing time step. As the play evolves, so do the models of other agents.

Robust Learning. A learning algorithm of the form (11.2) defines a systematic rule for how individual agents process available information to formulate a decision. Many of the learning algorithms in the existing literature provide guarantees on the asymptotic collective behavior provided that the agents follow these rules precisely. Here, we explore the robustness of such learning algorithms, i.e., the asymptotic guarantees on the collective behavior preserved when agents follow variations of the prescribed learning rules stemming from delays in information or asynchronous clock rates.

Equilibrium Selection. The price of anarchy and price of stability are two measures characterizing the inefficiency associated with Nash equilibria. The differences between these two measures follow from the fact that Nash equilibria are often not unique. This lack of uniqueness of Nash equilibria prompts the question of whether deriving distributed learning that favor certain types of Nash equilibria is attainable. Focusing on the framework of potential games, we will review one such algorithm that guarantees the collective behavior will lead to the specific Nash equilibria that optimize the potential function. Note that when utility functions are engineered, as in Example 4, a system designer can often ensure that the resulting game is a potential game where the action profiles that optimize the potential function coincide with the action profiles that optimize the system-level objective. (We reviewed one such methodology in Sect. 3.6.)

Universal Learning. All of the above learning algorithms provide asymptotic guarantees when attention is restricted to specific game structures, e.g., potential games or weakly acyclic games. Here, we focus on the derivation of learning algorithms that provide desirable asymptotic guarantees irrespective of the underlying game structure. Recognizing the previously discussed impossibility of natural and universal dynamics leading to Nash equilibria (Hart and Mas-Colell 2003), we shift our emphasis from convergence to Nash equilibria to convergence to the set of coarse correlated equilibrium. We introduce one such algorithm, termed *regret matching*, that guarantees convergence to the set of coarse correlated equilibrium irrespective of the underlying game structure. Lastly, we discuss the implications of such learning algorithms on the efficiency of the resulting collective behavior.

We will primarily gauge the quality of a learning algorithm by characterizing the collective behavior as time $t \to \infty$. When merging a particular distributed learning algorithm with an underlying game, the efficiency analysis techniques presented in Sect. 3.2 can then be employed to characterize the quality of the emergent collective behavior with regard to a given system-level objective.

4.1 Model-Based Learning

The central challenge in distributed learning is dealing with the fact that each agent's environment is inherently nonstationary in that the environment from the perspective of any agent consists of the behaviors of other agents, which are evolving. A common approach in distributed learning is to have agents make decisions in a myopic fashion, thereby neglecting the ramifications of an agent's current decision on the future behavior of the other agents. In this section we review two learning algorithms of this form that we categorize as model-based learning algorithms. In model-based learning, each agent observes the past behavior of the other agents and utilizes this information to develop a behavioral model of the other agents. Equipped with this behavioral model, each agent then performs a myopic best response seeking to optimize its expected utility. It is important to stress here that the goal is not to accurately model the behavior of the other agents in ensuing period. Rather, the goal is to derive systematic agent responses that will guide the collective behavior to a desired equilibrium.

4.1.1 Fictitious Play

One of the most well-studied algorithms of this form is fictitious play (Monderer and Shapley 1996). Here, each agent uses the empirical frequency of past play as a model for the behavior of the other agents at the ensuing time step. To that end, define the empirical frequency of play for each player $i \in N$ at time $t \in \{1, 2, \ldots\}$ as $q_i(t) = \{q_i^{a_i}\}_{a_i \in \mathcal{A}_i} \in \Delta(\mathcal{A}_i)$ where

$$q_i^{a_i}(t) = \frac{1}{t} \sum_{\tau=0}^{t-1} I\{a_i(\tau) = a_i\}, \tag{11.22}$$

and $I\{\cdot\}$ is the usual indicator function. At time t, each agent seeks to myopically maximize its expected utility given the belief that each agent $j \neq i$ will select its action independently according to a strategy $q_j(t)$. This update rule takes on the form

$$a_i(t) \in \arg\max_{a_i \in \mathcal{A}_i} \sum_{a_{-i} \in \mathcal{A}_{-i}} U_i(a_i, a_{-i}) \prod_{j \neq i} q_j^{a_j}(t). \tag{11.23}$$

The following theorem provided in Monderer and Shapley (1996) characterizes the long run behavior associated with fictitious play in potential games.

Theorem 4. *Consider any exact potential game G. If all players follow the fictitious play learning rule, then the players' empirical frequencies of play $q_1(t), \ldots, q_n(t)$ will converge to a Nash equilibrium of the game G.*

The fictitious play learning rule provides a mechanism to guide individual agent behavior in distributed control systems when the agents (i) can observe the previous action choices of the other agents in the system and (ii) have access to the structural form of their utility function. Further, fictitious play provides provable guarantees on the emergent collective behavior provided that the system can be modeled by an exact potential game. For example, consider the distributed routing problem given in Example 2 which can be modeled as a potential game irrespective of the number of agents, the number of edges, the topology of the network, or the edge-specific latency functions. Regardless of the structure of the routing problem, the fictitious play algorithm can be employed to drive the collective system behavior to a Nash equilibrium.

While the asymptotic guarantees associated with fictitious play in distributed routing problems is appealing, the implementation of fictitious play in such settings is problematic. First, each agent must be able to observe the specific behavior of all other agents in the network each period. Second, the choice of each agent at any time given in (11.23) requires (i) knowledge of the structural form of the agent's utility function and (ii) computing an expectation of its utility function, which involves evaluating a weighted summation over $|\mathcal{A}_{-i}|$ terms. In large-scale systems, such as distributed routing, each of these requirements could be prohibitive. Accordingly, research has attempted to alter the fictitious play algorithm to minimize such requirements while preserving the desirable asymptotic guarantees.

4.1.2 Variants of Fictitious Play

One of the first attempts to relax the implementation requirements associated with fictitious play centered on the computation of a best response given in (11.23). In Lambert et al. (2005), the authors proposed a sample-based approach for computing this best response, where each agent randomly drew samples of the other agents' behavior using their empirical frequencies of play and evaluated the average performance of each possible routing decision against the drawn samples. The choice with the best average performance was then substituted for the choice that maximized the agent's expected utility in (11.23), and the process was repeated. While simulations demonstrated reasonable performance even for limited samples, unfortunately preserving the theoretical asymptotic guarantees associated with fictitious play required that the number of samples drawn each period grew prohibitively large.

A second variant of fictitious play focused on the underlying asymptotic guarantees given in Theorem 4, which state that the empirical frequency of play converges to a Nash equilibrium. It is important to highlight this does not imply that the day-to-day behavior of the agents converges to a Nash equilibrium, e.g., the agents' day-to-day behavior could oscillate yielding a frequency of play consistent with a Nash equilibrium. Furthermore, the cumulative payoff may be less than the payoff associated with the limiting empirical frequencies. With this issue in mind, Fudenberg and Levine (1995) introduced a variant of fictitious play that assures a specific payoff consistency property against arbitrary environments, i.e., not just when other agents employ fictitious play.

4.1.3 Joint Strategy Fictitious Play with Inertia

The focus in model-based learning is not whether such models accurately reflect the behavior of the other agents. Rather, the focus is on whether systematic responses to potentially inaccurate models can guide the collective behavior to a desired equilibrium. The behavioral models used in fictitious play, i.e., assuming each agent will play a strategy independently according to the agent's empirical frequency of play, provided nice asymptotic guarantees but was prohibitive from an implementations perspective. Here, we consider a variant of fictitious play, termed joint strategy fictitious play (JSFP), which provides similar asymptotic guarantees while alleviating many of the computational and observational challenges associated with fictitious play (Marden et al. 2009). The main difference between fictitious play and joint strategy fictitious play resides in the behavioral model of the other agents. In joint strategy fictitious play, each agent presumes that the other players will select an action collectively in accordance with their empirical frequency of their past joint play. In two-player games, fictitious play and joint strategy fictitious play are equivalent. However, the learning algorithms yield fundamentally different behavior beyond two-player games.

We begin by defining the average hypothetical utility of agent $i \in N$ for each action $a_i \in \mathcal{A}$ as

$$\bar{U}_i^{a_i}(t) = \frac{1}{t}\sum_{\tau=0}^{t-1} U_i(a_i, a_{-i}(\tau)) = \frac{t-1}{t}\bar{U}_i^{a_i}(t-1) + \frac{1}{t}U_i(a_i, a_{-i}(t-1)). \tag{11.24}$$

Note that this average hypothetical utility is computed under the belief that the action choices of the other agents remain unchanged. Now, consider the decision-making rule where each agent $i \in N$ independently selects its action probabilistically according to the rule

$$a_i(t) = \begin{cases} \arg\max_{a_i \in \mathcal{A}_i} \bar{U}_i^{a_i}(t) & \text{with probability } (1-\epsilon), \\ a_i(t-1) & \text{with probability } \epsilon, \end{cases} \tag{11.25}$$

where $\epsilon > 0$ is referred to as the agent's inertia or probabilistic reluctance to change actions. Hence, with high probability, i.e., probability $(1-\epsilon)$, each agent selects the action that maximizes the agent's hypothetic utility.

The following theorem from Marden et al. (2009) characterizes the long run behavior of joint strategy fictitious play in potential games.

Theorem 5. *Consider any exact potential game G. If all players following the learning algorithm joint strategy fictitious play defined above, then the joint action profile will converge almost surely to a pure Nash equilibrium of the game G.*

Hence, JFSP with inertia provides similar asymptotic guarantees to fictitious play while minimizing the computational and observational burden on the agents. The name "joint strategy fictitious play" is derived from the fact that maximizing the average hypothetical utility in (11.24) is equivalent to maximizing an expected utility under the belief that all agents will play collectively according to the empirical frequency of their past joint play.

4.2 Robust Distributed Learning

Both fictitious play and joint strategy fictitious play are intricate decision-making rules that provide guarantees regarding the emergent collective behavior. A natural question that emerges when considering the practicality of such rules for control of networked systems is the robustness of these guarantees to common implementation issues including asynchronous clocks, noisy payoffs, and delays in information, among others. This section highlights that the framework of potential games, or more generally weakly acyclic games, is inherently robust to such issues.

We review the result in Young (2004) that deals with this exact issue. In particular, Young (2004) demonstrates the robustness of weakly acyclic games by identifying a broad family of learning rules, termed *finite memory better response processes*, with the property that any rule within this family will probably guide the collective behavior to a pure Nash equilibrium in any weakly acyclic game.

A finite memory better reply process with inertia is any learning algorithm of the following form: at each time t, each agent selects its action independently according to the rule

$$a_i(t) = \begin{cases} B_i^m(h^m(t)) & \text{with probability } (1-\epsilon), \\ a_i(t-1) & \text{with probability } \epsilon, \end{cases} \tag{11.26}$$

where $m \geq 1$ is the size of the agent's memory, $\epsilon > 0$ is the agent's inertia, $h^m(t) = \{a(t-1), a(t-2), \ldots, a(t-m)\}$ denotes the previous m action profiles, and $B_i^m : \mathcal{A}^m \to \Delta(\mathcal{A}_i)$ is the finite memory better reply process.[6] A finite memory better reply process $B_i^m(\cdot)$ can be any process that satisfies the following properties:

- If the history is saturated, i.e., $h^m(t) = \{\bar{a}, \bar{a}, \ldots, \bar{a}, \bar{a}\}$ for some action profile $\bar{a} \in \mathcal{A}$, then the strategy $p_i = B_i^m(h^m(t))$ must satisfy
  - If $\bar{a}_i \in \arg\max_{a_i \in \mathcal{A}_i} U_i(a_i, \bar{a}_{-i})$, then $p_i^{\bar{a}_i} = 1$ and $p_i^{a_i} = 0$ for all $a_i \neq \bar{a}_i$.
  - Otherwise, if $\bar{a}_i \notin \arg\max_{a_i \in \mathcal{A}_i} U_i(a_i, \bar{a}_{-i})$, then $p_i^{a_i} > 0$ if and only if $U_i(a_i, \bar{a}_{-i}) \geq U_i(\bar{a}_i, \bar{a}_{-i})$.
- If the history is not saturated, then the strategy $p_i = B_i^m(h^m(t))$ can be any probability distribution in $\Delta(\mathcal{A}_i)$.[7]

In summary, the only constraint imposed on a finite memory better reply process is that a better reply to saturated memory $\{a, \ldots, a\}$ is consistent with a better reply to the single action profile a.

[6] We write $a_i(t) = B_i^m(h^m(t))$ with the understanding that this implies that the action profile $a_i(t)$ is chosen randomly according to the probability distribution specified by $B_i^m(h^m(t))$.

[7] The actual definition of a finite better reply process considered in Young (2004) puts a further condition on the structure of B_i^m under the case where the memory is not saturated, i.e., the strategy assigns positive probability to any action with strictly positive regret. However, an identical proof holds for any B_i^m that satisfies the weaker conditions set forth in this chapter.

The following theorem from Young (2004) (Theorem 6.2) demonstrates the inherent robustness of weakly acyclic games.

Theorem 6. *Consider any weakly acyclic game G. If all agents follow a finite memory better reply process defined above, then the joint action profile will converge almost surely to a pure Nash equilibrium of the game G.*

One can view this result from two perspectives. The first perspective is that the system designer has extreme flexibility in designing learning rules for weakly acyclic games that guarantee the agents' collective behavior will converge to a pure Nash equilibrium. The second perspective is that perturbations of a nominal learning rule, e.g., agents updating asynchronously or responding to delayed or inaccurate histories, will also satisfy the conditions above and ultimately lead behavior to a Nash equilibrium as well. These perspectives provide the basis for our claim of robust distributed learning.

4.3 Equilibrium Selection in Potential Games

The preceding discussion focused largely on algorithms that ensured the emergent collective behavior constitutes a (pure) Nash equilibrium. In the case where there are multiple Nash equilibria, these algorithms provide no guarantees on which equilibrium is likely to emerge. Accordingly, characterizing the efficiency associated with the emergent collective behavior is equivalent to characterizing the efficiency associated with the worst performing Nash equilibrium, i.e., the price of anarchy.

In this section we explore the notion of equilibrium selection in distributed learning. That is, are there classes of distributed learning algorithms that converge to specific classes of equilibria? One motivation for pursuing such developments is the marginal cost utility, given in Theorem 3 with the design choice $\phi = W$, which ensures that the optimal allocation is a Nash equilibrium, i.e., the price of stability is 1. Accordingly, the focus of this section will be on learning dynamics that converge to the most efficient action profile in potential games, i.e., the action profile that maximizes the potential function.

4.3.1 Log-Linear Learning

We begin this subsection by describing a simple asynchronous best-reply process, where each agent chooses a best reply when given the opportunity to revise its strategy. Let $a(t)$ represent the action profile at time t. The action profile at time $t + 1$ is chosen as follows:

(i) An agent $i \in N$ is randomly picked to update its action according to a uniform distribution.

(ii) Agent i selects an action that is a best response to the action profile played by the other agents in the previous period, i.e.,

$$a_i(t+1) \in \arg\max_{a_i \in \mathcal{A}_i} U_i(a_i, a_{-i}(t)). \tag{11.27}$$

(iii) All other agents $j \neq i$ play their previous actions, i.e., $a_{-i}(t+1) = a_{-i}(t)$.
(iv) The process is then repeated.

It is straightforward to see that the above process will converge almost surely to a pure Nash equilibrium in any potential game by observing that $\phi(a(t+1)) \geq \phi(a(t))$ for all times t. Accordingly, the efficiency guarantees associated with the application of this algorithm to a potential game are in line with the price of anarchy of the game.

Here, a slight modification, or perturbation, is introduced of the above best-reply dynamics that ensures that the resulting behavior leads to the pure Nash equilibrium that optimizes the potential function, i.e., $a^{\text{opt}} \in \arg\max_{a \in \mathcal{A}} \phi(a)$. The algorithm, known as log-linear learning or the logit response dynamics (Alos-Ferrer and Netzer 2010; Blume 1993, 1997; Marden and Shamma 2012; Young 1998), follows the best-reply process highlighted above where step (ii) is replaced by a noisy best response. More formally, step (ii) is now of the form:

(ii) Agent i selects an action $a_i(t+1)$ according to a probability distribution $p_i(t) = \{p_i^{a_i}(t)\}_{a_i \in \mathcal{A}_i} \in \Delta(\mathcal{A}_i)$ that is of the form

$$p_i^{a_i}(t) = \frac{e^{(1/T)\cdot U_i(a_i, a_{-i}(t))}}{\sum_{\tilde{a}_i \in \mathcal{A}_i} e^{(1/T)\cdot U_i(\tilde{a}_i, a_{-i}(t))}}, \tag{11.28}$$

where the parameter $T > 0$ is referred to as the temperature.

A few remarks are in order regarding the update protocol specified in (11.28). First, when $T \to \infty$, the agent's strategy is effectively a uniform distribution over the agent's action set. Second, when $T \to 0^+$, the agent's strategy is effectively the best response strategy given in (11.27). Lastly, we present this algorithm (and the forthcoming Binary Log-Linear Learning) with regard to a fixed temperature parameter that is common to all agents. However, there are variations of this algorithm which allow for annealing of this temperature parameter that preserve the resulting asymptotic guarantees, e.g., Zhu and Martínez (2013).

The following theorem establishes the asymptotic guarantees associated with the learning algorithm log-linear learning in potential games (Blume 1993, 1997; Young 1998).

Theorem 7. *Consider any potential game G with potential function ϕ. If all players follow the learning algorithm log-linear learning with temperature $T > 0$, then the resulting process has a unique stationary distribution $\pi = \{\pi^a\}_{a \in \mathcal{A}} \in \Delta(\mathcal{A})$ of the form*

$$\pi^a = \frac{e^{(1/T)\cdot\phi(a)}}{\sum_{\tilde{a}\in\mathcal{A}} e^{(1/T)\cdot\phi(\tilde{a})}}. \tag{11.29}$$

The stationary distribution of the process given in (11.29) follows the same intuition as presented for the update protocol in (11.28). That is, when $T \to \infty$ the stationary distribution is effectively a uniform distribution over the joint action set $\mathcal{A}$. However, when $T \to 0^+$, all of the weight of the stationary distribution is concentrated on the action profiles that maximize the potential function ϕ. The above stationary distribution provides an accurate assessment of the resulting asymptotic behavior due to the fact that the log-linear learning process is both irreducible and aperiodic, hence (11.29) is the unique stationary distribution.

Merging log-linear learning with the marginal contribution utility design given in Theorem 3 leads to the following corollary.

Corollary 1. *Consider the class of resource allocation problems defined in Sect. 2.1 with agent set N, action sets $\{\mathcal{A}_i\}$, and a global objective $W : \mathcal{A} \to \mathbb{R}$. Consider the following game-theoretic control design:*

(i) *Assign each agent a utility function that captures the agent's marginal contribution to the global objective, i.e.,*

$$U_i(a) = W(a) - W(a_i^b, a_{-i}), \tag{11.30}$$

where $a_i^b \in \mathcal{A}_i$ is any fixed baseline action for agent i.

(ii) *Each agent follows the log-linear learning rule with temperature parameter $T > 0$.*

Then the resulting process has a unique stationary distribution $\pi(T) = \{\pi^a(T)\}_{a\in\mathcal{A}} \in \Delta(\mathcal{A})$ of the form

$$\pi^a(T) = \frac{e^{(1/T)\cdot W(a)}}{\sum_{\tilde{a}\in\mathcal{A}} e^{(1/T)\cdot W(\tilde{a})}}. \tag{11.31}$$

Observe that this design rule ensures that the resulting asymptotic behavior will be concentrated around the allocations that maximize the global objective W. This fact has made this design methodology an attractive option for several domains including wind farms, sensor networks, and coordination of unmanned vehicles, among others.

4.3.2 Binary Log-Linear Learning

The framework of log-linear learning imposes a fairly rigid structure on the update process of the agents. This structure mandates that (i) only one agent updates the action choice at any iteration, (ii) agents are able to select any action in their action set, and (iii) agents are able to assess their utility for any alternative action choice given the observed behavior of the other agents. In general, Alos-Ferrer and Netzer (2010) demonstrates that relaxing these structures arbitrarily can significantly alter the resulting asymptotic guarantees associated with log-linear learning. However, in each of the scenarios variations of log-linear learning can preserve the asymptotic guarantees while making the structure more amenable to engineering systems (Marden and Shamma 2012).

Here, we present a variation of log-linear learning that preserves the asymptotic guarantees associated with log-linear learning while accommodating restrictions in the agents' action sets. By restrictions in action sets, we mean that the set of actions available to a given agent is dependent on the agent's current action choice, and we express this dependence by the function $R_i : \mathcal{A}_i \to 2^{\mathcal{A}_i}$ where $a_i \in R_i(a_i)$ for all a_i. That is, if the choice of agent i at time t is $a_i(t)$, then the ensuing choice of the agent $a_i(t+1)$ must be contained in the set $R_i(a_i(t))$. Throughout this section, we consider restricted action sets that satisfy two properties:

(i) *Reversibility*: Let a_i, a_i' be any two action choices in $\mathcal{A}_i$. If $a_i' \in R_i(a_i)$ then $a_i \in R_i(a_i')$.
(ii) *Completeness*: Let a_i, a_i' be any two action choices in $\mathcal{A}_i$. There exists a sequence of actions $a_i = a_i^0, a_i^1, \ldots, a_i^m = a_i'$ with the property that $a_i^{k+1} \in R_i(a_i^k)$ for all $k \in \{0, \ldots, m-1\}$.

One motivation for considering restricted action sets of the above form is when the individual agents have mobility limitations, e.g., mobile sensor networks.

Note that the log-linear learning update rule given in (11.28) has full support on the agent's action set $\mathcal{A}_i$ thereby disqualifying this algorithm for use in the case where there are restrictions in action sets. Here, we seek to address the question of how to alter the algorithm so as to preserve the asymptotic guarantees, i.e., convergence in the stationary distribution to the action profile that maximizes the potential function. One natural variation would be to replace (11.28) with a strategy of the form: for any $a_i \in R_i(a_i(t))$

$$p_i^{a_i}(t) = \frac{e^{(1/T)\cdot U_i(a_i, a_{-i}(t))}}{\sum_{\tilde{a}_i \in R_i(a_i(t))} e^{(1/T)\cdot U_i(\tilde{a}_i, a_{-i}(t))}}, \tag{11.32}$$

and $p_i^{a_i}(t) = 0$ for any $a_i \notin R_i(a_i(t))$. However, such modifications can have drastic consequences on the resulting asymptotic guarantees. In fact, such a rule is not even able to guarantee that the potential function maximizer is in the support of the limiting distribution as $T \to 0^+$ (Marden and Shamma 2012).

Here, we introduce a variation of log-linear learning, termed binary log-linear learning with restricted action sets (Marden and Shamma 2012), that preserves these asymptotic guarantees. Binary log-linear learning follows the same setup as log-linear learning where step (ii) is now of the form:

(ii) Agent i selects a trial action $a_i^{\mathrm{t}} \in R_i(a_i(t))$ according to any distribution with full support on the set $R_i(a_i(t))$. Conditioned on the selection of this trial action, the agent selects the action $a_i(t+1)$ according to a probability distribution $p_i(t) = \{p_i^{a_i}(t)\}_{a_i \in \mathcal{A}_i} \in \Delta(\mathcal{A}_i)$ of the form

$$p_i^{a_i}(t) = \begin{cases} a_i(t) & \text{with probability } \frac{e^{(1/T)\cdot U_i(a(t))}}{e^{(1/T)\cdot U_i(a(t))} + e^{(1/T)\cdot U_i(a_i^{\mathrm{t}}, a_{-i}(t))}}, \\ a_i^{\mathrm{t}} & \text{with probability } \frac{e^{(1/T)\cdot U_i(a_i^{\mathrm{t}}, a_{-i}(t))}}{e^{(1/T)\cdot U_i(a(t))} + e^{(1/T)\cdot U_i(a_i^{\mathrm{t}}, a_{-i}(t))}}, \end{cases} \tag{11.33}$$

where $p_i^{a_i}(t) = 0$ for any $a_i \notin \{a_i(t), a_i^{\mathrm{t}}\}$.

Much like log-linear learning, for any temperature $T > 0$ binary log-linear learning can be modeled by an irreducible and aperiodic Markov chain over the state space $\mathcal{A}$; hence, there is a unique stationary distribution which we denote by $\pi(T) = \{\pi^a(T)\}_{a \in \mathcal{A}}$. While log-linear learning provides the explicit form of the stationary distribution $\pi(T)$, the value of log-linear learning centers on the fact that the support of the limiting distribution is precisely the set of potential function maximizers, i.e.,

$$\lim_{T \to 0^+} \pi^a(T) > 0 \;\Leftrightarrow\; a \in \arg\max_{a \in \mathcal{A}} \phi(a)$$

The action profiles contained in the support of the limiting distribution are termed the *stochastically stable states*. Accordingly, log-linear learning ensures that an action profile is stochastically stable if and only if it is a potential function maximizer.

The following theorem from Marden and Shamma (2012) characterizes the long run behavior of binary log-linear learning.

Theorem 8. *Consider any potential game G with potential function ϕ. If all players follow the learning algorithm binary log-linear learning with restricted action set and temperature $T > 0$, then an action profile is stochastically stable if and only if it is a potential function maximizer.*

This theorem demonstrates that a system designer can effectively deal with restrictions in action sets by appropriately modifying the learning rule. However, a consequence of this is that we are no longer able to provide a precise characterization of the stationary distribution as a function of the temperature parameter T. Unlike log-linear learning, binary log-linear learning applied to such a game does not satisfy reversibility unless there are additional constraints imposed on the agents' restricted action sets, i.e., $|R_i(a_i)| = |R_i(a_i')|$ for all $i \in N$ and $a_i, a_i' \in \mathcal{A}_i$. Hence, in this theorem we forgo a precise analysis of the stationary distribution in favor of a coarse analysis of the stationary distribution that demonstrates roughly the same asymptotic guarantees.

4.3.3 Beyond Asymptotic Guarantees

In potential games, both log-linear learning and binary log-linear learning ensure that the resulting collective behavior can be characterized by the action profiles that maximize the potential function when the temperature $T \to 0^+$. Here, we focus on the question of characterizing the convergence rates of this process. That is, how long does it take for the collective behavior to reach these desired equilibrium points.

Several negative results have emerged regarding the convergence rates of such algorithms (Daskalakis et al. 2009; Hart and Mansour 2010; Shah and Shin 2010). In particular, Hart and Mansour (2010) and Shah and Shin (2010) demonstrates that in general the amount of time that it may take to reach such an equilibrium could be exponential in both the number of agents and the cardinality of their action sets. Accordingly, research has shifted to identifying whether there are classes of games and variants of the above dynamics that exhibit more desirable guarantees on the convergence rates.

The following briefly highlights three domains where such positive results exist.

Symmetric Parallel Congestion Games. Consider the class of congestion games introduced in Example 2. A symmetric parallel congestion game is a congestion game where each agent $i \in N$ has an action set $\mathcal{A}_i = \mathcal{R}$; that is, any agent can choose any single edge from the set of available roads $\mathcal{R}$. In Shah and Shin (2010), the authors demonstrate that the mixing times associated with log-linear learning could grow exponentially with regard to the number of players n even in such limited scenarios. However, the authors introduce a variant of log-linear learning, which effectively replaces Step (i) of the algorithm (pick an updating player uniformly) with a new procedure which biases the selection rate of certain agents based on the current action profile a. This modification of log-linear learning provides similar asymptotic guarantees with far superior transient guarantees. In particular, this variant of log-linear learning provides a mixing time that is nearly linear in the number of agents for this class of congestion games.

Semi-Anonymous Potential Games. In symmetric parallel congestion games, all of the agents are anonymous (or identical) with regard to their impact on the potential function and their available action choices. More formally, we will call two agents $i, j \in N$ anonymous in a potential game if (i) $\mathcal{A}_i = \mathcal{A}_j$ and (ii) $\phi(a) = \phi(a')$ for any action profiles a, a' where $a'_i = a_j$, $a'_j = a_i$, and $a'_k = a_k$ for all $k \neq i, j$. Accordingly, let $C_1, \ldots C_m$ represent a minimal partition of N such that each set of agents C_k, $k \in \{1, \ldots, m\}$ is anonymous with respect to one another, i.e., any agents $i, j \in C_k$ are anonymous with respect to each other. The authors in Borowski et al. (2013) derive a variant of log-linear learning algorithm, similar to the algorithm for symmetric parallel congestion games in Shah and Shin (2010) highlighted above, that provides mixing times that are nearly linear in the number of agents n, but exponential in the number of indistinguishable groups of agents, m.

Graphical Coordination Games. Consider the family of graphical coordination games introduced in Example 3. In Montanari and Saberi (2009), the authors study the mixing times associated with log-linear learning in a special class of graphical coordination games where the underlying pairwise utility function constitutes a 2×2 symmetric utility function. In particular, the authors demonstrate that the structure of the network, in particular the min-cut of graph, is intimately related to the underlying speed of convergence. A consequence of this characterization is that the mixing times associated with log-linear learning is effectively linear in the number of agents when the underlying graph is sparse.

4.4 Universal Learning

The preceding sections presented algorithms that guarantee convergence to Nash equilibria (or potential function maximizers) for specific game structures, e.g., potential games or weakly acyclic games. Here, we focus on the question of whether there are universal algorithms that provide convergence to an equilibrium irrespective of the underlying game structure. With regard to Nash equilibrium, it turns out that such an objective is impossible as demonstrated by Hart and Mas-Colell (2003) which establishes that no natural dynamics converge to a Nash equilibrium in all games. Here, the phrase natural seeks to disqualify dynamics that can be thought of as an exhaustive search or utilizing a central coordinator. Nonetheless, by relaxing our equilibrium requirements focus from Nash equilibria to coarse correlated equilibria, such universal algorithms do exists. In the following, we survey the most well-known algorithm that achieves this objective and discuss its implications on the efficiency of this broader class of equilibria.

In this section we present an algorithm proposed in Hart and Mas-Colell (2000), referred to as *regret matching*, that guarantees convergence to the set of coarse correlated equilibrium. The informational demands and computations associated with the decision-making rule regret matching is very similar to those presented for the algorithm joint strategy fictitious play with inertia highlighted above. The main driver for each agent's strategy selection is the regret associated with each of its actions. For any time $t \in \{1, 2 \dots\}$, the regret of agent $i \in N$ for action $a_i \in \mathcal{A}_i$ is defined as

$$R_i^{a_i}(t) = \bar{U}_i^{a_i}(t) - \bar{U}_i(t), \tag{11.34}$$

where $\bar{U}_i(t) = \frac{1}{t} \sum_{\tau=0}^{t-1} U_i(a(\tau))$ is the average utility received by agent i up to time t and $\bar{U}_i^{a_i}(t) = \frac{1}{t} \sum_{\tau=0}^{t-1} U_i(a_i, a_{-i}(\tau))$ is the average utility that would have been received by agent i up to time t if the agent committed to action a_i all time steps and the behavior of the other agents were unchanged. Observe that $R_i^{a_i}(t) > 0$ implies that agent i could have received a higher average utility if the agent had committed to the action a_i for all previous time steps and the action choices of the other agents was unchanged.

The regret matching algorithm proceeds as follows: at each time $t \in \{1, 2, \ldots\}$, each agent $i \in N$ independently selects its action according to the strategy $p_i(t) \in \Delta(\mathcal{A}_i)$ of the form

$$p_i^{a_i}(t) = \frac{\left[R_i^{a_i}(t)\right]_+}{\sum_{\tilde{a}_i \in \mathcal{A}_i} \left[R_i^{\tilde{a}_i}(t)\right]_+} \tag{11.35}$$

where $[\cdot]_+$ denotes the projection to the positive orthant, i.e., $[x]_+ = \max\{x, 0\}$.

The following theorem characterizes the long run behavior of regret matching in any game.

Theorem 9. *Consider any finite game G. If all players follow the learning algorithm regret matching defined above, then the positive regret for any agent $i \in N$ and action $a_i \in \mathcal{A}_i$ asymptotically vanishes, i.e.,*

$$\lim_{t \to \infty} [R_i^{a_i}(t)]_+ = 0. \tag{11.36}$$

Alternatively, the empirical frequency of play converges to the set of coarse correlated equilibria.

The connection between the condition (11.36) and the definition of coarse correlated equilibria stems from the fact that an agent's regret and average utility can also be computed using the empirical frequency of play $z(t) = \{z^a(t)\}_{a \in \mathcal{A}}$ where

$$z^a(t) = \frac{1}{t} \sum_{\tau=0}^{t-1} I\{a(\tau) = a\}. \tag{11.37}$$

In particular, at any time $t \in \{1, 2, \ldots\}$ we have that

$$U_i(z(t)) = \sum_{a \in \mathcal{A}} U_i(a) z^a(t) = \bar{U}_i(t). \tag{11.38}$$

Further, defining the marginal distribution of the empirical frequency of play of all agents $j \neq i$ as $z_{-i}^{a_{-i}}(t) = \sum_{a_i \in \mathcal{A}_i} z^{(a_i, a_{-i})}(t)$, we have

$$U_i(a_i, z_{-i}(t)) = \sum_{a_{-i} \in \mathcal{A}_{-i}} U_i(a_i, a_{-i}) z_{-i}^{a_{-i}}(t) = \bar{U}_i^{a_i}(t). \tag{11.39}$$

Accordingly, if a sequence of play $a(0), a(1), \ldots, a(t-1)$, satisfies (11.36), then we know that the empirical frequency of play $z(t)$ satisfies

$$\lim_{t\to\infty} \{U_i(z(t)) - U_i(a_i, z_{-i}(t))\} \le 0, \quad \forall i \in N,\ a_i \in \mathcal{A}_i. \tag{11.40}$$

Hence, the limiting empirical frequency of play $z(t)$ is contained in the set of coarse correlated equilibria. Note that the convergence highlighted above does not state that the empirical frequency of play will converge to any specific correlated equilibrium; rather, it merely states that the empirical frequency of play will approach the set of coarse correlated equilibria.

Lastly, we presented a version of regret matching that provides convergence to the set of coarse correlated equilibria. Variants of the presented regret matching could also ensure convergence to the set of correlated equilibrium, which is a more rigid solution concept than presented in Definition 3. We direct the readers to Hart and Mas-Colell (2000) and Young (2004) for the details associated with this variation.

4.4.1 Equilibrium Selection of Correlated Equilibrium

The set of correlated equilibria is much larger than the set of Nash equilibria and can potentially be exploited to provide systems with better performance guarantees. One example of such a system is the Shapley game, which is a two-player game with utility functions of the form

		Agent 2 *A*	Agent 2 *B*	Agent 2 *C*
Agent 1	*A*	0, 0	0, 1	1, 0
	B	1, 0	0, 0	0, 1
	C	0, 1	1, 0	0, 0

Payoff Matrix

There are no pure Nash equilibria in this game and the unique (mixed) Nash equilibrium is when each agent i employs a strategy $p_i = (1/3, 1/3, 1/3)$, which yields an expected payoff of $1/3$ to each agent. However, there is also a coarse correlated equilibrium where the distribution z has a value $1/6$ on each of the six joint actions where some agent receives nonzero payoff; z has a value 0 for the other three joint actions. This coarse correlated equilibrium yields an expected utility of $1/2$ to each agent and is clearly more desirable. One could easily imagine other scenarios, e.g., team versus team games, where specific coarse correlated equilibrium could provide significant performance improvements over any Nash equilibrium.

The problem with regret matching for exploiting this potential opportunity is that behavior is not guaranteed to converge to any specific coarse correlated equilibrium. Accordingly, the efficiency guarantees associated with coarse correlated equilibria cannot be better than the efficiency bounds associated with pure Nash equilibria and can often be quite worse. With this issue in mind, recent work in Marden (2015) and Borowski et al. (2014) has sought to develop learning algorithms that converge to the efficient coarse correlated equilibrium, where efficiency is measured by the sum of the agents' expected utilities. Here, the algorithm introduced in Marden (2015) ensures that the empirical frequency of play will converge to the most efficient coarse correlated equilibrium, while Borowski et al. (2014) provides an algorithm that guarantees that the day-to-day behavior of the agents will converge to the most efficient correlated equilibrium. Both of these algorithms view convergence in a stochastic stability sense.

The motivation for these developments centers on the fact that joint randomization, which can potentially be characterized by correlated equilibria, can be key to providing desirable system-level behavior. One example of such a system is a peer-to-peer file sharing system where users engage in interactions with other users to transfer files of interest and satisfy demands (Wang et al. 2009). Here, Wang et al. (2009) demonstrates that the optimal system performance is actually characterized by the most efficient correlated equilibrium as defined above. Another example of such a system is the problem of access control for wireless communications, where there are a collection of mobile terminals that compete over access to a common channel (Altman et al. 2006). Optimizing system throughput requires a level of correlation between the transmission strategies of the mobiles so as to minimize the chance of simultaneous transmissions and failures. The authors in Altman et al. (2006) study the efficiency of correlated equilibria in this context. Identifying the role of correlated equilibrium (and learning strategies for attaining specific correlated equilibrium) warrants further research attention.

5 Conclusion

The goal of this chapter has been to highlight a potential role of game-theoretic learning in the design of networked control systems. We reviewed several classes of learning algorithms accentuating their performance guarantees and reliance on game structures.

It is important to reemphasize that game-theoretic learning represents just a single dimension of a game-theoretic control design. The other dimension centers on the assignment of objective functions to the individual agents. The structure of these agent objective functions not only dictate convergence guarantees associated with various game-theoretic learning algorithms but can also be exploited to characterize the efficiency of the resulting behavior. To that end, consider the assignment of agent objective functions that yields a potential game and has a given price of anarchy. Marrying this design with a learning algorithm that guarantees convergence to a pure Nash equilibrium in potential games yields a game-theoretic control design that ensures that the collective behavior will converge to a specific allocation (in particular a Nash equilibrium associated with the designed agent objective functions) and the efficiency of this allocation will be in line with the given price of anarchy.

Taking full advantage of this game-theoretic approach requires assigning agent objective functions that yield a potential game and optimize the price of anarchy over all such objective functions. Unfortunately, the existing literature provides no mechanism for accomplishing this goal as utility design for distributed engineering systems is currently not well understood. A reason for this gap is that agent objective function are traditionally modeled to reflect agent preferences in a given social system, e.g., a reasonable objective for drivers on a transportation network is minimizing experienced congestion. Hence, efficiency measures in games, such as the price of anarchy, are traditionally viewed from an analysis perspective with virtually no design component. Reversing this trend and deriving systematic methodologies for utility design in multiagent systems represents a significant opportunity for game-theoretic control moving forward.

References

Alos-Ferrer C, Netzer N (2010) The logit-response dynamics. Games Econ Behav 68:413–427

Altman E, Bonneau N, Debbah M (2006) Correlated equilibrium in access control for wireless communications. In: 5th International Conference on Networking

Babichenko Y (2012) Completely uncoupled dynamics and Nash equilibria. Games Econ Behav 76:1–14

Blondel VD, Hendrickx JM, Olshevsky A, Tsitsiklis JN (2005a) Convergence in multiagent coordination, consensus, and flocking. In: IEEE Conference on Decision and Control, pp 2996–3000

Blondel VD, Hendrickx JM, Olshevsky A, Tsitsiklis JN (2005b) Convergence in multiagent coordination, consensus, and flocking. In: Proceedings of the Joint 44th IEEE Conference on Decision and Control and European Control Conference (CDC-ECC'05), Seville

Blume L (1993) The statistical mechanics of strategic interaction. Games Econ Behav 5:387–424

Blume L (1997) Population games. In: Arthur B, Durlauf S, and Lane D (eds) The economy as an evolving complex system II. Addison-Wesley, Reading, pp 425–460

Borowski H, Marden JR, Frew EW (2013) Fast convergence in semi-anonymous potential games. In: Proceedings of the IEEE Conference on Decision and Control, pp 2418–2423

Borowski HP, Marden JR, Shamma JS (2014) Learning efficient correlated equilibria. In: Proceedings of the IEEE Conference on Decision and Control, pp 6836–6841

Cortes J, Martinez S, Karatas T, Bullo F (2002) Coverage control for mobile sensing networks. In: Proceedings of the IEEE International Conference on Robotics and Automation (ICRA'02), Washington, DC, pp 1327–1332

Daskalakis C, Goldberg PW, Papadimitriou CH (2009) The complexity of computing a Nash equilibrium. SIAM J Comput 39(1):195–259

Fudenberg D, Levine DK (1995) Consistency and cautious fictitious play. Games Econ Behav 19:1065–1089

Fudenberg D, Levine DK (1998) The theory of learning in games. MIT Press, Cambridge

Fudenberg D, Tirole J (1991) Game theory. MIT Press, Cambridge

Gopalakrishnan R, Marden JR, Wierman A (2014) Potential games are necessary to ensure pure Nash equilibria in cost sharing games. Math Oper Res 39(4):1252–1296

Hart S (2005) Adaptive heuristics. Econometrica 73(5):1401–1430

Hart S, Mansour Y (2010) How long to equilibrium? The communication complexity of uncoupled equilibrium procedures. Games Econ Behav 69(1):107–126

Hart S, Mas-Colell A (2000) A simple adaptive procedure leading to correlated equilibrium. Econometrica 68(5):1127–1150

Hart S, Mas-Colell A (2003) Uncoupled dynamics do not lead to Nash equilibrium. Am Econ Rev 93(5):1830–1836

Jadbabaie A, Lin J, Morse AS (2003) Coordination of groups of mobile autonomous agents using nearest neighbor rules. IEEE Trans Autom Control 48(6):988–1001

Kearns MJ, Littman ML, Singh SP (2001) Graphical models for game theory. In: Proceedings of the 17th Conference in Uncertainty in Artificial Intelligence, pp 253–260
Lambert III TJ, Epelman MA, Smith RL (2005) A fictitious play approach to large-scale optimization. Oper Res 53(3):477–489
Marden JR (2012) State based potential games. Automatica 48:3075–3088
Marden JR (2015) Selecting efficient correlated equilibria through distributed learning. In: American Control Conference, pp 4048–4053
Marden JR, Shamma JS (2012) Revisiting log-linear learning: asynchrony, completeness and a payoff-based implementation. Games Econ Behav 75(2):788–808
Marden JR, Shamma JS (2015) Game theory and distributed control. In: Young HP, Zamir S (eds) Handbook of game theory with economic applications, vol 4. Elsevier Science, pp 861–899
Marden JR, Wierman A (2013) Distributed welfare games. Oper Res 61:155–168
Marden JR, Arslan G, Shamma JS (2009) Joint strategy fictitious play with inertia for potential games. IEEE Trans Autom Control 54:208–220
Martinez S, Cortes J, Bullo F (2007) Motion coordination with distributed information. Control Syst Mag 27(4):75–88
Monderer D, Shapley LS (1996) Fictitious play property for games with identical interests. J Econ Theory 68:258–265
Montanari A, Saberi A (2009) Convergence to equilibrium in local interaction games. In: 50th Annual IEEE Symposium on Foundations of Computer Science
Murphey RA (1999) Target-based weapon target assignment problems. In: Pardalos PM, Pitsoulis LS (eds) Nonlinear assignment problems: algorithms and applications. Kluwer Academic, Alexandra
Nisan N, Roughgarden T, Tardos E, Vazirani VV (2007) Algorithmic game theory. Cambridge University Press, New York
Olfati-Saber R (2006) Flocking for multi-agent dynamic systems: algorithms and theory. IEEE Trans Autom Control 51:401–420
Olfati-Saber R, Murray RM (2003) Consensus problems in networks of agents with switching topology and time-delays. IEEE Trans Autom Control 49(9):1520–1533
Olfati-Saber R, Fax JA, Murray RM (2007) Consensus and cooperation in networked multi-agent systems. Proc IEEE 95(1):215–233
Roughgarden T (2005) Selfish routing and the price of anarchy. MIT Press, Cambridge
Roughgarden T (2015) Intrinsic robustness of the price of anarchy. J ACM 62(5):32:1–32:42
Shah D, Shin J (2010) Dynamics in congestion games. In: ACM SIGMETRICS, pp 107–118
Shamma JS (2014) Learning in games. In: Baillieul J, Samad T (eds) Encyclopedia of Systems and Control. Springer, London
Shoham Y, Powers R, Grenager T (2007) If multi-agent learning is the answer, what is the question? Artif Intell 171(7):365–377
Touri B, Nedic A (2011) On ergodicity, infinite flow, and consensus in random models. IEEE Trans Autom Control 56(7):1593–1605
Tsitsiklis JN (1987) Decentralized detection by a large number of sensors. Technical report. MIT, LIDS
Tsitsiklis JN, Bertsekas DP, Athans M (1986) Distributed asynchronous deterministic and stochastic gradient optimization algorithms. IEEE Trans Autom Control 35(9):803–812
Vetta A, "Nash equilibria in competitive societies, with applications to facility location, traffic routing and auctions," Proceedings of the 43rd Annual IEEE Symposium on Foundations of Computer Science, 2002, pp 416–425
Wang B, Han Z, Liu KJR (2009) Peer-to-peer file sharing game using correlated equilibrium. In: 43rd Annual Conference on Information Sciences and Systems, CISS 2009, pp 729–734
Wolpert D, Tumor K (1999) An overview of collective intelligence. In: Bradshaw JM (ed) Handbook of Agent Technology. AAAI Press/MIT Press, Cambridge, USA
Young HP (1998) Individual strategy and social structure. Princeton University Press, Princeton
Young HP (2004) Strategic learning and its limits. Oxford University Press, New York
Zhu M, Martínez S (2013) Distributed coverage games for energy-aware mobile sensor networks. SIAM J Control Optim 51(1):1–27

Network Games

12

S. Rasoul Etesami and Tamer Başar

Contents

Abstract

This chapter provides a general overview of the topic of network games, its application in a number of areas, and recent advances, by focusing on four

S. R. Etesami (✉)
Department of Industrial and Systems Engineering, University of Illinois at Urbana-Champaign, Urbana, IL, USA
e-mail: etesami1@illinois.edu

T. Başar
Coordinated Science Laboratory and Department of Electrical and Computer Engineering, University of Illinois at Urbana-Champaign, Urbana, IL, USA
e-mail: basar1@illinois.edu

T. Başar, G. Zaccour (eds.), *Handbook of Dynamic Game Theory*,
https://doi.org/10.1007/978-3-319-44374-4_10

major types of games, namely, congestion games, resource allocation games, diffusion games, and network formation games. Several algorithmic aspects and methodologies for analyzing such games are discussed, and connections between network games and other relevant topical areas are identified.

Keywords

Network games · Congestion games · Diffusion games · Potential games · Resource allocation · Network formation · Nash equilibrium · Price of anarchy

1 Introduction

Network games broadly refer to a class of games where the players and/or their actions are linked through an underlying network structure. Here, the network can capture a wide range of applications such as communication links, social relations, or even transportation roads. In many applications the network structure is not necessarily fixed and may dynamically change based on players' interactions. As an example, in certain social events such as political elections, individuals can be viewed as players whose individual goals are to vote (take an action) so as to select their favorite candidates. However, the vote of each player is a consequence of his[1] social interactions with his friends. Such social ties define the network structure among the players with an edge (or a link) between two players if and only if they are each other's friends. Once the players take their actions, it is possible that friends drift away from each other (as they vote differently and have less in common), or strangers become each other's friends (as now they have the same opinion about the candidates, which is a good starting point for friendship). As it is evident in this example, one of the key features of network games is the role of the network on the structure of the game which puts additional constraints on players' interactions; it determines which players can have direct influence on the payoff, and hence decision of a particular player.

Typically networks represent the interconnections between the players; however, there are many scenarios where the network describes the underlying structure on players' actions (rather than their payoffs directly). As an example, in a chess game, the players are not part of the network; however, at each possible move, their admissible actions can be described using an 8×8 lattice. In this chapter we will consider network games of both types and discuss several methods for analyzing them. One typical approach for studying network games is to translate them into the *normal* or *extensive* form by considering the set of all possible pure strategies which can be realized by players given the network constraints. While in some cases this approach can be helpful, in most cases such a representation is neither practical nor efficient.

[1]Throughout this chapter, we refer to players as "he," "she," or "it" ("his," "her," or "its") interchangeably, somewhat context dependent.

This is because in practice, the number of parameters that need to be specified for the normal or extensive representation grows exponentially with the size of the game, i.e., number of players, pure strategies, etc. In addition, games in normal form usually fail to capture directly and in a transparent way the network structure that is present in the strategic interactions. Such limitations mandate developing new tools and techniques which are more suitable for analyzing network games and yet keep the representation of the game as succinct and transparent as possible.

Network games have been extensively studied in the past literature and under various settings. To name a few, we mention here *congestion games* (Ackermann et al. 2008; Anantharam 2004; Awerbuch et al. 2008; Fabrikant et al. 2004; Milchtaich 1996; Rosenthal 1973; Tekin et al. 2012) with a wide range of applications in selfish routing (Blum et al. 2006; Correa et al. 2004; Menache and Ozdaglar 2011; Roughgarden 2002) where the network traffic congestion is selfishly controlled by vehicle owners who seek to minimize their travel costs; *caching and resource allocation games* (Chun et al. 2004; Etesami and Başar 2017a,b; Gopalakrishnan et al. 2012; Laoutaris et al. 2006; Maheswaran and Başar 2001; Marden and Roughgarden 2014; Menache and Ozdaglar 2011; Pacifici and Dan 2012; Pollatos et al. 2008), where a set of agents compete for the same set of resources over a network; and *cost sharing games*, where the objective is to share the costs among agents (e.g., the costs of creating a network or serving a market) while incentivizing them to cooperate despite their self-interests (Goemans et al. 2006; Nisan et al. 2007; Roughgarden and Schrijvers 2016).

Furthermore, network games have also been studied under the framework of *graphical games* where the players are located on the nodes of an underlying network and their payoffs are determined based on their own actions and neighbors' (Daskalakis and Papadimitriou 2006; Kearns et al. 2001; Nisan et al. 2007). *Diffusion games* constitute another type of well-studied network games in which the players' objective is to propagate a certain type of product or behavior in a desired way through the network (Alon et al. 2010; Goyal et al. 2014; Jackson and Yariv 2007; Montanari and Saberi 2010; Singer 2012; Small and Mason 2013; Tzoumas et al. 2012; Young 2006), with applications in product advertising, immunization, and virus spreading. Moreover, network formation games provide a popular framework for studying social and economic networks in order to understand how real-world networks (such as the Internet) develop when multiple selfish independent agents (e.g., ISPs) build pieces of the network to improve their own objective functions (Alon et al. 2013; Demaine et al. 2007; Fabrikant et al. 2003; Goyal and Vega-Redondo 2000; Jackson and Watts 2002; Kawald and Lenzner 2013). In addition, network games have proven quite useful for studying cyber-security networks (Alpcan and Başar 2010; Grossklags et al. 2008; Liang and Xiao 2012), opinion dynamics in social networks (Etesami and Başar 2015; Gionis et al. 2013), and distributed control (Li and Marden 2013), among many others (Dürr and Thang 2007; Eksin et al. 2013; Galeotti et al. 2010; Panagopoulou and Spirakis 2008).

As the class of network games covers a wide range of topics, in this chapter we mainly focus on four well-studied types of such games by presenting several key results in each area. More specifically, we consider in the chapter congestion games, resource allocation games, diffusion games, and network formation games and present some of the existing results from both a formulation side and an algorithmic aspect. We have left out of this overview a number of topics which would naturally fall under the heading of *network games*, the reason being that several of them have been covered exclusively in some other chapters of this *Handbook*. Four examples of *routing games* (Menache and Ozdaglar 2011; Roughgarden 2002), which would naturally fall under network games, and specifically under congestion games, have been discussed extensively in Chaps. 25 and 26 of the *Handbook* (Krichene et al. 2017; Shakkottai and Srikant 2017), including the Pigou network, Wardrop equilibrium, Braess paradox, flow control games, and Stackelberg equilibria in routing games. We will therefore not be covering these topics in this chapter. Some other topics in network games, not covered by the four main types this chapter is devoted to, are briefly mentioned in a section toward the end of the chapter.

2 Congestion Games

Congestion games are arguably one of the most studied types of network games first introduced by Rosenthal (1973). In the simplest form, congestion games are composed of a set of players and resources such that the payoff to a player depends on his choice of resource and the number of players that choose that specific resource. Since then, congestion games have been extensively used for modeling strategic problems where a set of players compete for the same set of resources such that the players' costs for using the same resource monotonically increase with the congestion on that resource. A simple example for such a scenario is the cost of traffic congestion for vehicle drivers (players), where in this case the resources can be interpreted as roads of the traffic network. As more vehicles decide to use a specific road, the congestion cost for the vehicles in that road increases.[2] In what follows, we formally introduce the class of congestion games and their specification to network congestion games.

Definition 1 (Congestion game). Let $\mathcal{E}$ be a finite set of resources and consider a set of n players, with action sets $\mathcal{A}_i \subseteq 2^{\mathcal{E}}, i \in [n] := \{1, 2, \dots, n\}$. For every $e \in \mathcal{E}$, a delay function $d_e(\cdot) : [n] \to \mathbb{Z}$ is a nondecreasing integer-valued function. Given an action profile $\boldsymbol{a} := (a_1, \dots, a_n) \in \mathcal{A}_1 \times \cdots \times \mathcal{A}_n$, let $n_e(\boldsymbol{a}) = |\{i \in [n] : e \in a_i\}|$ be the number of players in the action profile $\boldsymbol{a}$ who choose the resource e. The cost of player i for the action profile $\boldsymbol{a}$ is given by $c_i(\boldsymbol{a}) = \sum_{e \in a_i} d_e(n_e(\boldsymbol{a}))$.

In other words, the total delay (cost) that player i incurs is the sum of delays of resources used by player i, where the delay of a resource depends on the congestion

[2] Another example arises in the transmission of pockets in communication networks, where because of fixed bandwidth, an increase in the rate leads to increase in transmission time (i.e., delay) and thus increase in cost.

$n_e(\boldsymbol{a})$, the total number of players using that resource. The above formulation provides a general framework for modeling a wide range of problems as long as the players' costs depend only on the congestion in each resource and not on the specific type of players choosing that resource. In particular, such formulation can be easily adapted to incorporate the role of network structure into the game as is described in the following example.

Example 1 (Network congestion games). Let $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ be a directed network with vertex set $\mathcal{V}$ and edge set $\mathcal{E}$, such that each player $i \in [n]$ is associated with a pair of nodes $(s_i, t_i) \in \mathcal{V} \times \mathcal{V}$ who aims to travel from the source node s_i to the terminal node t_i. Therefore, the action set $\mathcal{A}_i$ for player i is given by all possible paths from s_i to t_i. Note that each path in $\mathcal{G}$ can be viewed as a subset of edges in $\mathcal{E}$, which means that one can view the resources in the network congestion game to be the set of all the edges, $\mathcal{E}$. Again by assuming proper delay functions on the edges (resources), one can view the network congestion game as a special case of the congestion game given in Definition 1, in which each player wants to choose a path which has the least traffic congestion along its edges.

In fact, congestion games (and in particular network congestion games) feature many nice properties, perhaps the most remarkable one being that they admit at least one *pure-strategy Nash equilibrium* (NE),[3] which is mainly due to their structure. More generally, congestion games are known to belong to the class of *exact potential* games which are guaranteed to have at least one pure-strategy Nash equilibrium (Monderer and Shapley 1996). Intuitively, a game is said to be an exact potential game if the incentive of all players to change (improve) their strategy can be captured through a single global function called the *potential function*.

Definition 2. Given an n-player game $\Gamma = ([n], \{\mathcal{A}_i\}, \{c_i\})$, let $\mathcal{A} = \mathcal{A}_1 \times \cdots \times \mathcal{A}_n$ be the set of action profiles of the players. The game is an exact potential game if there is an exact potential function $\Phi : \mathcal{A} \to \mathbb{R}$ such that $\forall a_{-i} \in \mathcal{A}_{-i}, \forall a_i, \hat{a}_i \in \mathcal{A}_i$ we have $\Phi(a_i, a_{-i}) - \Phi(\hat{a}_i, a_{-i}) = c_i(a_i, a_{-i}) - c_i(\hat{a}_i, a_{-i}), \forall i \in [n]$.

Note that based on Definition 2, any exact potential game admits at least one pure-strategy NE. This is simply because any minimizer of the exact potential function $\Phi(\cdot)$ delivers an action profile in which no player can decrease its cost further (otherwise that profile is not a minimizer anymore). It was first shown by Rosenthal (1973) that congestion games always admit an exact potential function and hence a pure-strategy NE. To see why this is true, let us consider the following potential function:

[3] A pure-strategy Nash equilibrium is an action profile from which no player has a unilateral incentive to change his strategy.

$$\Phi(\boldsymbol{a}) = \sum_{e \in \mathcal{E}} \sum_{j=1}^{n_e(\boldsymbol{a})} d_e(j).$$

Note that $\Phi(\boldsymbol{a})$ has no intuitive interpretation as "social cost"; it just accurately captures the definition of an exact potential function. To see this more clearly, let us assume that player i changes its action from $a_i \subseteq \mathcal{E}$ to $\hat{a}_i \subseteq \mathcal{E}$, while all other players keep their actions unchanged. We can write

$$\begin{aligned}\Phi(a_i, \boldsymbol{a}_{-i}) - \Phi(\hat{a}_i, \boldsymbol{a}_{-i}) = &\sum_{e \in a_i \setminus \hat{a}_i} \sum_{j=1}^{n_e(\boldsymbol{a})} d_e(j) + \sum_{e \in \hat{a}_i \setminus a_i} \sum_{j=1}^{n_e(\boldsymbol{a})} d_e(j) \\ &- \sum_{e \in a_i \setminus \hat{a}_i} \sum_{j=1}^{n_e(\hat{a}_i, \boldsymbol{a}_{-i})} d_e(j) - \sum_{e \in \hat{a}_i \setminus a_i} \sum_{j=1}^{n_e(\hat{a}_i, \boldsymbol{a}_{-i})} d_e(j), \end{aligned} \tag{12.1}$$

where the equality is due the fact that if a resource e belongs to both or neither of the actions $a_i, \hat{a}_i$, then its contributions to $\Phi(a_i, \boldsymbol{a}_{-i})$ and $\Phi(\hat{a}_i, \boldsymbol{a}_{-i})$ are the same. Moreover, $e \in a_i \setminus \hat{a}_i$ implies $n_e(\hat{a}_i, \boldsymbol{a}_{-i}) = n_e(\boldsymbol{a}) - 1$, and $e \in \hat{a}_i \setminus a_i$ implies $n_e(\hat{a}_i, \boldsymbol{a}_{-i}) = n_e(\boldsymbol{a}) + 1$. Substituting these two relations into (12.1), and simplifying the terms, we get

$$\begin{aligned}\Phi(a_i, \boldsymbol{a}_{-i}) - \Phi(\hat{a}_i, \boldsymbol{a}_{-i}) &= \sum_{e \in a_i \setminus \hat{a}_i} d_e(n_e(\boldsymbol{a})) - \sum_{e \in \hat{a}_i \setminus a_i} d_e(n_e(\boldsymbol{a}) + 1) \\ &= \sum_{e \in a_i \setminus \hat{a}_i} d_e(n_e(\boldsymbol{a})) - \sum_{e \in \hat{a}_i \setminus a_i} d_e(n_e(\hat{a}_i, \boldsymbol{a}_{-i})) \\ &= c_i(a_i, \boldsymbol{a}_{-i}) - c_i(\hat{a}_i, \boldsymbol{a}_{-i}),\end{aligned}$$

where the last equality is due to the definition of cost functions in the congestion game (Definition 1). This shows that $\Phi(\cdot)$ is an exact potential function for the congestion game, and hence it admits a pure-strategy NE.

In their seminal work, Monderer and Shapley (1996) proved the converse of the above result; namely, they proved that for any exact potential game, there is a congestion game with the same exact potential function.

Theorem 1. *Any exact potential game is isomorphic to a congestion game.*

The main idea behind establishing the result of Theorem 1 is to show that for a game with an exact potential function $\Phi(\cdot)$, one can construct an equivalent congestion game in which the players are the same, and the resources are interpreted as all possible subsets of actions which can be taken jointly by the players in the original exact potential game. Then, one can leverage the potential function $\Phi(\cdot)$ to construct well-defined delay and cost functions for the equivalent congestion game. More details can be found in Monderer and Shapley (1996).

2.1 Computing Pure-Strategy NE in Congestion Games

As we saw earlier, congestion games are guaranteed to have a pure-strategy NE. Therefore, one of the immediate questions is how to compute one of these equilibrium points efficiently. One way of doing that is to let the players in the congestion game reduce their costs by playing better actions in response to the most recent actions picked, and in some arbitrary order. Since a reduction in the cost of any player results in the same reduction in the value of the potential function $\Phi(\cdot)$ (which is nonnegative and bounded above), this process must eventually terminate in finite time to an action profile which is necessarily a pure-strategy NE. Otherwise, there must exist at least one player who can still reduce his cost and, hence, results in further reduction in the potential function which is a contradiction. Such a kind of argument where after finitely many improvements in the players' utilities (reduction in their cost) the sequence of action profiles will terminate to a pure-strategy NE is known as the *finite improvement path* property. Therefore, congestion games (or equivalently exact potential games) with finite action space possess the finite improvement path property, which can be used to find or approximate their pure-strategy NE points.

An important question now is to see whether the number of steps in a finite improvement path efficiently scales with the parameters of the game. In other words, it might be possible that the length of such improvement path can be exponentially large in terms of game parameters, which renders the applicability of such a method for finding pure-strategy NE points. Unfortunately, congestion games in their general form of Definition 1 can attain exponentially long improvement paths to their NE points, and the complexity of finding a pure-strategy NE in general congestion games is PLS complete[4] (Fabrikant et al. 2004). This means in plain terms that finding a pure-strategy NE in general congestion games is "as hard to compute as any object whose existence is guaranteed by a potential function". However, if we restrict our attention to a special subclass of congestion games, namely, *symmetric network congestion* games, then a pure-strategy NE can be found in polynomial time.

Definition 3. A *symmetric network congestion* game is a network congestion game (see Example 1) in which all the players have the same source-terminal pairs, i.e., $\exists\, (s,t) \in \mathcal{V} \times \mathcal{V}$ such that $(s_i, t_i) = (s,t), \forall i \in [n]$.

[4]Polynomial local search (PLS) is a complexity class that models the difficulty of finding a locally optimal solution to an optimization problem. A PLS-complete problem refers to a "hardest" problem in this complexity class.

Note that from Definition 3, it immediately follows that in the symmetric network congestion game, all the players share the same action set which is the edge set of all possible paths from source s to terminal t. Before we show how one can obtain a pure-strategy NE of the symmetric network congestion game, we first consider the following network optimization problem known as *min-cost flow* problem.

Min-cost Flow: In the min-cost flow problem, we are given a directed network $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, with a source $s \in \mathcal{V}$ and a terminal $t \in \mathcal{V}$, where each edge $e = (u, v) \in \mathcal{E}$ has a capacity $c(u, v) > 0$. A feasible flow of capacity $l > 0$ from the source s to the terminal t is a nonnegative vector $f : \mathcal{E} \to \mathbb{R}^{\geq 0}$ with support $\mathcal{E}$, which satisfies the following flow constraints:

- capacity constraints: $f(u, v) \leq c(u, v)$
- flow conservation: $\sum_{v \in \mathcal{V}} f(u, v) = 0, \forall u \neq s, t$
- skew symmetry: $f(u, v) = -f(v, u)$
- required flow: $\sum_{v \in \mathcal{V}} f(s, v) = \sum_{v \in \mathcal{V}} f(v, t) = l$.

For an edge (u, v) we let $b(u, v)$ be that edge flow cost and $\sum_{(u,v) \in \mathcal{E}} b(u, v) f(u, v)$ be the total flow cost. The objective of the min-cost flow problem is to find a feasible flow f with capacity l from s to t which has the minimum total cost. An important property of the min-cost-flow problem is that if edge capacities $c(u, v), (u, v) \in \mathcal{E}$ and flow capacity l are integer valued, then the min-cost flow problem admits an optimal integer flow $f^* : \mathcal{E} \to \mathbb{Z}^{\geq 0}$ which can be obtained efficiently in polynomial time. Using this property of the min-cost flow problem, the following theorem can be established.

Theorem 2. *There is a polynomial algorithm for finding a pure-strategy NE in symmetric network congestion games.*

This result is established by showing that for the symmetric network congestion games, finding an action profile which minimizes the potential function $\Phi(\boldsymbol{a}) = \sum_{e \in \mathcal{E}} \sum_{j=1}^{n_e(\boldsymbol{a})} d_e(j)$ (and hence constitutes a pure-strategy NE) can be reduced to solving the min-cost flow problem in polynomial time. Let $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ be the graph of a symmetric network congestion game with n players, source-terminal pair (s, t), and edge delay functions $\{d_e\}_{e \in \mathcal{E}}$. We replace each edge e in $\mathcal{G}$ by n parallel edges between the same nodes, each with capacity 1 and with flow costs $d_e(1), d_e(2), \ldots, d_e(n)$. We then claim that the optimal integer flow f^* with capacity n for this expanded network is a minimizer of the potential function $\Phi(\cdot)$. To see this clearly, let f^* be the optimal integer flow with capacity n. Since each edge of the new network has capacity 1, f^* can be split into n distinct routes from s to t. Consider every route as the action of a single player in the original symmetric network congestion game, and define the action profile $\boldsymbol{a}^*$ as some ordered collection of these routes (due to symmetry of the players, the order does not matter). Thus, $n_e(\boldsymbol{a}^*)$ represents the number of edges $e_1, e_2, \ldots, e_n$ on the new network which represent edge e on $\mathcal{G}$ and are used in the flow f^*. Since f^* is minimal, it must first use the cheapest edges in which case the contribution of the edges $e_1, \ldots, e_n$ to the total cost of f^* equals $\sum_{j=1}^{n_e(\boldsymbol{a}^*)} d_e(j)$. Therefore, the total cost of f^* equals $\Phi(\boldsymbol{a}^*) = \sum_{e\in\mathcal{E}} \sum_{j=1}^{n_e(\boldsymbol{a}^*)} d_e(j)$, which minimizes $\Phi(\cdot)$ (otherwise, the flow corresponding to minimizer of $\Phi(\cdot)$ will have a lower cost than f^*).

Theorem 2 provides a positive result on how to find a pure-strategy NE in a special class of congestion games, namely, symmetric network congestion games. Unfortunately, this class constitutes a very small subclass of congestion games, where the latter itself is a small subclass of games which admit pure-strategy NE. In fact, it has been shown in Fabrikant et al. (2004) that it is PLS complete to find a pure-strategy NE in:

- general congestion games,
- symmetric congestion games where the players share the same action sets,
- asymmetric network congestion games where the players have different pairs of source-terminal nodes (Example 1).

This again leaves open the question of existence of a polynomial time algorithm for finding a pure-strategy NE in congestion games.

Finally, we mention here that one can consider different variants of congestion games by modifying some of their underlying assumptions. As an example, in the definition of congestion games (Definition 1), it is assumed that players share the same delay functions. However, one can relax this assumption by introducing "player-specific" congestion games where the delay functions not only depend on resources but also are player specific. It was shown in Milchtaich (1996) that while such player-specific congestion games with identical action sets do not admit an exact potential function, they still possess a pure-strategy NE which can be constructed inductively in polynomial time.

2.2 Application of Congestion Games in Market Sharing

In this subsection, we provide an application of congestion games in market sharing. Broadly speaking, in market sharing games, there is a set of agents who want to provide service to their customers with a limited set of resources, while there are different request rates for different resources and providers. In this regard, one of the main challenges is to determine whether the outcome of interactions between service providers modeled as a repeated game converges to any stable outcome such as a Nash equilibrium and how long it will take to converge. In this subsection we consider a special form of market sharing game, first introduced by Goemans et al. (2006):

Market Sharing Game: Consider a market sharing game in which there are a set $[n]$ of n players each having a limited amount of budget $B_i, i \in [n]$, and a set $\mathcal{H}$ of

m markets. Each market $j \in \mathcal{H}$ is characterized by two parameters: q_j which is the query rate that market j is requested by its customers and C_j which is the cost of servicing that market. The connection between markets and players is captured by an undirected *bipartite* graph $\mathcal{G} = ([n] \cup \mathcal{H}, \mathcal{E})$, in which an edge $\{i, j\} \in \mathcal{E}$ between player $i \in [n]$ and market $j \in \mathcal{E}$ shows that player i is interested in (serving) market j. Each player i must decide what subset of markets to serve subject to its budget and network constraints. Therefore, a feasible action for player i, denoted by a_i, is a subset of markets that player i is interested in, so that $\sum_{j \in a_i} C_j \leq B_i$. We denote the set of all feasible actions of player i by $\mathcal{A}_i$ and the set of feasible action profiles of all the players by $\mathcal{A} = \mathcal{A}_1 \times \cdots \times \mathcal{A}_n$. Finally, given an action profile $\boldsymbol{a} \in \mathcal{A}$, the payoff received by player i is given by

$$u_i(\boldsymbol{a}) = \sum_{j \in a_i} \frac{q_j}{n_j}, \tag{12.2}$$

where n_j denotes the number of players in $\boldsymbol{a}$ who serve market j. The rationale behind such a payoff function is that all the players who are serving market j will share the total value of that market q_j equally among themselves. In this game, each agent wants to selfishly maximize its own payoff. An immediate consequence of this formulation is the following theorem.

Theorem 3. *Market sharing game is a special case of congestion games and hence admits a pure-strategy NE.*

The above theorem simply holds by noting that the Rosenthal potential function $\Phi(\boldsymbol{a}) = \sum_{j=1}^{m} \sum_{k=1}^{n_j} \frac{q_j}{k}$ adapted for the market sharing game serves as an exact potential function for this game. In fact, in the market sharing game, the market queries $\{q_j,\ j \in \mathcal{H}\}$ can be viewed as the resources, and the delay function associated with resource q_j is given by $d_{q_j}(k) = -\frac{q_j}{k}$ (here we are using minus since we are working with utilities rather than costs).

In practice, finding the best response for a player in the market sharing game is an NP-hard problem. This is because, given the set of actions of other players, a_{-i}, the best action of player i can be obtained by solving a knapsack problem where the value of market (item) j is equal to $\frac{q_j}{n_j}$ or $\frac{q_j}{n_j+1}$ depending on whether market j is currently being serviced by player i or not. The size of j in the knapsack instance is C_j, and the knapsack capacity equals to B_i. Therefore, even in the simplest case when we have only one player, finding a pure-strategy NE is equivalent to solving a general knapsack problem, which is NP-hard (although the existence of such a NE is guaranteed by Theorem 3). However, it turns out that if we restrict our attention to a special case of market sharing games where all markets have equal cost, then finding the best response action for each player reduces to a simple optimization problem.

Definition 4. The above game is a uniform market sharing game if all the markets have a uniform cost $C_j = C, \forall j \in \mathcal{H}$.

It is easy to see that in the uniform market sharing game, given the actions of other players, a_{-i}, player i can find its best response by simply serving the highest $\lfloor \frac{B_i}{C} \rfloor$ rewarding markets. In this case, one can find a pure-strategy NE of the uniform market sharing game efficiently as stated in the following theorem:

Theorem 4. *In the uniform market sharing game, a pure-strategy NE can be found after at most $m^2 n$ steps, where m is the total number of markets and n is the number of players.*

The idea behind proving Theorem 4 is to iteratively add and exchange markets to the action sets of players until a pure-strategy NE is achieved. This process is based on several rounds. The first stage of each round corresponds to adding only one new market to the action set of one player. The round then continues by a sequence of exchange stages where at each stage, a player exchanges only one market in his action set with another one in order to improve his payoff. A round ends when no player can further improve its payoff by this exchanging operation. More precisely:

- Start from the empty action profile where $a_i = \emptyset, \forall i \in [n]$.
- At the beginning of each round, select a player i who can serve an additional market, and let him add *only* one new market to his action set, i.e., player i changes his action from a_i to $a_i \cup \{j\}$, for some market $j \notin a_i$.
- After the adding stage at the very first stage of a round, the exchange stages start. At each exchange step, one player i updates his action by evicting a market $j \in a_i$ and inserting another market of his interest $k \notin a_i$ in order to maximize his payoff. Hence his action set will change from a_i to $a_i \setminus \{j\} \cup \{k\}$. The round ends when no further exchange step is possible.

Following these steps one can see that when neither adding nor exchanging stage is possible, the final action profile must be a pure-strategy NE, This implicitly uses the fact that we are dealing with a *uniform* market sharing game, and therefore any maximal action for a player can be obtained from any other maximal action by exchanging in and out two markets at a time. To bound the number of stages in the above iterative process, we note that since one player adds a market at the beginning of each round, the number of rounds cannot be greater than nm. Finally, one can show that each round can have at most $m - 1$ exchange stages which together with the initial adding stage implies that the length of each round is at most m. Thus, the total number of stages until the whole process terminates to a pure-strategy NE is at most $m^2 n$.

Now that existence of a pure-strategy NE in the market sharing game is guaranteed, we next turn our attention to "efficiency" of such equilibrium points.

In other words, we want to see how much lack of coordination among the players degrades the social optimality of the system. In this regard, one of the widely used metrics to measure the inefficiency of the Nash equilibrium points is known as the *price of anarchy* (PoA) which is the ratio of the optimal social welfare to the social welfare of the "worst" pure-strategy NE (Koutsoupias and Papadimitriou 1999). More precisely:

Definition 5. Given an n-player game with at least one pure-strategy NE over the finite action space $\mathcal{A} = \mathcal{A}_1 \times \cdots \times \mathcal{A}_n$, and utility functions $u_i(\cdot), i \in [n]$, let $u : \mathcal{A} \to \mathbb{R}$ be the social welfare function defined by $u(\boldsymbol{a}) := \sum_{i=1}^{n} u_i(\boldsymbol{a})$. Then, the price of anarchy of this game is defined to be

$$PoA = \frac{\max_{\boldsymbol{a} \in \mathcal{A}} \ u(\boldsymbol{a})}{\min_{\boldsymbol{a}^* \in NE} \ u(\boldsymbol{a}^*)},$$

where NE denotes the set of all pure-strategy NE points of the game. Note that we always have $PoA \geq 1$.[5]

In the following we show how to bound the PoA of the (not necessarily uniform) market sharing game. Let $\boldsymbol{a}^*$ and $\boldsymbol{a}^o$ be, respectively, a pure-strategy NE and an optimal action profile (i.e., the one which maximizes the social welfare over all action profiles). Now for any $i \in [n]$, let us define X_i to be the sum of all the market queries which are served by player i in the optimal action profile, but are not served by any player in the NE profile, i.e.,

$$X_i := \sum_{j \in a_i^o \setminus \cup_k a_k^*} q_j.$$

Then, we have $u_i(a^*) \geq X_i$. Otherwise, player i can deviate by changing his action from a_i^* to $a_i^o \setminus \cup_k a_k^*$, in which case its utility will increase to X_i, contradicting the fact that $\boldsymbol{a}^*$ is a NE. Next, we note that given any arbitrary action profile $\boldsymbol{a}$, by the definition of payoff function (12.2), the sum of the utility functions of all players (i.e., the social welfare) equals to $u(\boldsymbol{a}) = \sum_{j \in \cup_k a_k} q_j$. Thus, we can write

$$\begin{aligned} u(\boldsymbol{a}^o) - u(\boldsymbol{a}^*) &= \sum_{j \in \cup_k a_k^o} q_j - \sum_{j \in \cup_k a_k^*} q_j \\ &\leq \sum_{j \in (\cup_k a_k^o \setminus \cup_k a_k^*)} q_j \end{aligned}$$

[5]If instead of utility functions we were working with cost functions $c_i(\cdot), i \in [n]$, then the definition of PoA will change to $PoA = \frac{\max_{\boldsymbol{a}^* \in NE} \ c(\boldsymbol{a}^*)}{\min_{\boldsymbol{a} \in \mathcal{A}} \ c(\boldsymbol{a})}$, where $c : \mathcal{A} \to \mathbb{R}$ denotes the social cost function defined by $c(\boldsymbol{a}) = \sum_{i=1}^{n} c_i(\boldsymbol{a})$.

$$\leq \sum_{i=1}^{n} \sum_{j \in a_i^o \setminus \cup_k a_k^*} q_j = \sum_{i=1}^{n} X_i$$

$$\leq \sum_{i=1}^{n} u_i(a^*) = u(\boldsymbol{a}^*).$$

This shows that $\frac{u(\boldsymbol{a}^o)}{u(\boldsymbol{a}^*)} \leq 2$. As this relation holds for any Nash equilibrium profile a^*, we have the following theorem:

Theorem 5. *The PoA in the market sharing game is bounded above by 2.*

In fact, the result of Theorem 5 can be viewed as a special case of PoA bounds for so-called *valid utility* games (Vetta 2002). Valid utility games are a class of games where the utility functions satisfy certain conditions with sub-modular social welfare function.[6] In general it is known that the PoA of valid utility games is always bounded above by 2. Therefore, another way of proving Theorem 5 would be to show that the market sharing game is indeed a valid utility game.

3 Resource Allocation Games

As we saw in the previous section, congestion games constitute a subclass of games in which the existence of a pure-strategy NE is guaranteed. However there are many other types of games which are not congestion games and yet admit pure-strategy NE. In this section we provide three types of problems on network resource allocation to illustrate this matter further. The first two problems use a constructive argument to find a pure-strategy NE, while the last one uses a *generalized ordinal potential* function. In general, network resource allocation games are defined in terms of a set of available resources for each player, where the players are allowed to communicate through an undirected communication graph. Such a communication graph identifies the access cost among the players or the set of their feasible actions. We start by describing the following graphical resource allocation game proposed by Pacifici and Dan (2012):

3.1 A Graphical Resource Allocation

Consider a set of n nodes (players) and a set of resources $\mathcal{R}$. Every node is located at a vertex of an undirected graph $\mathcal{G} = ([n], \mathcal{E})$, called the influence graph. Each node i

[6]For a finite set $\mathcal{S}$, a set function $f(\cdot) : 2^{\mathcal{S}} \to \mathbb{R}$ is called sub-modular if $f(\mathcal{X} \cup \{s\}) - f(\mathcal{X}) \geq f(\mathcal{Y} \cup \{s\}) - f(\mathcal{Y})$ for any $\mathcal{X} \subseteq \mathcal{Y} \subseteq \mathcal{S}$, $s \in \mathcal{S}$.

allocates $k_i \in \mathbb{Z}^{>0}$ resources. The set of resources allocated by player i is described by its $|R|$ dimensional incident vector $\boldsymbol{a}_i = (a_i^1, \ldots, a_i^{|\mathcal{R}|}) \in \{0,1\}^{|\mathcal{R}|}$, where $a_i^r = 1$ if resource r is allocated by player i, and $a_i^r = 0$, otherwise. Therefore, one can view a_i as the action taken by player i which belongs to the action set $\mathcal{A}_i = \{a_i | \sum_r a_i^r \leq k_i\}$. A resource $r \in \mathcal{R}$ is called *i-busy* if it is allocated by at least one of player i's neighbors; otherwise it is called *i-free*. Moreover, we let v_{ir} be the value of resource r for player i. The payoff that a player i gets from allocating a resource r is influenced by the resource allocation of its neighboring nodes $\mathcal{N}(i)$ and is given by

$$u_i^r(1, a_{-i}^r) = \begin{cases} v_{ir} & \text{if } r \text{ is } i\text{-free}, \\ \delta_i v_{ir} & \text{if } r \text{ is } i\text{-busy}, \end{cases}$$

where $0 \leq \delta_i < 1$ is the cost of sharing for player i. A player i does not get any payoff from a resource r that it does not allocate, i.e., $u_i^r(0, a_{-i}^r) = 0$. Therefore, a player i gets the full value of a resource r if it allocates r while none of its neighbors allocates that resource. But, if resource r is already allocated by some of i's neighbors, then player i only receives a δ_i portion of the full value v_{ir}. Finally, the total payoff of player i equals the sum of payoffs over all possible resources, i.e., $u_i(a_i, a_{-i}) = \sum_{r \in \mathcal{R}} u_i^r(a_i^r, a_{-i}^r)$. Such a model has applications in selfish object replication on graphs, distributed radio spectrum allocation, and medium access control.

Remark 1. In the above graphical resource allocation game, the utility of a player is entirely specified by the actions of its neighbors. In the next subsection, we shall see a more complex resource allocation game where a player's action can depend heterogeneously on all others' in the network.

One can easily construct an example of the above graphical resource allocation game in which the sequential best replies of the players cycle. In other words, there are scenarios where starting from an initial action profile, there exists a sequence of players who sequentially play their best replies, and yet the game returns to its initial action profile. This eliminates the possibility of the existence of a potential function for this game as potential games have finite path improvement property (every sequence of best replies must be acyclic and converge to a pure-strategy NE). However, it seems that the requirement of having an exact potential function or having finite path improvement property for all sequences of best replies is far too strict to guarantee the existence of a NE. What if there exists at least one best reply path which can take us to a NE? This motivates introduction of the following definition:

Definition 6. A game is weakly acyclic under best replies if from every action profile $\boldsymbol{a}$, there is a best reply improvement path starting from $\boldsymbol{a}$ and ending in a pure-strategy NE.

Based on this definition, it has been shown by Pacifici and Dan (2012) that the graphical resource allocation game admits a pure-strategy NE:

Theorem 6. *The graphical resource allocation game is weakly acyclic and hence possesses a pure-strategy NE. Furthermore, from an arbitrary action profile, there exists a sequence of best replies by the players that reaches a NE in at most* $\sum_{i=1}^{n}\sum_{j\in\mathcal{N}(i)\cup\{i\}} k_j$ *steps.*

The main idea behind establishing Theorem 6 is to prioritize the players who first play their best responses. If there is no specific ordering of the players' best replies, it is possible that the generated sequence of action profiles along the best response sequence cycle without leading to a NE. To avoid such a situation, let $a_i(t)$ be a best response of player i at time step t and $E_i(t)$ and $I_i(t)$ be the sets of resources that player i evicts and inserts as a result of its best action $a_i(t)$, i.e., $E_i(t) = \{r|a_i^r(t-1) = 1, a_i^r(t) = 0\}$, and $I_i(t) = \{r|a_i^r(t-1) = 0, a_i^r(t) = 1\}$. Then, one can define four not mutually exclusive properties of $a_i(t)$, depending on whether the involved resources are i-busy or not as follows:

(1) $\exists r \in E_i(t)$, r is i-busy & $\exists r' \in I_i(t)$, r' is i-busy
(2) $\exists r \in E_i(t)$, r is i-busy & $\exists r' \in I_i(t)$, r' is i-free
(3) $\exists r \in E_i(t)$, r is i-free & $\exists r' \in I_i(t)$, r' is i-busy
(4) $\exists r \in E_i(t)$, r is i-free & $\exists r' \in I_i(t)$, r' is i-free

For instance, $a_i(t)$ having property (1) means that the best response of player i at time instant t results in evicting an i-busy resource r and inserting an i-busy resource r'. The rest of the proof follows by showing the following four steps: (i) Starting from an arbitrary initial action profile $\boldsymbol{a}(0)$, we let the players sequentially perform best responses of type (1) only. (ii) It can be shown that after performing at most $\sum_{j=1}^{n} k_j$ best replies with property (1), we reach an action profile $\boldsymbol{a}(t)$ in which there is no player that can perform a best reply that has property (1). (iii) Starting from the action profile $\boldsymbol{a}(t)$ obtained at the end of stage (ii), one can show that the only possibly best responses available for a player are those which have property (2) or (4). (iv) Finally, one can argue that a best reply of a player with property (2) or (4) strictly increases the utility of that player without decreasing any other player's utility. Since the players' utilities cannot increase indefinitely, this process will terminate after at most an extra $\sum_{i=1}^{n}\sum_{j\in\mathcal{N}(i)} k_j$ best replies of types (2) or (4) to an action profile where no player can play a strict best response, i.e., a pure-strategy NE.

Weakly acyclic games are especially interesting from a probabilistic and learning perspective. For example, it can be shown that random selection of players for performing best responses will converge in an expected finite time to a pure-strategy NE, even though the expected convergence time may be very long. In particular, weakly acyclic games are quite suitable for adopting distributed learning algorithms such as regret-based dynamics for finding their NE points (Marden et al. 2007).

3.2 Uncapacitated Selfish Caching Game

Caching files by the server nodes in a replication system is an effective method for improving the performance, availability of the files, and reliability of the systems. However, in many practical situations for obvious reasons, the server nodes are selfish elements who only wish to maximize their own utilities, regardless of the overall performance of the system (e.g., servers in different administrative domains utilize their local resources to better support their own clients). In this regard, one of the basic, yet descriptive, models for caching under competitive environment was given by Chun et al. (2004) as is described next.

Selfish Caching Game: Consider an undirected weighted network topology $\mathcal{G} = ([n], \mathcal{E})$ where each node represents a server (player). We assume that the distances on this network are symmetric and satisfy the triangle inequality, i.e., $d_{ij} + d_{jk} \geq d_{ik}$, where d_{ij} denotes the distance of two nodes i and j on the network $\mathcal{G}$. Each player i has a feasible set of actions, $\mathcal{A}_i$, where an action $a_i \in \mathcal{A}_i$ is a set of objects that player i can cache. Given the action profile $\boldsymbol{a} = (a_1, \ldots, a_n)$ chosen by all the players, the cost of player i is defined by

$$c_i(\boldsymbol{a}) = \sum_{o \in a_i} \alpha_o + \sum_{o \notin a_i} w_{io} d_{i\sigma_i(\boldsymbol{a},o)} \tag{12.3}$$

where α_o is the cost of caching object o, w_{io} is the demand rate that server i receives from its clients for object o, and $\sigma_i(\boldsymbol{a}, o)$ is the closest server to i that caches object o given the action profile $\boldsymbol{a}$. When no server caches the object, we define distance cost $d_{i\sigma_i(\boldsymbol{a},o)}$ to be very large so that at least one server will choose to cache the object o. This defines a noncooperative game among the players where each player wants to cache a subset of objects in order to minimize its own cost. Thus, the first question of interest here is to see whether this selfish caching game admits a pure-strategy NE.

There are often scenarios where analyzing the game as a whole is a challenging task. However, a closer look into the structure of the game reveals simple patterns which allows us to analyze the game for a basic case and build upon that to derive some conclusions for the general model. In fact, selfish caching game provides one of such instances. By a closer look into this game, one can easily see that since there is no capacity limit on cache size of the servers, one can look at each single object as a separate game and combine the pure-strategy Nash equilibria of these games to obtain a pure-strategy NE for the original multi-object game. More generally, it is known that if two games have pure-strategy NE, and their payoff functions are (in some precise sense not defined here) cross monotonic, then their union (same players, the union of the strategy spaces, and the same payoffs) is also guaranteed to have pure-strategy NE (Fabrikant et al. 2004). Therefore, we can first analyze the basic version of the selfish caching game with a single object and then glue their equilibrium points (if there exists any) to recover a NE for the multi-object game.

For the single object selfish caching game, which henceforth we refer to as the *basic* game, each server i has two strategies, namely, to cache or not to cache. Also, assume that the object under consideration is o. For simplicity we can represent the action of player i by a binary variable:

$$a_i = \begin{cases} 1 & \text{if player } i \text{ caches object } o, \\ 0 & \text{otherwise,} \end{cases}$$

in which case the cost (12.3) adapted for the basic game can be written as

$$c_i(\boldsymbol{a}) = \alpha_o a_i + w_{io} d_{i\sigma_i(\boldsymbol{a},o)}(1 - a_i). \tag{12.4}$$

Note that in the basic game, we have $\boldsymbol{a} \in \{0,1\}^n$. One can easily check that for the basic game with the cost function (12.4), an action profile $\boldsymbol{a}^* = (a_i^*, a_{-i}^*)$ constitutes a pure-strategy NE if and only if the following two conditions are satisfied:

(1) If $a_i^* = 0$ for some i, then $\exists j$ such that $a_j^* = 1$ and $d_{ij} \leq \frac{\alpha_o}{w_{io}}$.
(2) If $a_i^* = 1$ for some i, then $\forall j, a_j^* = 1$ we have $d_{ij} \geq \frac{\alpha_o}{w_{jo}}$.

The first condition guarantees that if object o is not cached by player i, it is because there is another server in his close vicinity who has the object (hence player i has no incentive to deviate). The second condition guarantees that if player i has object o in his cache, he has no incentive to drop the object because the object is not available in any of the other servers in his close vicinity. Having this characterization at hand, one can use an inductive argument to construct a pure-strategy NE for the basic game.

Theorem 7. *Pure-strategy Nash equilibria exist in the basic game.*

The idea is to inductively allocate object o to a new player while assuring that both conditions (1) and (2) are satisfied. For this purpose, let $\beta_i = \frac{\alpha_o}{w_{io}}, i \in [n]$, and consider the following iterative process:

- Initialize $\mathcal{T} = [n]$.
- Pick a server $j \in \mathcal{T}$ such that $\beta_j \leq \beta_i, \forall i \in \mathcal{T}$, and set $a_j = 1$.
- Let $Z(j) = \{i \in \mathcal{T} : d_{ij} \leq \beta_i\}$, and set $\mathcal{T} = \mathcal{T} \setminus Z(j)$.
- Iterate until $\mathcal{T} = \emptyset$. Set all a_i which are not assigned 1 in the above process to 0. Output $\boldsymbol{a} = (a_1, \ldots, a_n)$ as a pure-strategy NE.

The key idea in the above process is that at each time, we allocate the object o to a player j and eliminate all others who can access o through j with a lower cost. This set is precisely $Z(j)$. Moreover, because at each iteration j is the remaining node with the minimum β, no replica will be placed within distance β_j of any such j by this process. Therefore, at each stage of the above process, both conditions (1) and (2) are held, and the process terminates with a valid pure-strategy NE.

Now that existence of a pure-strategy NE in the selfish caching game is guaranteed, we next turn our attention to "efficiency" of such equilibrium points. To analyze PoA, let us further assume that all the demand rates are equal to 1, i.e., $w_{io} = 1, \forall i \in [n]$. To keep the analysis simple, we analyze PoA for the basic game. However the results can easily be generalized to multi-object selfish caching game as the basic games corresponding to different objects are uncorrelated due to no capacity constraints. Using the cost function (12.4), one can express the cost of the worst pure-strategy NE in a more closed form as

$$\max_{\boldsymbol{a}^* \in NE} c(\boldsymbol{a}^*) = \max_{\boldsymbol{a}^* \in NE} \{\alpha_o |\boldsymbol{a}^*| + \sum_{i=1}^{n} \min_{j: a_j^* = 1} d_{ij}\},$$

where $|\boldsymbol{a}^*|$ denotes the number of players who cache object o (i.e., the number of 1's in the action profile $\boldsymbol{a}^*$), $\min_{j:a_j^*=1} d_{ij}$ represents the distance to the closest replica (including i itself) from player i, and the set NE contains all pure-strategy NE profiles $\boldsymbol{a}^*$ characterized by conditions (1) and (2). Now using some case-dependent analysis based on parameter α_o or the network structure, one can derive general upper bounds on the PoA. In the following we provide two instances of such an analysis.

Let $d_{\min} = \min_{i \neq j} d_{ij}$ and $d_{\max} = \max_{i \neq j} d_{ij}$ be, respectively, the minimum and maximum possible distances between servers in the network.

- $\alpha_o < d_{\min}$: Every server caches object o for both NE and social optimum, i.e., $PoA = 1$.
- $\alpha_o > d_{\max}$: Since the placement cost is greater than any distance cost, in any NE only one server caches the object and other servers access it remotely. Therefore, the worst cost of a NE equals $\alpha_o + \max_j \sum_i d_{ij}$. However, the social optimum may still place multiple replicas. Since the optimal social cost always satisfies $\alpha_o \le \min_{\boldsymbol{a} \in \mathcal{A}} c(\boldsymbol{a}) \le \alpha_o + \min_j \sum_i d_{ij}$, in this case we have $\frac{\alpha_o + \max_j \sum_i d_{ij}}{\alpha_o + \min_j \sum_i d_{ij}} \le PoA \le \frac{\alpha_o + \max_j \sum_i d_{ij}}{\alpha_o}$; depending on the underlying network, we can even have $PoA = \Omega(n)$.

Such case-dependent analysis can be carried over for other range of parameters or network topologies. For instance, it has been shown in Chun et al. (2004) that for the complete graph, star, line, and D-dimensional grid, the PoA is bounded above by 1, 2, $O(\sqrt{n})$, and $O(n^{\frac{D}{D+1}})$, respectively. As it can be seen, depending on the underlying parameters, the equilibrium points of the selfish caching game can be quite efficient or inefficient. In general, games which benefit from having a low PoA are quite valuable for modeling chaotic allocation systems where the players act completely in a selfish manner. In the next subsection, we provide one of such class of games, with a richer structure to guarantee a very low PoA.

3.3 Capacitated Selfish Replication Game

In the previous two subsections, we discussed two types of games which are not exact potential games but still admit pure-strategy NE. In this section, we provide a larger class of games known as *generalized ordinal potential* games which include exact potential games as a special subclass. We provide an instance of such games in the context of selfish network resource allocation and study some of their algorithmic aspects such as finding (approximating) the equilibrium points or bounding the PoA. We begin with the following definition:

Definition 7. An n-player game $\Gamma = ([n], \{\mathcal{A}_i\}, \{c_i\})$ is called a generalized ordinal potential game if there exists a function $\Phi : \mathcal{A}_1 \times \cdots \times \mathcal{A}_n \to \mathbb{R}$ such that $\forall a_{-i} \in \mathcal{A}_{-i}, \forall a_i, \hat{a}_i \in \mathcal{A}_i$, we have

$$c_i(a_i, a_{-i}) - c_i(\hat{a}_i, a_{-i}) > 0 \;\Rightarrow\; \Phi(a_i, a_{-i}) - \Phi(\hat{a}_i, a_{-i}) > 0, \quad \forall i \in [n].$$

An immediate consequence of the above definition is that every generalized ordinal potential game with finite action space admits a pure-strategy NE. This is because starting from an arbitrary action profile, after a finite number of steps of single player improvement, we will reach a local minima of the generalized ordinal potential function which is a pure-strategy NE. Note that the existence of an exact potential function is a special case of the existence of a generalized ordinal potential function. While having generalized ordinal potential function is sufficient to guarantee existence of a pure-strategy NE, it usually does not provide an efficient way to find one of such equilibrium points. Therefore, an important open question is to see whether generalized ordinal potential games admit polynomial time algorithms for finding their pure-strategy NE points. To provide a more concrete example, we consider the following game known as *capacitated selfish replication* (CSR) game (Etesami and Başar 2017a,b; Gopalakrishnan et al. 2012). This game can be viewed as a capacitated version of the uncapacitated selfish caching game studied in the previous section, where the servers have limited capacity on their cache sizes. As we will see, this introduces more complication into the game as the cache constraints couple the players' actions much more than in the uncapacitated selfish caching game.

CSR Game: We start with a set $[n] = \{1, 2, \ldots, n\}$ of nodes (players) which are connected by an undirected *unweighted* graph $\mathcal{G} = ([n], \mathcal{E})$. We denote the set of all objects by $\mathcal{O} = \{o_1, o_2, \ldots, o_k\}$. For simplicity, but without much loss of generality, we assume that each node can hold only one resource in its cache. In other words, the action of player i, denoted by a_i, is a single element subset of $\mathcal{O}$, where the content of a_i is the object cached by player i. However, all the results can in fact be extended quite readily to CSR games with different capacities where a_i can be an arbitrary subset of $\mathcal{O}$ of a certain size (see Remark 2). Further, we assume that each node has access to all the objects. For a particular action profile (allocation) $\boldsymbol{a} = (a_1, a_2, \ldots, a_n)$, we define the sum cost function $c_i(\boldsymbol{a})$ of the ith player as follows:

$$c_i(\boldsymbol{a}) = \sum_{o \in \mathcal{O} \setminus a_i} d_{\mathcal{G}}(i, \sigma_i(\boldsymbol{a}, o)), \tag{12.5}$$

where as before $\sigma_i(\boldsymbol{a}, o)$ is i's nearest node holding o in $\boldsymbol{a}$, and $d_{\mathcal{G}}(\cdot, \cdot)$ denotes the graphical distance between two nodes in the graph $\mathcal{G}$. Finally, if some resource o is missing in an allocation profile $\boldsymbol{a}$, we define the cost of each player for that specific resource to be very large. Therefore, for $n \geq |\mathcal{O}|$, this incentivizes at least one of the players to allocate the missing resources in the network. In the case where $n < |\mathcal{O}|$, all the players will allocate different resources and the game becomes trivial; hence, we can simply assume that $n \geq |\mathcal{O}|$. In Fig. 12.1 we have illustrated the CSR game for $n = 11$ players and $|\mathcal{O}| = 3$ resources.

Remark 2. For the CSR game with different capacities, one can introduce a new undirected network which transfers games with different cache sizes to one with unit size caches. This can be done simply by replacing each player i with capacity L_i with L_i new sub-nodes as a clique, namely, $i_1, i_2, \ldots, i_{L_i}$, each with cache size 1. We treat all of these sub-nodes in the clique as node i and very naturally connect them

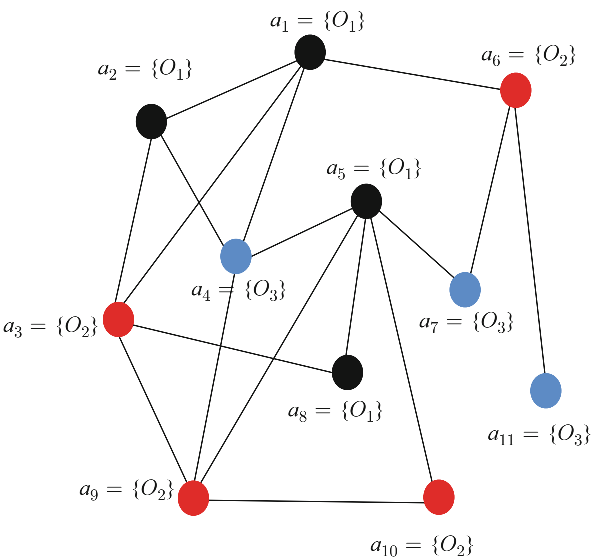


Fig. 12.1 CSR game with $n = 11$ players and $O = \{o_1, o_2, o_3\}$ resources

to all the other sub-nodes where i was connected before. This makes an equivalence between the CRG games with different cache sizes and that with unit cache sizes; see (Etesami and Başar 2017b, Sect. VI).

Based on Remark 2, we analyze here only the CSR game with unit cache size. Let us first consider a simple case where the network $\mathcal{G}$ has a tree structure, i.e., there is no cycle in the network. In this case, the following simple iterative process can deliver a pure-strategy NE of the CSR game after only n steps (Etesami and Başar 2017a).

Constructing NE on a Tree: Start from an arbitrary node of $\mathcal{G}$ as a root and label it by 1. Define ℓth level to be the set of all nodes in $\mathcal{G}$ which are at a distance ℓ from node 1. At the ith step of the algorithm, we choose a player at the highest level who has not been chosen before; we label it by i, and let her allocate one resource based on her best response with respect to all the agents who have updated before, i.e., $1, 2, \ldots, i-1$. Ties are broken arbitrarily.

However, when the network $\mathcal{G}$ has a cycle, the situation becomes more complicated. In this case, we argue that CSR game is a generalized ordinal potential game and hence admits a pure strategy NE. To this end, for a given allocation profile $\boldsymbol{a}$, let the *radius* of agent i, denoted by $r_i(\boldsymbol{a})$, be the distance between node i and the nearest node other than her holding the same resource as i, i.e., $r_i(\boldsymbol{a}) = \min_{j \neq i, a_j = a_i} d_{\mathcal{G}}(i, j)$. If there does not exist such a node, we simply define $r_i(\boldsymbol{a}) = D$, where D is the diameter of the network (i.e., the maximum graphical distance between any pair of nodes in $\mathcal{G}$). Using this convention, one can see an equivalence between decrease in cost and increase in radius for player i, when the actions of the remaining players are fixed. This is because given two allocation profiles $\boldsymbol{a}$ and $\tilde{\boldsymbol{a}}$, which only differ in the ith coordinate, using (12.5) and the definition of radius, one can easily see that $c_i(\boldsymbol{a}) - c_i(\tilde{\boldsymbol{a}}) = r_i(\tilde{\boldsymbol{a}}) - r_i(\boldsymbol{a})$. In addition, for $\ell = 1, 2, \ldots, D$, let $n_\ell(\boldsymbol{a})$ be the number of players whose radii are ℓ, i.e., $n_\ell(\boldsymbol{a}) = |\{i : r_i(\boldsymbol{a}) = \ell\}|$, and define the radius vector $\boldsymbol{n}(\boldsymbol{a})$ to be $\boldsymbol{n}(\boldsymbol{a}) := (n_1(\boldsymbol{a}), \ldots, n_D(\boldsymbol{a}))$. Then we have the following theorem:

Theorem 8. *The radius vector $\boldsymbol{n}(\boldsymbol{a})$ lexicographically decreases after every strictly better response by a player. In particular, $\Phi(\boldsymbol{a}) := \sum_{\ell=1}^{D} n_\ell(\boldsymbol{a}) \times n^{D-\ell}$ serves as a generalized ordinal potential function for the CSR game, where its value decreases after each strictly better response of a player.*

As a result of Theorem 8, the CSR game always admits a pure-strategy NE. However, note that the generalized ordinal potential function $\Phi(\boldsymbol{a})$ can take exponentially large values (up to $\mathcal{O}(n^D)$) and hence cannot be used to obtain or characterize the NE points efficiently. Therefore, in the following and motivated by the structure of $\Phi(\boldsymbol{a})$, we provide two simple algorithms where the first one finds a pure-strategy NE in $O(n^2)$ when the network structure is dense with respect to the number of objects, while the second one provides a constant approximation of a NE in quasi-polynomial time $O(n^{\ln D})$ over general networks.

A closer look into the structure of the generalized ordinal potential function $\Phi(\boldsymbol{a}) := \sum_{\ell=1}^{D} n_\ell(\boldsymbol{a}) \times n^{D-\ell}$ shows that if the players were updating based on their best responses at each stage, then the speed of convergence of the best response dynamics would highly depend on the radius of the updating agent at each time instant. In particular, the lower the radius of updating player, the steeper would be the decrease in the value of the function $\Phi(\cdot)$. Since such a function is always nonnegative, the iterations of the best response dynamics must terminate to a minimizer of this function which is a valid NE. Based on this observation, and in order to incorporate the role of updating radii into the algorithm, we introduce a slightly different version of the best response dynamics as shown in Algorithm 1. Based on this algorithm, one can establish the following result (Etesami and Başar 2017b):

Algorithm 1 Least best response

Given a CSR game, at each time $t = 1, 2, \ldots$, and from the set of all players who want to update, select a player with the least radius and let her update be based on her strictly best response. Ties are broken arbitrarily.

Theorem 9. *Let d_m denote the minimum degree of the graph in the CSR game with n players. Then, the least best response dynamics will converge to a pure-strategy NE in no longer than T steps, where*

$$T = \begin{cases} n^2 & \text{if } \; d_m \geq |\mathcal{O}|, \\ 3n & \text{if } \; |\mathcal{O}| \leq 4. \end{cases}$$

The idea of establishing the above bounds is to show that once players start to play their best responses based on Algorithm 1, those who have radii smaller than the current updating player will still play their best responses. As a result, after at most every constant number of stages, one more player will play his best response, and the dynamics must terminate to a pure-strategy NE. Now if none of the conditions in Theorem 9 are satisfied, the following algorithm proves useful to find an ϵ-approximation of a pure-strategy NE.

Algorithm 2 ϵ-best response

Given a network $\mathcal{G} = ([n], \mathcal{E})$, a real number $\epsilon > 1$, a set of available resources $\mathcal{O}$, and an arbitrary initial allocation profile $\boldsymbol{a}(0)$, at every time instance t select an agent i who can increase its radius $r_i(\boldsymbol{a}(t))$ by a factor of at least ϵ, i.e., $r_i(\boldsymbol{a}(t+1)) \geq \epsilon r_i(\boldsymbol{a}(0))$, and let her play her best response. Ties are broken arbitrarily.

Theorem 10. *Given a real number $\epsilon > 1$, the ϵ-best response algorithm terminates after at most $\mathcal{O}\left(n^2 D^{\log_\epsilon n}\right)$ steps with an allocation profile $\hat{\boldsymbol{a}}$ which is an ϵ-approximation of a pure-strategy NE, i.e., an allocation profile in which no player can reduce its cost by a factor of more than $\frac{1}{\epsilon}$.*

Using the equivalence between reduction in cost and increase in the radius, and by the definition of the ϵ-best response dynamics, it is not hard to see that if the ϵ-best response algorithm terminates, it must end at an action profile where no player can further increase its cost by a factor more than ϵ (hence, reduce its cost by a factor more than $\frac{1}{\epsilon}$), i.e., an ϵ-approximated NE. Therefore, the key part of the argument in establishing Theorem 10 is to bound the number of iterations in Algorithm 2. This can be done using the following nonnegative potential function:

$$R(\boldsymbol{a}(t)) := \sum_{k=1}^{D} \frac{n_k(\boldsymbol{a}(t))}{k^{\log_\epsilon n}},$$

where D is the diameter of the network and $\boldsymbol{a}(t)$ denotes the action profile at the t-th iteration of Algorithm 2. One can show that after each time of running the ϵ-best response dynamics, the value of $R(\cdot)$ decreases by at least $\frac{1}{nD^{\log_\epsilon n}}$. Since $R(\cdot)$ is upper bounded by n, it cannot decrease by more than $n^2 D^{\log_\epsilon n}$ times, which shows that the dynamics must terminate after at most these many steps.

Next, we turn our attention to the efficiency of the Nash equilibrium points of the CSR game. As we saw in the previous subsection, the uncapacitated selfish caching game could have a wide range of PoA, depending on the parameters of the problem. Surprisingly, the following theorem shows that this is not the case for the capacitated selfish replication game:

Theorem 11. *In the CSR game, we have $PoA \leq 3$.*

To see how we can upper bound the PoA in CSR game, we need to somehow connect the social cost of a NE with that of the optimal one. Therefore, we first seek to find a characterization for an arbitrary NE. Let $\boldsymbol{a}^*$ be an arbitrary NE and consider a specific node i with equilibrium radius $r_i(\boldsymbol{a}^*)$. We note that all the objects must appear at least once in $B_{\mathcal{G}}(i, r_i(\boldsymbol{a}^*))$ which is a graphical ball containing all the nodes whose graphical distances to i are less than or equal to $r_i(\boldsymbol{a}^*)$. In fact, if a specific object is missing in $B_{\mathcal{G}}(i, r_i(\boldsymbol{a}^*))$, then node i can increase its radius by updating its current object to that specific missing object, thereby decreasing its cost. But this is in contradiction with $\boldsymbol{a}^*$ being a NE. Now, given the equilibrium profile $\boldsymbol{a}^*$, let us define $\hat{r}_i$ to be the smallest integer such that $B_{\mathcal{G}}(i, \hat{r}_i)$ contains at least two resources of the same type, i.e.,

$$\hat{r}_i = \min \left\{ r \in \mathbb{N} : \forall j, k \in B_{\mathcal{G}}(i, r-1), a_j^* \neq a_k^*, \text{and } \exists j_0, k_0 \in B_{\mathcal{G}}(i, r), a_{j_0}^* = a_{k_0}^* \right\}.$$

Now we claim that all the resources must appear at least once in $B_{\mathcal{G}}(i, 3\hat{r}_i)$. To see this and by the above definition, let $j_0 \neq k_0 \in B_{\mathcal{G}}(i, \hat{r}_i)$ be such that $a_{j_0}^* = a_{k_0}^*$. This means that the equilibrium radius of node j_0, i.e., $r_i(\boldsymbol{a}^*)_{j_0}$ is at most $d_{\mathcal{G}}(j_0, i) + d_{\mathcal{G}}(i, k_0) \leq 2\hat{r}_i$. On the other hand, by the argument at the beginning of the proof, all the resources must appear at least once in $B_{\mathcal{G}}(j_0, 2\hat{r}_i)$. But since

$B_{\mathcal{G}}(j_0, 2\hat{r}_i) \subseteq B_{\mathcal{G}}(i, 3\hat{r}_i)$, this shows that $B_{\mathcal{G}}(i, 3\hat{r}_i)$ must include all the resources at least once.

Next, let us denote an optimal allocation profile by $\boldsymbol{a}^o$ and the cost of node i in the optimal allocation and at NE by $c_i(\boldsymbol{a}^o)$ and $c_i(\boldsymbol{a}^*)$, respectively. Now for the Nash equilibrium $\boldsymbol{a}^*$, and since by the definition of $\hat{r}_i$ there are no two similar resources in $B_{\mathcal{G}}(i, \hat{r}_i - 1)$, and all the resources appear at least once in $B_{\mathcal{G}}(i, 3\hat{r}_i)$, we can write

$$c_i(\boldsymbol{a}^*) = \sum_{o \in \mathcal{O}\setminus\{a_i\}} d_{\mathcal{G}}(i, \sigma_i(\boldsymbol{a}^*, o)) \leq \sum_{j \in B_{\mathcal{G}}(i, \hat{r}_i - 1)} d_{\mathcal{G}}(i, j) + 3\hat{r}_i \left(|\mathcal{O}| - 1 - |B_{\mathcal{G}}(i, \hat{r}_i - 1)|\right). \tag{12.6}$$

On the other hand, for the cost of node i in the optimal allocation P^o, we have

$$c_i(\boldsymbol{a}^o) = \sum_{o \in \mathcal{O}\setminus\{a_i\}} d_{\mathcal{G}}(i, \sigma_i(\boldsymbol{a}^o, o)) \geq \sum_{j \in B_{\mathcal{G}}(i, \hat{r}_i - 1)} d_{\mathcal{G}}(i, j) + \hat{r}_i \left(|\mathcal{O}| - 1 - |B_{\mathcal{G}}(i, \hat{r}_i - 1)|\right), \tag{12.7}$$

where the inequality holds since node i has to pay at least $\sum_{j \in B_{\mathcal{G}}(i, \hat{r}_i - 1)} d_{\mathcal{G}}(i, j)$ for the first $|B_{\mathcal{G}}(i, \hat{r}_i - 1)|$ closest objects and to pay at least $\hat{r}_i$ for the remaining $(|O| - 1 - |B_{\mathcal{G}}(i, \hat{r}_i - 1)|)$ resources. Comparing relations (12.6) and (12.7), we get

$$\frac{c_i(\boldsymbol{a}^*)}{c_i(\boldsymbol{a}^o)} \leq \frac{\sum_{j \in B_{\mathcal{G}}(i, \hat{r}_i - 1)} d_{\mathcal{G}}(i, j) + 3\hat{r}_i \left(|\mathcal{O}| - 1 - |B_{\mathcal{G}}(i, \hat{r}_i - 1)|\right)}{\sum_{j \in B_{\mathcal{G}}(i, \hat{r}_i - 1)} d_{\mathcal{G}}(i, j) + \hat{r}_i \left(|\mathcal{O}| - 1 - |B_{\mathcal{G}}(i, \hat{r}_i - 1)|\right)},$$

which is bounded from above by 3. Thus, for every equilibrium $\boldsymbol{a}^*$ and for all $i \in [n]$, we have $c_i(\boldsymbol{a}^*) \leq 3c_i(\boldsymbol{a}^o)$. Summing both sides of this inequality over all $i \in [n]$, we get $c(\boldsymbol{a}^*) \leq 3c(\boldsymbol{a}^o)$, i.e., $PoA \leq 3$.

As we mentioned earlier, PoA is an important metric to evaluate the efficiency of the NE points of a game. Therefore, Theorem 11 shows that CSR game has highly efficient equilibrium points despite the complete selfish nature of its players. In the next sections, we shall use the PoA metric to evaluate the efficiency of the equilibrium points for other types of network games and derive analogous bounds for their PoA.

4 Diffusion Games

One of the widely studied models in network games, and in particular within the context of social networks, is the diffusion model, where the goal is to propagate a certain type of product or behavior in a desired way through the network. Examples include online advertising for companies' products, propagation of rumors and computer viruses, and epidemics. One of the challenges within this area is to understand how to control the diffusion process in a proper way by possibly targeting and investing on the most influential entities in the underlying social network. This problem becomes even more complicated when there are several parties who plan to maximize their benefits by spreading their own products. This brings up the notion of competitive diffusion over networks, where game theoretic tools seem quite appealing and effective. However, depending on the objective and the complex nature of social networks which might be woven with rational decisions, one can find various models aimed at capturing the idea of competition diffusion over networks. In this section we overview some of the earlier important models within this context and elaborate on the analytic tools for analyzing them. We shall start by describing the following deterministic diffusion game model first proposed by Alon et al. (2010).

4.1 Deterministic Competitive Diffusion Game

Consider an undirected network $\mathcal{G} = ([n], \mathcal{E})$ of n nodes and two players (types) A and B with positive integer budgets l_A and l_B, respectively. Initially at time $t = 0$, each player decides to choose a subset of nodes with size equal to its budget and place his own seeds. In other words, the action of each player $i \in \{A, B\}$ is a subset of nodes $a_i \subseteq [n]$ initially selected by that player such that $|a_i| = l_i$. After that, a discrete time diffusion process unfolds among uninfected nodes as follows:

- If at some time step t an uninfected node is neighbor to infected nodes of only one type (A or B), it will adopt that type at the next time step $t + 1$.
- If an uninfected node is connected to nodes of both types at some time step t, it will change to a gray node at the next time step $t + 1$ and does not adopt any type afterward.

This process continues until no new adoption happens. Finally, the utility of each player will be the total number of infected nodes of its own type at the end of the process. We assume that if both players initially place their seeds on the same node, that node will change to gray (hence, no one will benefit from it). We want to emphasize the fact that when a node changes to gray, not only will it not adopt any type at the next time step but also may block the streams of diffusion to other uninfected nodes. In fact, the existence of such gray nodes in the evolution of the process makes any prediction about the outcome of the diffusion process very difficult.

Remark 3. In fact, the above model can readily be extended to the case when there are more than two players. In this case, one can introduce the same process as above such that an uninfected node will adopt type i at time $t + 1$ if and only if type i is the only existing type among its neighbors at time step t.

It has been shown in Alon et al. (2010) and Small and Mason (2013) that the deterministic competitive diffusion game may or may not admit pure-strategy NE depending on the topology of the network $\mathcal{G}$ and the number of the players. For instance, one can easily check that if the underlying graph $\mathcal{G}$ has a tree structure, the deterministic diffusion game with two players admits a pure-strategy NE; this result no longer holds when there are three or more players even with the tree network structure. Therefore, one of the immediate questions with regard to the deterministic diffusion game is to see under what network structures such a game admits a pure-strategy NE. The following theorem provides a somehow negative response to this question:

Theorem 12. *Given a graph $\mathcal{G}$ and $m \geq 2$ players, deciding whether the deterministic diffusion game admits a pure-strategy NE on $\mathcal{G}$ is NP complete.*

The above theorem suggests that unless P = NP, there is no efficient way of deciding whether the deterministic diffusion game admits a pure-strategy NE. The idea behind establishing the result of Theorem 12 is to reduce the existence of a pure-strategy NE in diffusion game to a solution of the *3-partitioning problem* which is known to be an NP-complete problem. In the 3-partitioning problem, we are given integers $\alpha_1, \alpha_2, \ldots, \alpha_{3m}$ and a β such that $\frac{\beta}{4} < \alpha_i < \frac{\beta}{2}$ for every $1 \leq i \leq 3m$, $\sum_{i=1}^{3m} \alpha_i = m\beta$ and have to partition them into disjoint sets $P_1, \ldots, P_m \subseteq \{1, 2, \ldots, 3m\}$ such that for every $1 \leq j \leq m$, we have $\sum_{i \in P_j} \alpha_i = \beta$. Therefore, one can carefully design a network which encodes the structures of the 3-partitioning problem such that the existence of a pure-strategy NE of the m-player (single seed) deterministic diffusion game over such a network is equivalent to a solution of the encoded 3-partitioning problem. For more details, see Etesami and Başar (2016).

While Theorem 12 provides a hardness result on the existence of a pure-strategy NE, if we restrict our attention to specific networks, it is still possible to obtain some positive results. To provide some examples, let us first consider the following networks: (i) An $r \times s$ lattice is a graph $L_{r \times s}$ with vertex set $V = \{(x, y) \in \mathbb{Z}^2 : 0 \leq x \leq r, 0 \leq y \leq s\}$ such that each node is adjacent to those whose Euclidean distances are 1 from it. (ii) A k-dimensional hypercube is a graph Q_k with vertex set $\{0, 1\}^k$ such that two k-tuples are adjacent if and only if they differ in exactly one coordinate. (iii) An Erdos-Renyi graph $\mathcal{G}(n, p)$ is an n node random graph where the edges emerge independently with probability $p \in (0, 1)$. Based on the above network structures, we have the following result:

Proposition 1. *For the two-player deterministic diffusion game with single seed placement (i.e., $l_1 = l_2 = 1$), we have*

- *An action profile (a_1, a_2) is a NE over the lattice if and only if a_1 and a_2 are adjacent nodes in the most centric square or edges of the lattice.*
- *An action profile (a_1, a_2) is a NE over the hypercube Q_k if and only if the graphical distance between a_1 and a_2 is an odd number.*

- *Given arbitrary constants* $\alpha, c > 1$*, let* $p \in [\frac{c \ln n}{n}, \sqrt{\frac{(1+c)\ln n}{n^{\alpha}}}]$*. Then, as* $n \to \infty$*, for every random seed placement over* $\mathcal{G}(n, p)$*, we have* $\mathbb{E}[U_1] = \mathbb{E}[U_2] \geq \frac{1}{5p}$*, where* $\mathbb{E}[U_1]$ *and* $\mathbb{E}[U_2]$ *denote the expected utilities of players 1 and 2, respectively. In particular,* $\frac{U_1}{\mathbb{E}[U_1]} = \frac{U_2}{\mathbb{E}[U_2]} \to 1$*, almost surely.*

While Proposition 1 characterizes the set of pure-strategy NE for two special types of networks, it is still interesting to identify broader classes of networks which are guaranteed to admit pure-strategy NE. In particular, it would be interesting to see how imbalance of the initial budgets will affect the final outcomes. For instance, while for single seed placement over Erdos-Renyi networks both players can obtain a payoff of at least $\frac{1}{5p}$, it seems that slightly increasing a player's budget can substantially favor him in winning the whole set of nodes. Finally, we want to mention that one of the main reasons that makes the analysis of pure-strategy NE in deterministic diffusion game more difficult is the deterministic adoption rule in the structure of the game. In the subsequent subsections, we shall see other types of network diffusion games with probabilistic adoption rules that will allow us to analyze such games in more detail.

4.2 Competitive Contagion in Networks

In this subsection, we provide a rich class of competitive strategies, which depend in subtle ways on the *stochastic* dynamics of adoption, the relative budgets of the players, and the underlying structure of the network.

Initial Seed Allocation: Consider a game with two players R(ed) and B(lue) on a (possibly directed) network $\mathcal{G} = ([n], \mathcal{E})$. Each player $p \in \{R, B\}$ has a budget $k_p \in \mathbb{Z}^{\geq 0}$ and initially chooses an allocation of his budget over the set of vertices. Hence, the action of player p can be expressed as a vector $a_p = (a_{p1}, \ldots, a_{pn})$, where $a_{pi} \in \mathbb{Z}^{\geq 0}$ and $\sum_{i=1}^{n} a_{pi} = k_p$. Therefore, the set of all pure strategies of player p, denoted by $\mathcal{A}_p$, equals to the set of all such integer-valued vectors. A mixed strategy for player p is a probability distribution σ_p on $\mathcal{A}_p$. We denote the set of all possible mixed strategies for player p by $\mathcal{M}_p$. The two players initially choose their mixed strategies (σ_R, σ_B) and, based on that, select their actions. Denoting any realized initial allocation (action) of the players by (a_R, a_B), let $\mathcal{V}(a_R) := \{i \in [n] : a_{Ri} > 0\}$, $\mathcal{V}(a_B) := \{i \in [n] : a_{Bi} > 0\}$, and $\mathcal{V}(a_R, a_B) = \mathcal{V}(a_R) \cup \mathcal{V}(a_B)$ be the set of vertices invested by player R, player B, and both players, respectively. For any fixed initial budget allocation (a_R, a_B), the initial state s_i of vertex $i \in [n]$ is in $\{B, R\}$, if and only if $s_i \in \mathcal{V}(a_R \cup a_B)$. Moreover, $s_i = R$ with probability $\frac{a_{Ri}}{a_{Ri}+a_{Bi}}$, and $s_i = B$ with probability $\frac{a_{Bi}}{a_{Ri}+a_{Bi}}$. Following the initial allocation of seeds, the stochastic diffusion rule determines how these R and B infections propagate over the network $\mathcal{G}$. The adoption process is based on a discrete finite time model, where the state of a vertex i at time instant $t = 1, 2, \ldots, T$ is denoted by $s_{it} \in \{U, R, B\}$, where U stands for Uninfected, R stands for infection by red player, and B stands for infection by blue player.

Stochastic Adoption Rule: There is an update schedule which determines the order in which vertices are considered for state updates, assuming that once an uninfected node becomes infected, it will not change its state in the future. While one can consider a variety of updating rules within this framework, the one which we consider here is based on the so-called *switching-selection* model (Goyal et al. 2014). This is a two-stage adoption rule where first it is determined whether an uninfected node will switch to an infected node, and if yes, the type of infection will be selected. More precisely, let us denote the fraction of node i's neighbors infected by R and B at the time of update by α_R and α_B, respectively. Moreover, let us consider two functions: An increasing function $f(\cdot)$ (the switching function) which maps $\alpha_R + \alpha_B$ to $[0, 1]$ with $f(0) = 0, f(1) = 1$. A selection function $g(\cdot)$ which maps $\frac{\alpha_R}{\alpha_R+\alpha_B}$ to $[0, 1]$ such that $g(0) = 0, g(1) = 1$, and $g(y)+g(1-y) = 1, \forall y \in [0, 1]$. One can think of the switching function f as specifying how rapidly adoption increases with the fraction of neighbors who have adopted, regardless of their type R or B, while the selection function g specifies the probability of infection by each type given that a node is susceptible to infection. Given these switching-selection functions which are the same across all the vertices, the stochastic adoption process takes place as follows:

- With probability $f(\alpha_R + \alpha_B)$, node i becomes infected, and with probability $1 - f(\alpha_R + \alpha_B)$ node i remains uninfected and the update ends.
- If node i is determined to be infected, it will adopt type R with probability $g(\frac{\alpha_R}{\alpha_R+\alpha_B})$ and type B with probability $g(\frac{\alpha_B}{\alpha_R+\alpha_B})$.

Remark 4. One can describe the two-step adoption rule using switching-selection functions f and g by only a single-step update rule and using a *generalized adoption* function h with range $[0, 1]$. This is because for local fractions of red and blue neighbors α_R and α_B, if we denote the probability of adopting R by $h(\alpha_R, \alpha_B)$ (symmetrically the probability of adopting B by $h(\alpha_B, \alpha_R)$), then we simply have $f(\alpha_R + \alpha_B) = h(\alpha_R, \alpha_B) + h(\alpha_B, \alpha_R)$, and $g(\frac{\alpha_R}{\alpha_R+\alpha_B}) = \frac{h(\alpha_R,\alpha_B)}{h(\alpha_B,\alpha_R)+h(\alpha_R,\alpha_B)}$. In other words, having the generalized adoption function h, one can always decompose h into a two-step process by defining the switching-selection functions as above. Therefore, from now we only work with a single-step update rule using the generalized adoption function h.

Payoff Functions: As we described above, the two players simultaneously choose some number of vertices to initially invest (seed). After that, the above stochastic adoption rule unfolds which determines how each player's type spreads over the network. Here, the final goal for each player is to find the best initial investment strategy in order to maximize its (expected) total number of final adoptions. More precisely, for $p \in \{R, B\}$, let X_p be a random variable denoting the number of infected nodes by player p at the end of the process. Given the initial seeding strategy (σ_R, σ_B), the expected payoff of player p is given by $u_p(\sigma_R, \sigma_B) = \mathbb{E}[X_p|(\sigma_R, \sigma_B)]$, where the expectation is with respect to all the randomization in the players' strategies for choosing initial seeds, as well as the randomization of the stochastic adoption process. Finally, a mixed-strategy profile (σ_R^*, σ_B^*) constitutes a mixed-strategy Nash equilibrium for the competitive contagion game if $\sigma_p^*, p \in \{R, B\}$ maximizes player p's payoff given the strategy of the other player σ_{-p}^*. Note that here we are considering *mixed-strategy* NE whose existence is already guaranteed by Nash's theorem (Başar and Olsder 1999).

One of our goals for studying the above competitive contagion game is to see how the choice of generalized adoption function (or switching-selection functions) will affect the efficiency of the resulting Nash equilibrium points. To this end, we again leverage the notion of PoA adapted for the competitive contagion game given by

$$PoA := \frac{\max_{(a_R,a_B)\in \mathcal{A}_R\times\mathcal{A}_B} \mathbb{E}[X_R + X_B|(a_R, a_B)]}{\min_{(\sigma_R,\sigma_B)\in NE} \mathbb{E}[X_R + X_B|(\sigma_R, \sigma_B)]},$$

where NE denotes the set of all mixed-strategy Nash equilibria of the contagion game.

Definition 8. We say a generalized adoption function h is competitive if for all $\alpha_R, \alpha_B \in [0, 1]$, we have $h(\alpha_R, \alpha_B) \le h(\alpha_R, 0)$. In other words, a player always has equal or higher infection probability in the absence of the other player. Moreover, we say it is additive if $h(\alpha_R, \alpha_B) + h(\alpha_B, \alpha_R) = f(\alpha_R + \alpha_B)$ for some increasing function f, i.e., h permits interpretation as a switching function.

It is worth noting that based on the above definition, switching-selection formulation as a special case of generalized adoption formulation always satisfies the additivity property. As an example, in the switching-selection formulation, if f is concave and g is a linear function, we automatically get both additivity and competitiveness. We only need to check the competitiveness of h. Since $f(\alpha_R + \alpha_B) = h(\alpha_B, \alpha_R) + h(\alpha_R, \alpha_B)$, we can write

$$\frac{\alpha_R}{\alpha_R + \alpha_B} = g(\frac{\alpha_R}{\alpha_R + \alpha_B}) = \frac{h(\alpha_R, \alpha_B)}{h(\alpha_B, \alpha_R) + h(\alpha_R, \alpha_B)} = \frac{h(\alpha_R, \alpha_B)}{f(\alpha_R + \alpha_B)},$$

where the first equality is by linearity of g and the second equality is due to Remark 4. Thus we have $h(\alpha_R, \alpha_B) = \frac{\alpha_R f(\alpha_R+\alpha_B)}{\alpha_R+\alpha_B}$. Therefore, competitiveness of h is equivalent to showing that

$$h(\alpha_R, \alpha_B) = \frac{\alpha_R f(\alpha_R + \alpha_B)}{\alpha_R + \alpha_B} \le f(\alpha_R) = h(\alpha_R, 0), \quad \forall \alpha_R, \alpha_B \in [0, 1],$$

or $\frac{f(\alpha_R+\alpha_B)}{\alpha_R+\alpha_B} \leq \frac{f(\alpha_R)}{\alpha_R}$ which clearly holds by concavity of f. Now we are ready to state the main result of this subsection:

Theorem 13. *Given any network $\mathcal{G}$ with an additive and competitive generalized adoption function $h(\cdot)$, the PoA of the competitive contagion game is at most* 4.

We only sketch here the main steps in proving the above theorem. The result uses the following two intermediate statements which can be shown using a coupling technique between two random processes:

(1) Given an additive and competitive generalized adoption function h, for *any* initial seeding profile a_R, a_B, we have $\mathbb{E}[X_R|(a_R, a_B)] \leq \mathbb{E}[X_R|(a_R, \emptyset)]$ and $\mathbb{E}[X_B|(a_R, a_B)] \leq \mathbb{E}[X_B|(\emptyset, a_B)]$. In other words, the utility of each player without the presence of the other one is nondecreasing. Thus, by linearity of expectation, $\mathbb{E}[X_R + X_B|(a_R, a_B)] \leq \mathbb{E}[X_B|(\emptyset, a_B)] + \mathbb{E}[X_R|(a_R, \emptyset)]$.
(2) For an additive generalized adoption function h and *any* initial seeding profile a_R, a_B, the total payoffs for both players must be at least that for the red player alone, i.e., $\mathbb{E}[X_R + X_B|(a_R, a_B)] \geq \mathbb{E}[X_R|(a_R, \emptyset)]$.

Now using (1) and (2), we argue how to show $PoA \leq 4$. Let (a_R^o, a_B^o) be the optimal initial seeding action and (a_R^*, a_B^*) be an arbitrary pure-strategy NE (the extension to the mixed-strategy NE can be done analogously). Also, without loss of generality, assume that the red player has the larger budget, i.e., $k_R = |a_R^o| = |a_R^*| \geq |a_B^*| = |a_B^o| = k_B$. Then

$$\begin{aligned}\mathbb{E}[X_R|(a_R^o, a_B^*)] + \mathbb{E}[X_B|(a_R^o, a_B^*)] &= \mathbb{E}[X_R + X_B|(a_R^o, a_B^*)] \\ &\geq \mathbb{E}[X_R|(a_R^o, \emptyset)] \geq \frac{1}{2}\mathbb{E}[X_R + X_B|(a_R^o, a_B^o)],\end{aligned}$$

where equality is by linearity of expectation, the first inequality is by (2), and the last inequality is by (1) and the fact that $\mathbb{E}[X_R|(a_R^o, \emptyset)] \geq \mathbb{E}[X_B|(\emptyset, a_B^o)]$ (as red player has the larger initial budget). Therefore, we have either $\mathbb{E}[X_R|(a_R^o, a_B^*)] \geq \frac{1}{4}\mathbb{E}[X_R + X_B|(a_R^o, a_B^o)]$ or $\mathbb{E}[X_B|(a_R^o, a_B^*)] \geq \frac{1}{4}\mathbb{E}[X_R + X_B|(a_R^o, a_B^o)]$. If the first case holds, then using the definition of a NE, we have

$$\mathbb{E}[X_R+X_B|(a_R^*, a_B^*)] \geq \mathbb{E}[X_R|(a_R^o, a_B^*)] \geq \frac{1}{4}\mathbb{E}[X_R + X_B|(a_R^o, a_B^o)]$$

which shows $PoA \leq 4$. Otherwise, if $\mathbb{E}[X_B|(a_R^o, a_B^*)] \geq \frac{1}{4}\mathbb{E}[X_R + X_B|(a_R^o, a_B^o)]$, we can write

$$\mathbb{E}[X_R+X_B|(a_R^*, a_B^*)] \geq \mathbb{E}[X_B|(\emptyset, a_B^*)] \geq \mathbb{E}[X_B|(a_R^o, a_B^*)] \geq \frac{1}{4}\mathbb{E}[X_R+X_B|(a_R^o, a_B^o)]$$

where the first inequality is by property (2) and the second inequality is due to property (1). This again shows that $PoA \leq 4$.

Finally, we briefly describe how one can establish the statements in (1) or (2). This can be done using a coupling argument, where the idea is to consider two independent different random processes and couple them together such that at each stage, the coupled process mimics the adoption dynamics of each of the individual processes. The key point is that while such a coupling generates the same marginal sample paths as the individual processes, it will allow us to compare their outcomes. For instance, to show that $\mathbb{E}[X_R|(a_R, a_B)] \leq \mathbb{E}[X_R|(a_R, \emptyset)]$, let the solo red process simulate the adoption process when only the red player with initial seeding a_R is present in the game. Moreover, let the joint process be the one in which both red and blue players with initial seeding actions (a_R, a_B) are present. Define a coupled process $< U_v, W_v >$ where U_v faithfully represents the state of a vertex v in the solo red process and W_v represents the state of a vertex in the joint process and such that $W_v = R$ implies $U_v = R$. Existence of such a coupling guarantees that for any update-by-update basis in any run or sample path of the coupled dynamics, the number of red-infected nodes of the solo red process is not less than that in the joint process which must also hold in expectation over runs, yielding the statement. In fact, the additive and competitiveness properties of the generalized adoption function h allow one to show the existence of such a coupling. All the other statements in (1) or (2) can also be shown using a similar coupling argument where more details can be found in Goyal et al. (2014).

Although Theorem 13 provides a positive result which shows that under additive and competitive adoption rule, the lack of coordination between the two players does not hurt much the social welfare; unfortunately this property no longer holds once we consider more general adoption functions. The following theorem provides one such negative result:

Theorem 14. *Let the selection function be linear $g(y) = y$ and $\delta \in (0, 1)$. Then for any $M > 0$ and either of the following non-concave switching functions*

$$f(x) = \begin{cases} 0 & \text{if } x \leq \delta, \\ 1 & \text{if } x > \delta, \end{cases} \qquad\qquad f(x) = x^r, r > 1,$$

there exists a network $\mathcal{G}$ for which $PoA > M$.

In fact, Theorem 14 suggests that concavity of the switching function is very critical for having a bounded PoA. In other words, even a slight convexity in the structure of the switching function can result in an unbounded PoA. Therefore, understanding the influence of switching/selection functions on the PoA of the contagion game would be an interesting problem. In particular, it would be useful to understand the structural properties of the NE points for the competitive contagion game in further detail. Further, the competitive contagion game is given for only two players, namely, red and blue. However, multiplicity of players will bring more complications into the analysis which deserves further investigation to see how the PoA bounds might chance.

4.3 Coordination-Based Diffusion Game

In the previous two models for the network diffusion game, we considered a scenario where the diffusion process at each run is the result of a fixed deterministic or probabilistic law. Situations will arise where the diffusion process itself is a consequence of strategic interactions of the players. In other words, the players themselves are part of the diffusion process. In the following, we consider a coordination-based diffusion game where the agents play a coordination game with their local neighbors in order to decide what product to choose.

Game Model: A game is played in periods $t = 1, 2, \dots$ among a set $[n]$ of players over an undirected weighted network $\mathcal{G} = ([n], \mathcal{E}, \{w_{ij}\})$, where each node represents a player. Each edge $i, j \in \mathcal{E}$ has a weight w_{ij} which shows the influence magnitude of player i on player j, and vice versa. Each player has two alternative actions $\{A, B\}$. As before, we denote player i's action by $a_i \in \{A, B\}$ and the entire actions profile (state) of the game by $\boldsymbol{a} \in \{A, B\}^n$. Given an action profile $\boldsymbol{a}$, the utility of player i is composed of two components: (i) the *individual* component of payoff $v_i(a_i)$ which results from the agent's idiosyncratic preferences for A or B irrespective of other agents and (ii) the *social* component of payoff $\sum_{j \in \mathcal{N}_i} w_{ij} u_j(a_i, a_j)$ resulting from the externalities created by the choices of other agents. Here, $\mathcal{N}_i$ denotes the set of neighbors of player i in the network, and $u_j(a_i, a_j)$ can be viewed as the payoff function of a two-person coordination game between player i and its neighbor j in which each player has the actions A and B. The symmetric payoff matrix of this 2×2 coordination game which is the same for all pairs of players is shown in Fig. 12.2 where we assume that there are increasing returns from coordination. This means that matching the partner's choice is better than not matching, i.e., $\alpha_2 > \alpha_3$, and $\alpha_1 > \alpha_4$. Finally, the total payoff to player i for a given state $\boldsymbol{a}$ equals to

$$U_i(\boldsymbol{a}) = \sum_{j \in \mathcal{N}_i} w_{ij} u_j(a_i, a_j) + v_i(a_i). \tag{12.8}$$

Fig. 12.2 Coordination game matrix between pair of players with $\alpha_2 > \alpha_3$ and $\alpha_1 > \alpha_4$

	A	B
A	α_1, α_1	α_3, α_4
B	α_4, α_3	α_2, α_2

Our goal for analyzing such a game is to understand how each action type A or B propagates as a result of players' interactions over time or whether we can say anything about the limiting behavior and convergence time of their dynamics. Such a coordination-based diffusion game has been addressed in several past works such as Young (2006) or Montanari and Saberi (2010), with proposed solutions mainly based on Ising model in statistical mechanics (to be defined soon).

Example 2. Consider a social network capturing friendship relations among individuals (players) where edge weights represent the strength or weakness of friendships. Assume that each player wants to choose a cell phone contract with either AT&T (A) or Verizon (B). People have different tastes about choosing either of such cell phone service providers. Moreover, as AT&T or Verizon has special offers for their own customers (e.g., free text message between their customers), individuals have more incentive to choose the service provider that most of their friends are using. This determines the social component of players' payoffs which is due to the externalities created by others' decisions. Therefore, the trade-off between individual and social components determines what service provider will be more widespread and how fast it will take place.

Before we get into analysis of the dynamics of the iterative coordination-based diffusion game, we first note that the single-stage game is an exact potential game. This is because for a given state of the game $\boldsymbol{a}$, if we define

$$W_{AA}(\boldsymbol{a}) := \sum_{\{i,j\}\in\mathcal{E}:a_i=a_j=A} w_{ij},$$

$$W_{BB}(\boldsymbol{a}) := \sum_{\{i,j\}\in\mathcal{E}:a_i=a_j=B} w_{ij},$$

$$v(\boldsymbol{a}) = \sum_{i=1}^{n} v_i(a_i),$$

then the function

$$\Phi(\boldsymbol{a}) := (\alpha_1 - \alpha_4)W_{AA}(\boldsymbol{a}) + (\alpha_2 - \alpha_3)W_{BB}(\boldsymbol{a}) + v(\boldsymbol{a}) \tag{12.9}$$

serves as an exact potential function for the single-stage coordination-based diffusion game (one can easily check that $U_i(A, a_{-i}) - U_i(B, a_{-i}) = \Phi(A, a_{-i}) - \Phi(B, a_{-i}), \forall i \in [n]$). Now in order to study the action adoption dynamics over the course of the time $t = 1, 2, \ldots$, let us assume that each player updates his action at random times following a Poisson arrival process. Hence, without loss of generality, we may assume that each player on the average updates once per unit time. These updating processes are assumed to be independent among the players, so that the probability that two players update at the same time is negligible.

We study noisy best response dynamics in this environment, in which when a player updates his action in the next time step he will play his best response with probability close to 1. However, there is still a small probability that the player makes a mistake and chooses the alternative action. More precisely, if a_{-i} denotes the current choice of players other than i, then the probability that i plays action A in the next time step is given by

$$\mathbb{P}\{i \text{ chooses } A\} = \frac{e^{\beta[U_i(A,a_{-i})-U_i(B,a_{-i})]}}{1+e^{\beta[U_i(A,a_{-i})-U_i(B,a_{-i})]}}, \tag{12.10}$$

and $\mathbb{P}\{i \text{ chooses } B\} = 1 - \mathbb{P}\{i \text{ chooses } A\}$. Here, β is the noise parameter where the larger β means the higher chance that the player chooses his best response (i.e., the action with a higher payoff). For instance, $\beta = \infty$ corresponds to the strict best response dynamics. Note that using the potential function (12.9), one can rewrite the updating rule (12.10) as

$$\mathbb{P}\{i \text{ chooses } A\} = \frac{e^{\beta[\Phi(A,a_{-i})-\Phi(B,a_{-i})]}}{1+e^{\beta[\Phi(A,a_{-i})-\Phi(B,a_{-i})]}}, \tag{12.11}$$

which does not depend on the specific payoff function of player i, but rather a global potential function Φ. Therefore, one can view the evolution of the game dynamics based on update rule (12.10) as the evolution of a discrete time Markov chain, where the states correspond to all possible 2^n action profiles $\boldsymbol{a} \in \{A, B\}^n$ and the transition probability of going from state (a_i, a_{-i}) to (A, a_{-i}) is given by (12.10) (or equivalently (12.11)). It is known that the stationary distribution of such a Markov chain with logistic transition probabilities (12.11) is given by the following Gibbs distribution:

$$\pi^{\beta}(\boldsymbol{a}) := \frac{e^{\beta\Phi(\boldsymbol{a})}}{\sum_{\hat{\boldsymbol{a}}\in\{A,B\}^n} e^{\beta\Phi(\hat{\boldsymbol{a}})}},$$

which is the long-run relative frequency of each state $\boldsymbol{a}$. In fact, one can show that when β is sufficiently large (i.e., there is little noise in the adjustment process), the long-run distribution will be concentrated almost entirely on the states with high potential, which are referred to as *stochastically stable* states (Young 2006).

In what follows, we study the convergence time of the game dynamics to its Gibbs stationary distribution. In other words, we want to see, starting from any arbitrary initial state of the game, how long it will take until the diffusion process comes close to the stochastically stable states. In fact, one can show that the convergence time to stochastically stable states heavily depends on the structure of the underlying network. Therefore, in order to provide more concrete results, we restrict our attention to a special class of graphs known as *close-knit family* which is defined next:

Definition 9. Let $\mathcal{S} \subseteq [n]$ be a nonempty subset of nodes of a graph $\mathcal{G}$ without any isolated vertex. $\mathcal{S}$ is called r-close-knit if

$$\min_{\mathcal{S}' \subseteq \mathcal{S}} \frac{d(\mathcal{S}, \mathcal{S}')}{\sum_{i \in \mathcal{S}'} d_i} \geq r,$$

where $r > 0$ is a positive constant, $d(\mathcal{S}, \mathcal{S}')$ denotes the number of edges $\{i, j\}$ in $\mathcal{G}$ such that $i \in \mathcal{S}$ and $j \in \mathcal{S}$, and d_i denotes the degree of node i in graph $\mathcal{G}$.

Definition 10. Given $r \in (0, \frac{1}{2})$ and $k \in \mathbb{Z}^{>0}$, a graph $\mathcal{G}$ is (r, k)-close-knit if every node belongs to a subset of nodes of size at most k that is at least r-close-knit. A family of graphs $\mathcal{F}$ is close-knit if for every $r \in (0, \frac{1}{2})$, there exists a positive integer k_r such that every graph in $\mathcal{F}$ is (r, k_r)-close-knit.

As an example one can easily check that the families of cycles or square lattices are close-knit. Since the sizes of graphs in a close-knit family are different, to provide a general convergence time bound for all the graphs in such a family, in the following we shall assume that all the graph edges have unit weight and the individual components of all the payoffs are zero. This means that players have no individual preference over A or B, and the externalities determine the final outcome. In such a case, the potential function becomes $\Phi(\boldsymbol{a}) = (\alpha_1 - \alpha_4)W_{AA}(\boldsymbol{a}) + (\alpha_2 - \alpha_3)W_{BB}(\boldsymbol{a})$, which is maximized in a state where all the players choose A if $\alpha_1 - \alpha_4 > \alpha_2 - \alpha_3$ or in a state where all the players choose B if $\alpha_2 - \alpha_3 > \alpha_1 - \alpha_4$. That is, the adoption process will select either of these two states in the long run, which are called *risk-dominant equilibrium*. Next, the following definition provides a metric to measure the convergence time to stochastically stable risk-dominant equilibria.

Definition 11. Given $\delta \in (0, 1)$, noise parameter β, graph $\mathcal{G}$, and the initial state of the game $\boldsymbol{a}(0)$, we let $T(\beta, \mathcal{G}, \boldsymbol{a}(0), \delta)$ be the expected value of the first time t such that, starting from the initial state $\boldsymbol{a}(0)$, with probability of at least $1 - \delta$, at least $1 - \delta$ fraction of the players are using risk-dominant strategy at time t and thereafter. Finally, the δ-*inertia* of the process is the maximum of the above expected time over all initial states, i.e., $T(\beta, \mathcal{G}, \delta) := \max_{\boldsymbol{a}(0) \in \{A,B\}^n} T(\beta, \mathcal{G}, \boldsymbol{a}(0), \delta)$.

Based on the above definition of δ-inertia, the following theorem provides a uniform upper bound on the waiting time until the adoption process comes close to a risk-dominant equilibrium (the state having maximum potential), provided that the network belongs to a close-knit family.

Theorem 15. *Given $\delta > 0$, a close-knit family of graphs $\mathcal{F}$, and a 2×2 coordination matrix game with risk-dominant equilibrium, there exists β_δ such that for every $\beta \geq \beta_\delta$, the δ-inertia of the adoption process in the coordination-based diffusion game is uniformly bounded from above for all the graphs in family $\mathcal{F}$.*

In fact, a more generalized version of the above theorem for a slightly different setting has been given in Montanari and Saberi (2010) where the convergence rate of the adoption process to the risk-dominant equilibrium has been characterized as a function of the network structure. For instance, it has been shown that innovation spreads much more slowly on well-connected network structures dominated by long-range links than in low-dimensional ones dominated, for example, by geographic proximity.

As we close this subsection, we discuss briefly an alternative way of capturing the network structure into the diffusion game process. There are many situations in which the network structure is not fully known, but some of its statistics (e.g., the degree distribution) is a common knowledge to all the players. In such cases one can again define coordination-based adoption rules for propagation of innovation over this abstract network of players in which the individuals only have access to the common statistics of the network. Under such a setting, *Bayesian Nash equilibrium* seems to be a more appropriate solution concept to determine the long-run behavior of the diffusion process, as now the players are also faced with the uncertainty of their neighbors and of their numbers. We refer to Jackson and Yariv (2007) for one such formulation, with several results on the stability and other properties of the equilibrium points.

5 Network Formation Games

In all the network games that we have seen so far, the underlying assumption has been that the network has a fixed topology, and the players pick their actions over that topology. In other words, the players' actions over the course of game dynamics do not change the network structure. In this section, we relax this assumption by considering a class of network games where the network structure itself is a consequence of players' actions. Such games are especially valuable in understanding the nature of social, economic, and Internet-like networks where the connections between individuals are subject to change as a result of their interactions. The concept of network formation game was first introduced by Jackson and Wolinsky (1996) (see also Jackson and Watts 2002). Since then various models have been proposed to capture the essence of network evolution in the presence of strategic agents. In the following we consider a basic model for network formation games, which was first proposed by Fabrikant et al. (2003) and has received considerable attention in the past decade.

A Basic Network Creation Game: Consider a set $[n] = \{1, 2, \ldots, n\}$ of n players where an action for player i, denoted by a_i is a subset of $[n] \setminus \{i\}$. Therefore, the action set of player i is $\mathcal{A}_i = 2^{[n]\setminus\{i\}}$. Given an action profile $\boldsymbol{a} = (a_1, \ldots, a_n)$, we let the graph $\mathcal{G}[\boldsymbol{a}]$ be an undirected network with vertex set $[n]$ and edge set $\cup_{i=1}^{n}\{i\} \times a_i$. In other words, a player's action determines the set of nodes that player i is connected to in $\mathcal{G}[\boldsymbol{a}]$. Given a network structure $\mathcal{G}[\boldsymbol{a}]$ formed as a result of players' actions $\boldsymbol{a}$, the cost of player i is given by

$$c_i(\boldsymbol{a}) = \alpha|a_i| + \sum_{j=1}^{n} d_{\mathcal{G}[\boldsymbol{a}]}(i, j), \tag{12.12}$$

where $d_{\mathcal{G}[\boldsymbol{a}]}(i, j)$ denotes the graphical distance between players i and j in the graph $\mathcal{G}[\boldsymbol{a}]$ and $\alpha > 0$ is the cost of creating an extra edge. The idea behind this formulation is that agents aim to bear the cost of least possible number of edges, but at the same time, they want to achieve a good connection to all other agents in the network.

As we will see soon, the network creation game always admits a pure-strategy NE, although due to the nonconstructiveness of the Nash equilibrium, no distributed algorithm for finding such networks is known (later we will take a more constructive approach to find pure-strategy NE points by restricting the players' feasible actions). In fact, despite the simple structure of this game, even computing the best response for a player is an NP-hard problem. To see this more clearly, let us assume that $\alpha \in (1, 2)$ and consider the network formed by all other players other than i. Assuming that player i has no incoming edge, he has to decide what subset of nodes to choose and connect to in order to minimize his cost c_i. Since $\alpha < 2$, in the best response of player i, every node j must be at most two hops away from i. Otherwise, player i can make a link with j and reduce his cost. This means that the best response of player i must be a dominating set for the rest of the graph.[7] Moreover, since $\alpha > 1$, having more than the required number of edges would only increase player i's cost. Therefore, the best response for player i is to choose a minimum size dominating set of the rest of the graph and connect to them. As finding the minimum size domination set is an NP-complete problem, so is finding the best response for player i.

Despite the hardness of computing best response strategies, one can easily check that the following networks form a pure-strategy NE for different ranges of $\alpha > 0$.

- $\alpha < 1$: In this case the complete graph is the unique NE.
- $\alpha \geq 1$: In this case the star graph is a NE (although not the only one).

5.1 PoA of the Network Creation Game

Here, we turn our attention to the PoA analysis of this basic network creation game. Let us first consider a simple case when $\alpha \geq n^2$ is very large. In this case, one can easily check that every pure-strategy NE must have a tree structure. This is because first of all, the equilibrium network must be connected; otherwise, two players belonging to different components will have infinite cost which incentivize them to form a link. Moreover, no player will be involved in a cycle; otherwise,

[7]A dominating set of a graph is a subset of its vertices such that each edge has at least one end point in that set.

that player can deviate by severing one of his links to exit that cycle and reduce his cost by at least $\alpha - (1 + 2 + \ldots + (n-1)) > 0$. This shows that for a wide range of parameters, the equilibrium structure is a tree. Therefore, it is important to understand the PoA for tree structures. The following theorem says that indeed tree Nash equilibrium structures have small PoA (Fabrikant et al. 2003).

Theorem 16. *For any tree Nash equilibrium, we have $PoA \leq 5$.*

Now that we know that the PoA for the tree Nash equilibrium points is bounded above by a constant, we may use this fact to obtain constant PoA for other ranges of α in which the pure-strategy NE is not necessarily a tree structure. In fact, it seems feasible to establish constant upper bounds for a wide range of values for the parameter α, where the idea is to upper bound the social cost of an arbitrary Nash equilibrium graph by its underlying breadth-first search (BFS) trees which we know has a small cost. In the following theorem, we provide one such result for the case when $\alpha \leq \sqrt{\frac{n}{2}}$, which was given by Demaine et al. (2007):

Theorem 17. *For $\alpha \leq \sqrt{\frac{n}{2}}$, the PoA of the network creation game is at most 6.*

To see how one can show a constant upper bound for this range of values of α, let us consider an arbitrary NE of the network formation game, $\boldsymbol{a}^*$. We show that for any arbitrary vertex v_0 in $\mathcal{G}[\boldsymbol{a}^*]$, we have

$$c(\boldsymbol{a}^*) \leq 2\alpha(n-1) + nDist_{\boldsymbol{a}^*}(v_0) + (n-1)^2, \tag{12.13}$$

where $Dist_{\boldsymbol{a}^*}(v_0) := \sum_{j=1}^{n} d_{\mathcal{G}[\boldsymbol{a}^*]}(v_0, j)$. In the equilibrium graph $\mathcal{G}[\boldsymbol{a}^*]$, let us consider a breadth-first search (BFS) tree with root v_0 and denote it by $T(v_0)$. This tree is one that starts at the root v_0 and explores the neighbor nodes first, before moving to the next level; it is just a layered version of $\mathcal{G}[\boldsymbol{a}^*]$ with distinguished tree edges. Let us consider an arbitrary vertex $i \neq v_0$ and denote the number of tree edges built by i by n_i. Now if player i deviates by severing all of its non-tree edges in the equilibrium and connect itself to the root v_0, its cost will be at most $(n_i + 1)\alpha + Dist_{\boldsymbol{a}^*}(v_0) + (n-1)$. This is because node i has exactly $n_i + 1$ neighbors in this new strategy. Moreover, since only non-tree edges were deleted, $Dist_{\boldsymbol{a}^*}(v_0)$ is not affected by i's new strategy. The new edge between i and v_0 ensures that the shortest path distance between i and any other vertex j is by at most 1 larger than the shortest path distance between v_0 and j. Hence, the distance cost of i to all other nodes after its deviation is at most $Dist_{\boldsymbol{a}^*}(v_0) + (n-1)$. Now, since the deviated cost of player i cannot be more than its equilibrium cost, we have $c_i(\boldsymbol{a}^*) \leq (n_i+1)\alpha + Dist_{\boldsymbol{a}^*}(v_0) + (n-1)$. Summing this inequality for all $i \neq v_0$, and noting that $\sum_{i \neq v_0}(n_i + 1) + n_{v_0} = 2(n-1)$ (since it counts the edges of $T(v_0)$ exactly twice), we arrive at the inequality (12.13). Now given this inequality, one can show that if the BFS tree $T(v_0)$ has depth d, i.e., it has d different levels, then the PoA of the $\mathcal{G}[\boldsymbol{a}^*]$ is at most $d+1$ (i.e., one more than the depth of its underlying BFS tree). This is simply because for a d-depth tree, $Dist_{\boldsymbol{a}^*}(v_0) \leq (n-1)d$. On the other hand, for $\alpha \leq \sqrt{\frac{n}{2}}$, one can easily check that the optimal social cost is achieved for the star graph, which is at least $\alpha(n-1) + n(n-1)$. Thus, the PoA is at most

$$\frac{2\alpha(n-1) + n(n-1)d + (n-1)^2}{\alpha(n-1) + n(n-1)} \leq d + 1.$$

Therefore, we only need to show that for $\alpha \leq \sqrt{\frac{n}{2}}$, the depth of the BFS tree in $\mathcal{G}[\boldsymbol{a}^*]$ is at most 5. For this purpose, denote the set of all vertices in the equilibrium graph $\mathcal{G}[\boldsymbol{a}^*]$ that have distances at most 2 from v_0 by $\mathcal{N}_2(v_0)$. One can then show that $|\mathcal{N}_2(v_0)| > \frac{n}{2\alpha}$, since, otherwise, player v_0 can connect itself to one of the nodes in $\mathcal{N}_2(v_0)$ and reduce its cost. Now suppose that, to the contrary, there is a vertex i at a distance of at least 6 from v_0. Then i can connect to player v_0 to decrease its distance to all the nodes in $\mathcal{N}_2(v_0)$ by at least 1. Because player i has not established such a link, we conclude that $\alpha > |\mathcal{N}_2(v_0)|$. Since we already know that $|\mathcal{N}_2(v_0)| > \frac{n}{2\alpha}$, we must have $\alpha > \frac{n}{2\alpha}$, which implies $\alpha > \sqrt{\frac{n}{2}}$. This contradiction shows that node v_0 is at a distance of at most 5 from every other node in the equilibrium network $\mathcal{G}[\boldsymbol{a}^*]$. Thus, the depth of $T(v_0)$ is at most 5, which implies that $PoA \leq 5 + 1 = 6$.

In fact there have been substantial work to study the PoA of the network creation game for other range of values of the parameter α, which mainly uses the BFS tree structure in the Nash equilibrium graph to upper bound its social cost. Without going into more details, we only mention here that constant upper bound for the PoA of the network creation game is known for all range of $\alpha \in \mathbb{R}^{\geq 0} \setminus (n^{1-\epsilon}, 12n \log n)$, where ϵ can be any fixed number in $(0, 1)$. However, for the range $\alpha \in (n^{1-\epsilon}, 12n \log n)$, the best known upper bound on the PoA is the sublinear bound of $O(2^{\sqrt{\log n}})$ given by Demaine et al. (2007).

5.2 Local Search for Network Creation Game

As we have seen in the previous subsection, one of the limitations of the network creation game is nonexistence of a constructive method for finding its pure-strategy NE structures. In this subsection we study the dynamic behavior of a slightly different version of this game with local search action space and show that under specific assumptions, players can selfishly improve upon their actions until a pure-strategy NE is achieved. The results of this section are based on Alon et al. (2013), Lenzner (2011), and Kawald and Lenzner (2013).

Definition 12. In the *swap network creation game* or simply swap game (SG), an admissible action for player i in the network $\mathcal{G}[\boldsymbol{a}]$ is to choose a subset of verticies $\hat{a}_i$ such that $|a_i| = |\hat{a}_i|$ and $|a_i \cap \hat{a}_i| = |a_i|$.

Intuitively, admissible actions in the SG are actions which replace one neighbor j of i in the graph $\mathcal{G}[\boldsymbol{a}]$ by another vertex k. This corresponds to "swapping" the edge $\{i, j\}$ from j toward k, which constitutes replacement of edge $\{i, j\}$ by edge $\{i, k\}$.

Note that in the SG game, the costs of the players are defined as before (12.12), and the only difference between the SG game and the basic network creation game is the restriction on the feasible moves of the players.

Theorem 18. *Starting from a tree network, the SG game is a generalized ordinal potential game. In particular, any sequence of better responses by players will converge after at most $O(n^3)$ steps to a pure-strategy NE.*

First, we note that starting from a tree T as the initial network, after each swap action by a player, the resulting network will still be a tree. Now, let us consider an arbitrary player i and assume that he swaps one of his edges from $\{i, j\}$ to $\{i, k\}$. Let T' and T'' be the two sub-trees obtained after removing the edge $\{i, j\}$ from T such that $j \in T'$ and $i \in T''$. Then one can see that Δc_i, the difference in player i's cost after and before this swap, equals to

$$\Delta c_i = \sum_{v \in T'} d_{T'}(j, v) - \sum_{v \in T'} d_{T'}(k, v). \tag{12.14}$$

This is simply because player i has the same number of edges in both T and $T \setminus \{i, j\} \cup \{i, k\}$, and hence his edge self-costs before and after this deviation are the same. The only difference is in the distance term of his cost which precisely equals to (12.14).

Next let us define a generalized ordinal potential function $\Phi := \sum_{i=1}^{n} c_i$ to be the social cost. We invoke how the edge swapping of player i affects this function. Note that any node in the sub-tree, T' will feel the same distance cost before and after this swap. However, any node in the sub-tree T'' will feel a difference of exactly Δc_i in its cost after the swap. Therefore, denoting the number of vertices in T'' by $|T''|$, one can see that the difference in the potential function after and before the swap, denoted by $\Delta\Phi$, equals to $\Delta\Phi = |T''|\Delta c_i$. This shows that Φ is a generalized ordinal potential function for the SG, and hence any sequence of better responses must terminate at a pure-strategy NE. Moreover, from all initial trees with n nodes, a path of length $n-1$ has the maximum potential of $2\alpha(n-1)+\frac{n^3-n}{3}$. Since after every strictly better response by a player the value of the potential function decreases by at least 1, and this function is bounded below by $2\alpha(n-1)$, after at most $O(n^3)$ steps, the better response dynamics will terminate at a pure-strategy NE.

In fact, one can take one step further to show that starting from a tree network, the only unique equilibrium in the SG must be the star network. Otherwise, in the equilibrium tree, there exist four nodes i_1, i_2, i_3, and i_4 connected with a path such that the distance between i_1 and i_4 is exactly 3. Now one can show that either i_1 can swap its edge from $\{i_1, i_2\}$ to $\{i_1, i_3\}$ and reduce its cost, or if not, i_4 can reduce its cost by swapping $\{i_4, i_3\}$ and $\{i_4, i_2\}$. This shows that any NE must be a tree of diameter 2, i.e., a star graph.

It is worth noting that Theorem 18 holds only if the game is played over the tree. However, for general networks this result no longer holds, as one can show that the sequence of best responses may cycle. Therefore, the SG over general networks does not admit a generalized potential function. Thus, any treatment of the SG dynamics on general graphs requires fundamentally different techniques. We also mention here that one can use similar analysis as in the case of basic network creation game to obtain upper bounds for the PoA of the SG within various ranges of values of α, (Alon et al. 2013). As a final remark, one can relax the swapping assumption in the SG to the case where players are allowed to swap *multiple* edges at a time. In this case, and starting from a tree network, one can use structural insights to obtain a linear-time algorithm for computing the best response of players. By contrast, computing a best response in general graphs can be shown to be an NP-hard problem (by reduction to p-median problem Lenzner 2011), if more than one edge can be swapped at a time.

5.3 Coordination-Based Network Formation Game

In this subsection we discuss an alternative formulation for network formation games. In the basic network creation game or SG that we have addressed earlier, the action set of the players were restricted to the subset of nodes that they want to form a link with. However, one can generalize these action sets to the case when not only a player selects who to be connected with, but also they will play a game with those neighbors to maximize their payoffs (Goyal and Vega-Redondo 2000; Jackson and Watts 2002). The following formulation proposed by Jackson and Watts (2002) provides one of such instances.

Coordination & Network formation: Consider a set $[n]$ of n players where each player i has an action set of the form $\mathcal{A}_i := 2^{[n]\setminus\{i\}} \times \{A, B\}$, where A and B are two alternatives for each player. These n players play over a network which is determined by the first coordinate of their actions. More precisely, let us denote the action profile of player i by $a_i = (s_i, r_i)$, where $s_i \subseteq [n] \setminus \{i\}$, and $r_i \in \{A, B\}$. Given the action profile $\boldsymbol{a}$, the game network is defined to be $\mathcal{G}_{\boldsymbol{a}} := ([n], \cup_{i=1}^{n}\{i\} \times s_i)$, and the payoff to player $i \in [n]$ is given by

$$u_i(\boldsymbol{a}) = \sum_{j \neq i} \mathbb{I}_{ij}(\mathcal{G}_{\boldsymbol{a}})[v_i(r_i, r_j) - f(n_i)], \tag{12.15}$$

where $v_i(r_i, r_j)$ is a payoff that depends on the actions chosen (for instance, it can be determined based on a two-player coordination game with a 2×2 symmetric payoff matrix as in Fig. 12.2). $\mathbb{I}_{ij}(\mathcal{G}_{\boldsymbol{a}})$ is the edge indicator function where $\mathbb{I}_{ij}(\mathcal{G}_{\boldsymbol{a}}) = 1$ if there exists an edge between i and j in the network $\mathcal{G}_{\boldsymbol{a}}$, and $\mathbb{I}_{ij}(\mathcal{G}_{\boldsymbol{a}}) = 0$, otherwise. Finally, n_i denotes the degree of node i in the formed network $\mathcal{G}_{\boldsymbol{a}}$, and $f(\cdot)$ is a nondecreasing function. In this game, the players first choose who to play with and then play a coordination game with their neighbors in order to maximize their utilities. As an example, in the context of social networks, people not only have freedom to choose their friends, but also they interact with them somehow to maximize their own payoffs.

Now one can consider various dynamics for studying the behavior of such a game when it is played repeatedly by the players. For instance, denoting the action profile of players at time t by $\boldsymbol{a}^t$, at the next period, one player is chosen uniformly at random and updates his strategy myopically, best responding to what the other players with whom he interacts did in the previous period. Also, there is a small probability $\epsilon \in (0, 1)$ that the player chooses another action other than his best response. This induces a Markov chain over the game action profiles (states), which converges to a unique stationary distribution for any initial action profile $\boldsymbol{a}^0$. Following this setting, one can study the stochastically stable equilibrium points of the game, which are the game states having nonzero probability in the steady-state distribution, for arbitrarily small probabilities of trembles ϵ. In the following, we provide one such result which is due to Jackson and Watts (2002):

Proposition 2. *Let $f(\cdot)$ be the constant function, i.e., $f(n_i) = k, \forall n_i$, and assume that the payoff functions $v_i(\cdot, \cdot)$ are determined based on the two-player coordination game with payoff matrix given in Fig. 12.2. Then*

- *If $(\alpha_1-k)(\alpha_2-k) < 0$, then the unique stochastically stable state is the complete graph with all players playing A or B, except when $\frac{\alpha_2-\alpha_3}{\alpha_1-\alpha_4+\alpha_2-\alpha_3} \leq \frac{2}{n-1}$, $\alpha_1-k < 0$, and $\alpha_2 - k < 0$ in which case, the empty graph with all players playing A is also stochastically stable.*
- *If $\alpha_3 - k > 0$ and $\alpha_4 - k > 0$, then the unique stochastically stable state is the complete graph with all players playing A.*

As a final remark, we mention here that one can consider different application arena for network formation games. For example, network formation games also arise in wireless communications, where relay stations (RSs) form tree structures interconnecting them for most efficient connection to a base station Saad et al. (2011). In particular, in this network formation process, RSs may also take into account security considerations (see Saad et al. 2012), where the RSs interact and choose their secure communication paths in the uplink of a wireless multi-hop network, in the presence of eavesdroppers.

6 Some Other Types of Network Games

In this section, we provide a brief overview on various other types of network games. Network games have been widely used in studying traffic assignment problems as well as flow congestion control within the general framework of routing games. Such games have been frequently addressed in the earlier literature (Başar 2007; Menache and Ozdaglar 2011) with many applications in telecommunications (Altman et al. 2006), load balancing (Nisan et al. 2007), smart grid (Etesami et al. 2017), and transportation networks (Krichene et al. 2017). In fact, several important concepts and solutions in routing games such as the *Pigou network*, *Wardrop equilibrium*, and *Braess paradox* have been discussed extensively in another chapter of this *Handbook* related to communication networks (Shakkottai and Srikant 2017).

Network games have also been studied within the context of *aggregative games* in engineering as well as economic literatures (Koshal et al. 2016; Parise et al. 2015). Aggregative games generally refer to a class of games in which every player's payoff is a function of its own strategy and the aggregate (e.g., sum) of all others'. In fact, it can be shown that under certain conditions, the aggregative games are potential games and, hence, admit pure-strategy Nash equilibria (Jensen 2010). *Rent-seeking* games constitute another application domain for network games in which players manipulate or utilize the underlying social or political networks in order to derive the economic activities toward their own benefits (Murray 2012). We also refer to Bramoullé et al. (2014) and Jackson (2005, 2010), and the references therein, for other applications of network games in social and economic networks.

Finally, network games have recently emerged under a new research thrust, that of bioengineering. Such bio-networked games refer to a class of games which are inspired from biological systems. For instance, as we saw earlier in this chapter, one of the motivations for studying diffusion games was epidemics or virus spread, which have biological origins. For more information on such bio-inspired network games and an overview of their applications, we refer interested reader to Altman (2014).

7 Conclusions

This chapter has presented an overview of noncooperative network games with various formulations and applications in different fields. The focus in the chapter has been mainly on algorithmic and computational aspects and for four major types of network games: congestion games, resource allocation games, diffusion games, and network formation games. Connections between these different classes of games as well as to other network games have been discussed. Existence and uniqueness of pure-strategy Nash equilibria have been studied, and whenever possible efficient algorithms for obtaining one such equilibrium point have been provided. In particular, the efficiency of Nash equilibrium under various settings has been studied. The methodologies provided in this chapter provide a rich collection of tools and techniques for analyzing other similar types of network games.

As it has been mentioned in the *Introduction*, there are many other types of network games which are quite relevant to the subject of this chapter. Here we have mainly focused on network games with a finite number of players and discrete action spaces. However, most of the models can be extended to games with continuum action spaces or with a large population (continuum) of players. Some of these topics, and others as pointed out in the chapter, and particularly in the previous section, have been discussed on other chapters of the *Handbook*. We mention in this context the applications of network games in opinion dynamics and social networks (Bolouki et al. 2017), routing games in communication networks (Shakkottai and Srikant 2017), and mean field games (Caines et al. 2017).

References

Ackermann H, Röglin H, Vöcking B (2008) On the impact of combinatorial structure on congestion games. J ACM (JACM) 55(6):25

Alon N, Feldman M, Procaccia AD, Tennenholtz M (2010) A note on competitive diffusion through social networks. Inf Process Lett 110(6):221–225

Alon N, Demaine ED, Hajiaghayi MT, Leighton T (2013) Basic network creation games. SIAM J Discret Math 27(2):656–668

Alpcan T, Başar T (2010) Network security: a decision and game-theoretic approach. Cambridge University Press, Leiden

Altman E (2014) Bio-inspired paradigms in network engineering games. J Dyn Games (JDG) 1(1):1–15

Altman E, Boulogne T, El-Azouzi R, Jiménez T, Wynter L (2006) A survey on networking games in telecommunications. Comput Oper Res 33(2):286–311

Anantharam V (2004) On the Nash dynamics of congestion games with player-specific utility. In: IEEE conference on decision and control, vol 5. IEEE, pp 4673–4678

Awerbuch B, Azar Y, Epstein A, Mirrokni VS, Skopalik A (2008) Fast convergence to nearly optimal solutions in potential games. In: Proceedings of the 9th ACM conference on electronic commerce. ACM, pp 264–273

Başar T (2007) Control and game-theoretic tools for communication networks. Appl Comput Math 6(2):104–125

Başar T, Olsder GJ (1999) Dynamic noncooperative game theory. SIAM, Philadelphia

Blum A, Even-Dar E, Ligett K (2006) Routing without regret: on convergence to Nash equilibria of regret-minimizing algorithms in routing games. In: Proceedings of the twenty-fifth annual ACM symposium on principles of distributed computing. ACM, pp 45–52

Bramoullé Y, Kranton R, D'amours M (2014) Strategic interaction and networks. Am Econ Rev 104(3):898–930

Bolouki S, Nedić A, Başar T (2017) Social networks. In: Başar T, Zaccour G (eds) Handbook of dynamic game theory. Springer International Publishing, Cham

Caines PE, Huang M, Malhamé RP (2017) Mean field games. In: Başar T, Zaccour G (eds) Handbook of Dynamic Game Theory. Springer International Publishing, Cham

Chun B-G, Chaudhuri K, Wee H, Barreno M, Papadimitriou CH, Kubiatowicz J (2004) Selfish caching in distributed systems: a game-theoretic analysis. In: Proceedings of the twenty-third annual ACM symposium on principles of distributed computing. ACM, pp 21–30

Correa JR, Schulz AS, Stier-Moses NE (2004) Selfish routing in capacitated networks. Math Oper Res 29(4):961–976

Daskalakis C, Papadimitriou CH (2006) Computing pure Nash equilibria in graphical games via Markov random fields. In: Proceedings of the 7th ACM conference on electronic commerce. ACM, pp 91–99

Demaine ED, Hajiaghayi MT, Mahini H, Zadimoghaddam M (2007) The price of anarchy in network creation games. In: Proceedings of the twenty-sixth annual ACM symposium on principles of distributed computing. ACM, pp 292–298

Dürr C, Thang N (2007) Nash equilibria in Voronoi games on graphs. In: Algorithms-ESA, pp 17–28

Eksin C, Molavi P, Ribeiro A, Jadbabaie A (2013) Learning in network games with incomplete information: asymptotic analysis and tractable implementation of rational behavior. IEEE Signal Process Mag 30(3):30–42

Etesami SR, Başar T (2015) Game-theoretic analysis of the Hegselmann-Krause model for opinion dynamics in finite dimensions. IEEE Trans Autom Control 60(7):1886–1897

Etesami SR, Başar T (2016) Complexity of equilibrium in competitive diffusion games on social networks. Automatica 68:100–110

Etesami SR, Başar T (2017a) Price of anarchy and an approximation algorithm for the binary-preference capacitated selfish replication game. Automatica 76:153–163

Etesami SR, Başar T (2017b) Pure Nash equilibrium in a binary capacitated resource allocation game. IEEE Trans Control Netw Syst. https://doi.org/10.1109/TCNS.2016.2631440, http://ieeexplore.ieee.org/abstract/document/7752896/. Published online: 22 Nov 2016

Etesami SR, Saad W, Mandayam N, Poor HV (2017) Smart routing in smart grids. In: Proceedings of the 56th IEEE Conference on Decision and Control, Melbourne, 12–15 Dec. https://arxiv.org/abs/1705.03805

Fabrikant A, Luthra A, Maneva E, Papadimitriou CH, Shenker S (2003) On a network creation game. In: Proceedings of the twenty-second annual symposium on principles of distributed computing. ACM, pp 347–351

Fabrikant A, Papadimitriou C, Talwar K (2004) The complexity of pure Nash equilibria. In: Proceedings of the Thirty-sixth Annual ACM Symposium on Theory of Computing. ACM, pp 604–612

Galeotti A, Goyal S, Jackson MO, Vega-Redondo F, Yariv L (2010) Network games. Rev Econ Stud 77(1):218–244

Gionis A, Terzi E, Tsaparas P (2013) Opinion maximization in social networks. In: Proceedings of the 2013 SIAM international conference on data mining. SIAM, pp 387–395

Goemans MX, Li L, Mirrokni VS, Thottan M (2006) Market sharing games applied to content distribution in ad hoc networks. IEEE J Sel Areas Commun 24(5):1020–1033

Gopalakrishnan R, Kanoulas D, Karuturi NN, Rangan CP, Rajaraman R, Sundaram R (2012) Cache me if you can: capacitated selfish replication games. In: LATIN. Springer, pp 420–432

Goyal S, Vega-Redondo F (2000) Learning, network formation and coordination. Technical report, Tinbergen Institute Discussion Paper

Goyal S, Heidari H, Kearns M (2014) Competitive contagion in networks. Games Econ Behav. https://doi.org/10.1016/j.geb.2014.09.002, http://www.sciencedirect.com/science/article/pii/S0899825614001365?via%3Dihub

Grossklags J, Christin N, Chuang J (2008) Secure or insure? A game-theoretic analysis of information security games. In: Proceedings of the 17th International Conference on World Wide Web. ACM, pp 209–218

Jackson MO (2005) A survey of network formation models: stability and efficiency. In: Group formation in economics: networks, clubs, and coalitions. Cambridge University Press, Cambridge, pp 11–49

Jackson MO (2010) Social and economic networks. Princeton University Press, Princeton

Jackson MO, Watts A (2002) On the formation of interaction networks in social coordination games. Games Econ Behav, 41(2):265–291

Jackson MO, Wolinsky A (1996) A strategic model of social and economic networks. J Econ Theory 71(1):44–74

Jackson MO, Yariv L (2007) Diffusion of behavior and equilibrium properties in network games. Am Econ Rev 97(2):92–98

Jensen MK (2010) Aggregative games and best-reply potentials. Econ Theory 43(1):45–66

Kawald B, Lenzner P (2013) On dynamics in selfish network creation. In: Proceedings of the twenty-fifth annual ACM symposium on parallelism in algorithms and architectures. ACM, pp 83–92

Kearns M, Littman ML, Singh S (2001) Graphical models for game theory. In: Proceedings of the seventeenth conference on uncertainty in artificial intelligence. Morgan Kaufmann Publishers Inc., pp 253–260

Krichene W, Reilly JD, Amin S, Bayen AM (2017) Stackelberg routing on parallel transportation networks. In: Başar T, Zaccour G (eds) Handbook of Dynamic Game Theory. Springer International Publishing, Cham

Koshal J, Nedić A, Shanbhag UV (2016) Distributed algorithms for aggregative games on graphs. Oper Res 64(3):680–704

Koutsoupias E, Papadimitriou C (1999) Worst-case equilibria. In: STACS

Laoutaris N, Telelis O, Zissimopoulos V, Stavrakakis I (2006) Distributed selfish replication. IEEE Trans Parallel Distrib Syst 17(12):1401–1413

Lenzner P (2011) On dynamics in basic network creation games. In: Algorithmic game theory. Springer, Berlin/London, pp 254–265
Li N, Marden JR (2013) Designing games for distributed optimization. IEEE J Sel Top Sign Proces 7(2):230–242
Liang X, Xiao Y (2012) Game theory for network security. IEEE Commun Surv Tutor 15(1):472–486
Maheswaran RT, Başar T (2001) Decentralized network resource allocation as a repeated noncooperative market game. In: Proceedings of the 40th IEEE conference on decision and control, vol 5. IEEE, pp 4565–4570
Marden JR, Roughgarden T (2014) Generalized efficiency bounds in distributed resource allocation. IEEE Trans Autom Control 59(3):571–584
Marden JR, Arslan G, Shamma JS (2007) Regret based dynamics: convergence in weakly acyclic games. In: Proceedings of the 6th International Joint Conference on Autonomous Agents and Multiagent Systems. ACM, p 42
Menache I, Ozdaglar A (2011) Network games: theory, models, and dynamics. Synth Lect Commun Netw 4(1):1–159
Milchtaich I (1996) Congestion games with player-specific payoff functions. Games Econ Behav 13(1):111–124
Monderer D, Shapley LS (1996) Potential games. Games Econ Behav 14(1):124–143
Montanari A, Saberi A (2010) The spread of innovations in social networks. Proc Natl Acad Sci 107(47):20196–20201
Murray C (2012) Markets in political influence: rent-seeking, networks and groups. Technical report, University Library of Munich
Nisan N, Roughgarden T, Tardos E, Vazirani VV (2007) Algorithmic game theory, vol 1. Cambridge University Press, Cambridge/New York
Pacifici V, Dan G (2012) Convergence in player-specific graphical resource allocation games. IEEE J Sel Areas Commun 30(11):2190–2199
Panagopoulou P, Spirakis P (2008) A game theoretic approach for efficient graph coloring. In: Algorithms and computation, pp 183–195
Parise F, Gentile B, Grammatico S, Lygeros J (2015) Network aggregative games: distributed convergence to Nash equilibria. In: 54th Annual Conference on Decision and Control (CDC). IEEE, pp 2295–2300
Pollatos GG, Telelis OA, Zissimopoulos V (2008) On the social cost of distributed selfish content replication. In: International conference on research in networking. Springer, pp 195–206
Rosenthal RW (1973) A class of games possessing pure-strategy Nash equilibria. Int J Game Theory 2(1):65–67
Roughgarden T (2002) Selfish routing. Technical report, Ph.D. Dissertation, Cornell University, Ithaca
Roughgarden T, Schrijvers O (2016) Network cost-sharing without anonymity. ACM Trans Econ Comput 4(2):8
Saad W, Han Z, Başar T, Debbah M, Hjorungnes A (2011) Network formation games among relay stations in next generation wireless networks. IEEE Trans Commun 59(9):2528–2542
Saad W, Zhou X, Maham B, Başar T, Poor HV (2012) Tree formation with physical layer security considerations in wireless multi-hop networks. IEEE Trans Wirel Commun 11(11):3980–3991
Shakkottai S, Srikant R (2017) Communication networks: pricing, congestion control, routing and scheduling. In: Başar T, Zaccour G (eds) Handbook of Dynamic Game Theory. Springer International Publishing, Cham
Singer Y (2012) How to win friends and influence people, truthfully: influence maximization mechanisms for social networks. In: Proceedings of the fifth ACM international conference on web search and data mining. ACM, pp 733–742
Small L, Mason O (2013) Information diffusion on the iterated local transitivity model of online social networks. Discret Appl Math 161(10):1338–1344
Tekin C, Liu M, Southwell R, Huang J, Ahmad SHA (2012) Atomic congestion games on graphs and their applications in networking. IEEE/ACM Trans Netw (TON) 20(5):1541–1552

Tzoumas V, Amanatidis C, Markakis E (2012) A game-theoretic analysis of a competitive diffusion process over social networks. In: Proceedings of the 8th International Conference on Internet and Network Economics. Springer, pp 1–14

Vetta A (2002) Nash equilibria in competitive societies, with applications to facility location, traffic routing and auctions. In: Proceedings of the 43rd Annual IEEE Symposium on Foundations of Computer Science, 2002. IEEE, pp 416–425

Young HP (2006) The diffusion of innovations in social networks. In: The economy as an evolving complex system III: current perspectives and future directions, vol 267. Oxford University Press, Oxford/New York

Cooperative Differential Games with Transferable Payoffs

13

Leon A. Petrosyan and Georges Zaccour

Contents

Abstract

In many instances, players find it individually and collectively rational to sign a long-term cooperative agreement. A major concern in such a setting is how to ensure that each player will abide by her commitment as time goes by. This will occur if each player still finds it individually rational at any intermediate instant of time to continue to implement her cooperative control rather than switch to

This work was supported by the Saint Petersburg State University under research grant No. 9.38.245.2014, and SSHRC, Canada, grant 435-2013-0532.

L. A. Petrosyan (✉)
Faculty of Applied Mathematics-Control Processes, St Petersburg State University, St Petersburg, Russia
e-mail: spbuoasis7@petrlink.ru

G. Zaccour
Department of Decision Sciences, GERAD, HEC Montréal, Montreal, QC, Canada
e-mail: georges.zaccour@gerad.ca

T. Başar, G. Zaccour (eds.), *Handbook of Dynamic Game Theory*,
https://doi.org/10.1007/978-3-319-44374-4_12

a noncooperative control. If this condition is satisfied for all players, then we say that the agreement is time consistent. This chapter deals with the design of schemes that guarantee time consistency in deterministic differential games with transferable payoffs.

Keywords
Cooperative differential games · Time consistency · Strong time consistency · Imputation distribution procedure · Shapley value · Core

1 Introduction

Solving a static cooperative game is typically done using a two-step procedure. First, one determines the collectively optimal solution, and next, one distributes the total payoff among the players, using one of the many available cooperative game solutions, e.g., core, Shapley value, nucleolus. A common denominator to all these solutions is the axiom of *individual rationality*, which states that no player will accept an allocation that leaves her with less than what she could secure by not participating in the agreement.[1] In a dynamic cooperative game, one must ensure that all parties will abide by the agreement as time goes by. This will occur if each player's cooperative payoff-to-go dominates the noncooperative payoff-to-go at any intermediate instant of time. This property is known as *dynamic individual rationality* (DIR) or *time consistency*. The sustainability of cooperation in differential games is the focus of this chapter.

It frequently happens that players (e.g., partners in a supply chain; union and management; political parties; countries; spouses; etc.) agree to cooperate over a long-term horizon, say $[t_0, T]$. By cooperation, we mean that the parties coordinate their strategies in view of optimizing a collective performance index (profit, cost, welfare, happiness, etc.). Although coordination may imply some loss of freedom for the parties in terms of their choice of actions, the rationale for coordination stems, on balance, from the collective and individual gains it generates compared to noncooperation.

A relevant question is why economic and social agents bind themselves in long-term contracts instead of keeping all their options open and cooperating one period at a time? A first answer to this question is that negotiating to reach an acceptable arrangement is costly (not only in monetary terms but also in time, emotion, effort, etc.), and therefore, it makes sense to avoid frequent renegotiation whenever feasible. Second, some problems are inherently dynamic. For instance, curbing polluting emissions in the industrial and transport sectors requires making heavy investments in cleaner technologies, changing consumption habits, etc., which clearly cannot be achieved overnight. Any environmental agreement between players (countries, regions, etc.) must be long-lived to allow for such adjustments to take place.

[1] This property has also been referred to as the stand-alone test.

Despite the fact that long-term cooperation can potentially bring high collective dividends, it is an empirical fact that some agreements are abandoned before their maturity date. In a dynamic game setting, if a cooperative contract breaks down before its intended end date, we say that it is *time inconsistent*. Suppose that a game has been played cooperatively until some instant of time $\tau \in (t_0, T)$. Haurie (1976) offered two reasons why an initially agreed-upon solution may become unacceptable to one or more players at time τ:

(i) If the players agree to renegotiate the original agreement at time τ, it is not certain that they will want to continue with the agreement. In fact, they will not choose to go on with the original agreement if it is not a solution to the cooperative game that starts out at time τ.

(ii) Suppose that a player considers deviating, that is, as of time τ she will use a strategy that is different from the cooperative one. Actually, a player should do this if deviating gives her a payoff in the continuation game that is greater than the one she stands to receive through continued cooperative play.

Such instabilities arise simply because, in general, the position of the game at an intermediate instant of time will differ from the initial position. In particular, individual rationality may fail to apply when the game reaches a certain position, despite the fact that it was satisfied at the outset. This phenomenon is notable in differential games and in state-space games as such, but not in repeated games. Once the reason behind such instabilities is well understood, the research agenda becomes to attempt to find mechanisms, incentives, side payments, etc., that can help prevent breakdowns from taking place. This is our aim here.

The rest of the chapter is organized as follows: In Sect. 2 we briefly review the relevant literature, and in Sect. 3 we introduce the ingredients of deterministic differential games. Section 4 is a refresher on cooperative games. Section 5 deals with time consistency, and Sect. 6 with strong time consistency. In Sect. 7 we treat the case of random terminal time, and we briefly conclude in Sect. 8.

2 Brief Literature Background

Broadly, two approaches have been developed to sustain cooperation over time in differential games, namely, seeking a cooperative equilibrium and designing time-consistent solutions.[2]

In the cooperative equilibrium approach, the aim, as the name suggests, is to embed the cooperative solution with an equilibrium property that renders it self-

[2]This chapter focuses on how a cooperative solution in a differential game can be sustained over time. There is a large literature that looks at the dynamics of a cooperative solution, especially the core, in different game settings, but not in differential games. Here, the environment of the game changes when a coalition deviates; for instance, the set of players may vary over time. We refrain from reviewing this literature and direct the interested reader to Lehrer and Scarsini (2013) and the references therein.

enforcing (stable). Studies in repeated games have developed conditions under which individually rational outcomes can be sustained as equilibria over time. One folk theorem result states that any individually rational payoff vector can be supported as a Nash equilibrium outcome in an infinitely repeated game if players are sufficiently farsighted (Friedman 1986). This raises the question of whether individually rational outcomes exist. In repeated games with complete information and perfect monitoring, the answer is yes since the players face the same game at every stage. In state-space games, the situation is different. In a discrete-time setup, a stochastic game includes a state variable that evolves over time, as a product of the initial conditions, the players' actions, and a transition law. The latter may be deterministic, in which case the game is a multistage game (or a difference game). A folk theorem for stochastic games is given in Dutta (1995), but no general theorems seem to exist for differential games. The reason is that the presence of a state variable complicates folk theorem analysis: a deviation by a player yields a one-shot gain (as in a repeated game), but it also changes the state, now and in all future periods. However, particular results exist for situations in which Pareto-optimal outcomes are supported by trigger strategies in differential games (Dockner et al. 2000; Haurie et al. 2012).[3] Such strategies embody (effective) punishments that deprive any player the benefits of a defection, and the threats of punishments are credible which ensures that it is in the best interest of the player(s) who did not defect to implement a punishment. Early examples of contributions in this area include Haurie and Tolwinski (1985), Tolwinski et al. (1986), Haurie and Pohjola (1987), and Haurie et al. (1994). The books by Dockner et al. (2000) and Haurie et al. (2012) provide a comprehensive introduction to cooperative equilibria in differential games.

Note that the folk theorems are for infinite-horizon games. Enforcing cooperation in finite-horizon games is more difficult, not to say generally elusive. The reason is that defection from the agreement at the last stage is individually rational and this deviation cannot be punished. Using a backward-induction argument, it is easy to show that the unique subgame-perfect equilibrium in repeated and multistage games is to implement Nash equilibrium controls at each stage of the finite game. This clear-cut theoretical result has not always received empirical support, and in fact, experiments show that cooperation may be realized, at least partially, in finite-horizon games (see, e.g., Angelova et al. 2013). The literature has came out with different ways to cope with the difficulties in enforcing cooperation in finite-horizon dynamic games. See, e.g., Radner (1980), Benoit and Krishna (1985), Eswaran and Lewis (1986), Mailath et al. (2005), Flesch et al. (2014), Flesch and Predtetchinski (2015), and Parilina and Zaccour (2015a).

[3]There are (rare) cases in which a cooperative outcome "by construction" is in equilibrium.This occurs if a game has a Nash equilibrium that is also an efficient outcome. However, very few differential games have this property. The fishery game of Chiarella et al. (1984) is an example. Martín-Herrán and Rincón-Zapatero (2005) and Rincón-Zapatero et al. (2000) state conditions for Markov-perfect equilibria to be Pareto optimal in a special class of differential games.

Having in mind the same objective of embedding the cooperative solution with an equilibrium property, a series of papers considered incentives within a Stackelberg dynamic game context (deterministic and stochastic). In this setting, the incentive problem is one-sided, that is, the follower is incentivized to align its objective with that of the leader (hence enforcement of cooperation). Early contributions include Başar and Olsder (1980), Zheng and Başar (1982), Başar (1984, 1985), and Cansever and Başar (1985). Ehtamo and Hämäläinen (1986, 1989, 1993) considered two-sided incentive strategies in two-player differential games. A player's incentive strategy is a function of the other player's action. In an incentive equilibrium, each player implements her part of the agreement if the other player also does. In terms of computation, the determination of an incentive equilibrium requires solving a pair of optimal control problems, which is in general relatively easy to do. A main concern with incentive strategies is their credibility, since it may happen that the best response to a deviation from cooperation is to stick to cooperation rather than also deviating. In such a situation, the threat of punishment for a deviation is an empty one. In applications, one can derive the conditions that the parameter values must satisfy to have credible incentive strategies. For a discussion of the credibility of incentive strategies in differential games with special structures, see Martín-Herrán and Zaccour (2005, 2009). In the absence of a hierarchy in the game, a further drawback of incentive equilibria is that the concept is defined for only two players. An early reference for incentive design problems with one leader and n followers, as well as design problems with multiple levels of hierarchy is Başar (1983). See also Başar (1989) for stochastic incentive design multiple levels problems. Incentive strategies and equilibria have been applied in a number of areas, including environmental economics (see, e.g., Breton et al. 2008 and De Frutos et al. 2015), marketing (see, e.g., Martín-Herrán et al. 2005, Buratto and Zaccour 2009, and Jørgensen and Zaccour 2002b), and closed-loop supply chains (De Giovanni et al. 2016).

The second line of research, which was originally active in differential games before moving on to other classes of games, is to request that the cooperative agreement be time consistent, i.e., it satisfies the dynamic individual rationality property. The issue here is: will a cooperative outcome that is individually rational at the start of the game continue to be so as the game proceeds? The starting point is that players negotiate and agree on a cooperative solution and on the actions it prescribes for the players. Clearly, the agreement must satisfy individual rationality at the initial position of the game, since otherwise, there would be no scope for cooperation at all.

In a nutshell, we say that a cooperative solution is time consistent at the initial position of the game if, at any intermediate position, the cooperative payoff-to-go of each player dominates her noncooperative payoff-to-go. It is important to note here that the comparison between payoffs-to-go is carried out along the cooperative state trajectory, which means that the game has evolved cooperatively till the time of comparison. Kaitala and Pohjola (1990, 1995) proposed the concept of agreeability, which requires cooperative payoff-to-go dominance along any state trajectory, that is, not only the cooperative state trajectory. From the definitions of time consistency and agreeability, it is clear that the latter implies the former. In the class of *linear-state differential games*, Jørgensen et al. (2003) show that if the cooperative solution is time consistent, then it is also agreeable. Further, Jørgensen et al. (2005) show that there is also equivalence between time consistency and agreeability in the class of *homogenous linear-quadratic differential games* (HLQDG).[4]

There exists a large applied and theoretical literature on time consistency in cooperative differential games. The concept itself was initially proposed in Petrosyan (1977); see also Petrosjan and Danilov (1979, 1982, 1986). In these publications in Russian, as well as in subsequent books in English (Petrosyan 1993; Petrosjan and Zenkevich 1996), and in Petrosyan (1977), time consistency was termed dynamic stability. For an implementation of a time-consistent solution in different deterministic differential game applications, see, e.g., Gao et al. (1989), Haurie and Zaccour (1986, 1991), Jørgensen and Zaccour (2001, 2002a), Petrosjan and Zaccour (2003), Petrosjan and Mamkina (2006), Yeung and Petrosjan (2001, 2006), Kozlovskaya et al. (2010), and Petrosyan and Gromova (2014). Interestingly, whereas all these authors are proposing a time-consistent solution in a differential game, the terminology may vary between papers. For a tutorial on time consistency in differential games, see Zaccour (2008). Finally, time consistency in differential games with a random terminal duration is the subject of Petrosyan and Shevkoplyas (2000, 2003) and Marín-Solano and Shevkoplyas (2011).

3 A Differential Game

The description of a deterministic differential game played on a time interval $[t_0, T]$ involves the following elements:

1. A set of players $N = \{1, \ldots, n\}$.
2. For each player $i \in N$, a control variable $u_i \in U_i$, where U_i is a compact set of admissible control values for player i. Let $u(t) = (u_1(t), \ldots, u_n(t))$.
3. A state variable $x \in \mathbb{R}^n$.[5] The evolution of the state is governed by the following differential equation:

$$\dot{x} = f(x, u_1, \ldots, u_n), \quad x(t_0) = x_0, \tag{13.1}$$

where $x_0 \in \mathbb{R}^n$ is the initial given state value.

[4]Such games have the following two characteristics: (i) The instantaneous-payoff function and the salvage-value function are quadratic with no linear terms in the state and control variables; (ii) the state dynamics are linear in the state and control variables.

[5]Typically, one considers $x(t) \in X \subseteq \mathbb{R}^n$, where X is the set of admissible states. To avoid unecessary complications for what we are attempting to achieve here, we assume that the state space is $\mathbb{R}^n$.

4. A payoff functional for player $i, i \in N$,

$$K_i(x_0, T - t_0; u_1, \ldots, u_n) = \int_{t_0}^{T} h_i(x(t), u_1(t), \ldots, u_n(t))dt + H_i(x(T)), \tag{13.2}$$

   where $h_i(\cdot)$ is player i's instantaneous payoff and function $H_i(\cdot)$ is her terminal payoff (or reward or salvage value).
5. An information structure, that is, the piece of information that players consider when making their decisions. Here, we retain a feedback information structure, which means that the players base their decisions on the position of the game $(t, x(t))$.

We make the following assumptions:

Assumption 1. *For each feasible players' strategy, there exists a unique, and extensible on* $[t_0, \infty)$, *solution of the system* (13.1).

Assumption 2. *Functions h_i and f are continuously differentiable in x and u. The function H_i, $i \in N$, is continuously differentiable in x.*

Assumption 3. *Functions h_i and H_i are positive for all values of u and x.*

Assumptions 1 and 2 are made to avoid dealing with the case where the feedback strategies are nonsmooth functions of x, which implies that we may lose the uniqueness property of the trajectory generated by such strategies, according to the state equations. Although this case is clearly of interest, we wish to focus here on the issues directly related to sustainability of cooperation rather than being forced to deal with technicalities that would deviate the reader's attention from the main messages. Assumption 3 is made to simplify the presentation of some of the results. It is, by no means, a restrictive assumption, as it is always possible to add a large positive number to each player's objective without altering the results.

We shall refer to the differential game described above by $\Gamma(x_0, T - t_0)$, with $T - t_0$ being the duration of the game. We suppose that there is no inherent obstacle to cooperation between the players and that their payoffs are transferable. More specifically, we assume that before the game actually starts, the players agree to play the control vector $u^* = \left(u_1^*, \ldots, u_n^*\right)$,which is given by

$$u^* = \arg\max_u \sum_{i=1}^{n} K_i(x_0, T - t_0; u_1, \ldots, u_n).$$

We shall refer to u^* as the cooperative control vector and to the resulting x^* as the cooperative state and to $x^*(t)$ as the cooperative state trajectory. The corresponding total cooperative payoff is given by

$$TCP = \sum_{i=1}^{n} \left(\int_{t_0}^{T} h_i(x^*(t), u_1^*(t), \ldots, u_n^*(t))dt + H_i(x^*(T)) \right).$$

A pending issue is how to divide TCP among the players. As mentioned in the introduction, this can be done by using one of the available solutions of a cooperative game. We recall the basics of cooperative games below. Further, we note that the amount $K_i(x_0, T - t_0; u_1^*, \ldots, u_n^*)$ is player i's payoff before any side payment is made, which does not correspond in general to the outcome that this player will indeed pocket in the game.

For a comprehensive coverage of differential games, the interested reader may consult one of the available textbooks on the subject, e.g., Başar and Olsder (1995), Dockner et al. (2000), Engwerda (2005), and Haurie et al. (2012).

4 A Refresher on Cooperative Games

We recall in this section some basic elements of cooperative game theory. These are presented while keeping in mind what is needed in the sequel.

A dynamic cooperative game of duration $T - t_0$ is a triplet (N, v, L_v), where N is the set of players; v is the characteristic function that assigns to each coalition $S, S \subseteq N$, a numerical value,

$$v(S; x_0, T - t_0) : P(N) \to \mathbb{R}, \qquad v(\varnothing; x_0, T - t_0) = 0,$$

where $P(N)$ is the power set of N and $L_v(x_0, T - t_0)$ is the set of imputations, that is:

$$L_v(x_0, T - t_0) = \Big\{ (\xi_1(x_0, T - t_0), \ldots, \xi_m(x_0, T - t_0))$$
$$\text{such that } \xi_i(x_0, T - t_0) \geq v(\{i\}; x_0, T - t_0)$$
$$\text{and} \quad \sum_{i \in N} \xi_i(x_0, T - t_0) = v(N; x_0, T - t_0) \Big\}.$$

The definition of the set of imputations involves two conditions, namely,

$$\text{individual rationality} : (\xi_i(x_0, T - t_0) \geq v(\{i\}; x_0, T - t_0)),$$

and

$$\text{collective rationality}: \left(\sum_{i \in N} \xi_i \left(x_0, T - t_0\right) = v \left(N; x_0, T - t_0\right)\right).$$

Individual rationality means that no player will accept an allocation or imputation that gives her less than what she could secure by acting alone. Collective rationality means that the total collective gain should be allocated, that is, no deficits or subsidies are considered. To make the connection with what we wrote earlier, observe that

$$v \left(N; x_0, T - t_0\right) = \sum_{i=1}^{n} K_i (x_0, T - t_0; u_1^*, \ldots, u_n^*),$$

and that player i will get some $\xi_i \left(x_0, T - t_0\right)$, which will not necessarily be equal to $K_i (x_0, T - t_0; u_1^*, \ldots, u_n^*)$.

The characteristic function (CF) measures the power or the strength of a coalition. Its precise definition depends on the assumption made about what the left-out players—that is, the complement subset of players $N \backslash S$—will do (see, e.g., Ordeshook 1986 and Osborne and Rubinstein 1994). A number of approaches have been proposed in the literature to compute $v\left(\cdot\right)$. We briefly recall some of these approaches.

α-CF: In their seminal book, Von Neumann and Morgenstern (1944) defined the characteristic function as follows:

$$v^{\alpha} \left(S; x_0, T - t_0\right) := \max_{u_S = (u_i : i \in S)} \min_{u_{N \backslash S} = (u_j : j \in N \backslash S)} \sum_{i \in S} K_i \left(x_0, T - t_0; u_S, u_{N \backslash S}\right)$$

That is, $v^{\alpha}(\cdot)$ represents the maximum payoff that coalition S *can guarantee for itself* irrespective of the strategies used by the players in $N \backslash S$.

β-CF: The β characteristic function is defined as

$$v^{\beta} \left(S; x_0, T - t_0\right) := \min_{u_{N \backslash S} = (u_j : j \in N \backslash S)} \max_{u_S = (u_i : i \in S)} \sum_{i \in S} K_i \left(x_0, T - t_0; u_S, u_{N \backslash S}\right),$$

that is, $v^{\beta}(\cdot)$ gives the maximum payoff that coalition S *cannot be prevented from getting* by the players in $N \backslash S$.

γ-CF: The γ characteristic function, which is attributed to Chander and Tulkens (1997), is defined as the partial equilibrium outcome of the noncooperative game between coalition S and left-out players acting individually, that is, each player not belonging to the coalition best-replies to the other players' strategies. Formally, for any coalition $S \subseteq N$, we have

$$v^{\gamma}(S; x_0, T - t_0) := \sum_{i \in S} K_i \left(x_0, T - t_0; u_S, \{u_j\}_{j \in N \setminus K} \right) \tag{13.3}$$

$$u_S := \text{argmax}_{u_S} \sum_{i \in S} K_i \left(x_0, T - t_0; u_S, \{u_j\}_{j \in N \setminus K} \right)$$

$$u_j := \text{argmax}_{u_j} K_j \left(x_0, T - t_0; u_S, u_j, \{u_l\}_{l \in N \setminus \{S \cup j\}} \right),$$

$$\text{for all } j \in N \setminus K.$$

δ-CF: The δ characteristic function, which is attributed to Petrosjan and Zaccour (2003), assumes that the left-out players do not react strategically to the formation of the coalition, but implement their Nash equilibrium actions determined in the n-player noncooperative game (or any other fixed actions). Formally, for any coalition $S \subseteq N$,

$$v^{\delta}(S; x_0, T - t_0) := \sum_{i \in S} K_i \left(x_0, T - t_0; u_S, \{\tilde{u}_j\}_{j \in N \setminus S} \right) \tag{13.4}$$

$$u_S := \text{argmax}_{u_S} \sum_{i \in S} K_i \left(x_0, T - t_0; u_S, \{\tilde{u}_j\}_{j \in N \setminus S} \right)$$

$$\tilde{u}_j := \text{argmax}_{u_j} K_j \left(x_0, T - t_0; u_j, \{u_l\}_{l \in N \setminus \{j\}} \right), \text{ for all } j \in N.$$

Remark 1. We make the following observations:

1. The α and β characteristic functions assume that left-out players form an anti-coalition. The γ and δ characteristic functions do not make such assumption, but suppose that these players act individually.
2. The α and β characteristic functions are superadditive, that is,

$$v(S \cup Q, x_0, T - t_0) \geq v(S, x_0, T - t_0) + v(Q, x_0, T - t_0),$$

$$\text{for al } S, Q \subset N, S \cap Q = \varnothing.$$

   The γ and δ characteristic functions were introduced in the context of games with negative externalities and have been shown to be superadditive in this context (see Reddy and Zaccour 2016).
3. The γ and δ characteristic functions coincide for linear-state differential games (see Zaccour 2003).
4. In the absence of externalities, i.e., if the payoffs to the members of coalition S are independent of the actions of the nonmembers ($N \setminus S$), then $v(S; x_0, T - t_0)$ would be the result of an optimization problem and not an equilibrium one.

5. The positiveness of payoff functions $K_i, i = 1, \ldots, n$ implies positiveness of the characteristic function. If $v(\cdot)$ is superadditive, then

$$v(S'; x_0, T - t_0) \geq v(S; x_0, T - t_0),$$

for any $S, S' \subset N$ such that $S \subset S'$, i.e., the superadditivity of the function v in S implies that this function is monotone in S.

4.1 Solution Concepts

The set of imputations $L_v(x_0, T - t_0)$ can be interpreted as the set of admissible solutions (allocations). Once it is defined, the next step is to choose a particular imputation from that set or to select a subset of $L_v(x_0, T - t_0)$. To do so, game theorists have proposed different solution concepts, which are typically defined by a series of axioms or requirements that the allocation(s) must satisfy, e.g., fairness and stability. We distinguish between solution concepts that select a unique imputation in $L_v(x_0, T - t_0)$, e.g., Shapley value and the nucleolus, and those that select a subset of imputations, e.g., the core and stable set. The two most used solution concepts in applications of cooperative games are the Shapley value and the core. We recall their definitions and use them later on. Denote by s the number of players in coalition S.

Definition 1. The Shapley value is an imputation $Sh = (Sh_1, \ldots Sh_n)$ defined by

$$\begin{aligned} &Sh_i(x_0, T - t_0) \\ &= \sum_{\substack{S \subset N \\ i \in S}} \frac{(n-s)!(s-1)!}{n!} \left(v(S; x_0, T - t_0) - v(S \setminus \{i\}; x_0, T - t_0)\right). \end{aligned} \tag{13.5}$$

Being an imputation, the Shapley value satisfies individual rationality, i.e., $Sh_i(x_0, T - t_0) \geq v(\{i\}; x_0, T - t_0))$ for all $i \in N$. The term $(v(S; x_0, T - t_0) - v(S \setminus \{i\}; x_0, T - t_0))$ measures the marginal contribution of player i to coalition S. Thus, the Shapley value allocates to each player the weighted sum of her marginal contributions to all coalitions that she may join. The Shapley value is the unique imputation satisfying three axioms: fairness (identical players are treated in the same way), efficiency $\left(\sum_{i \in N} Sh_i(x_0, T - t_0) = v(N; x_0, T - t_0)\right)$, and linearity (if v and w are two characteristic functions defined for the same set of players, then $Sh_i(v + w) = Sh_i(v) + Sh_i(w)$ for all $i \in N$).

The core is the set of all undominated imputations. Gillies (1953) showed that for an imputation $\xi(x_0, T - t_0) = (\xi_1(x_0, T - t_0), \ldots, \xi_n(x_0, T - t_0))$ to be in the core, it must satisfy the condition

$$\sum_{i \in S} \xi_i \left(x_0, T - t_0\right) \geq v(S; x_0, T - t_0), \forall S \subseteq N.$$

In other words, the above condition states that an imputation is in the core if it allocates to each possible coalition an outcome that is at least equal to what this coalition can secure by acting alone. Consequently, the core is defined by

$$C = \left\{ \xi \left(x_0, T - t_0\right), \text{ such that } \sum_{i \in S} \xi_i \left(x_0, T - t_0\right) \geq v(S; x_0, T - t_0), \forall S \subset N, \right.$$
$$\left. \text{and } \sum_{i \in N} \xi_i \left(x_0, T - t_0\right) = v(N; x_0, T - t_0) \right\}.$$

Note that the core may be empty, may be a singleton, or may contain many imputations.

Remark 2. Note that the Shapley value and the core were introduced for the whole game, that is, the game with initial state x_0 and duration $T - t_0$. Clearly, we can define them for any subgame $\Gamma(x_t, T - t)$.

5 Time Consistency of the Cooperative Solution

5.1 Preliminaries

To implement a time-consistent solution, we need to keep track of what happens at any position $\left(x_t^*, T - t\right)$ of the game. From now on, we shall use $\Gamma_v(x_0, T - t_0)$ to refer to the overall cooperative game and $\Gamma_v(x_t, T - t)$ for the cooperative subgame starting out in state x_t and of duration $T - t$. The subscript v refers to the characteristic function, which can be any of those defined above. We denote by $W_v(x_0, T - t_0) \subset \Gamma_v(x_0, T - t_0)$, the subset of imputations selected according to the chosen solution, e.g., core and Shapley value. Similarly, $W_v(x_t, T - t)$ will refer to a subset of $\Gamma_v(x_t, T - t)$. Let $x^*(t)$, $t \in [t_0, T]$ be the trajectory resulting from cooperation. As this trajectory is generated by joint optimization, we shall refer to it as the optimal trajectory. (For simplicity, we henceforth assume that such a trajectory exists.)

We assume that $W_v(x_0, T - t_0) \neq \emptyset$; otherwise, the problem of time consistency is emptied. Note that this condition is always fulfilled for some cooperative game solutions, such as the Shapley value or the nucleolus, but not necessarily for the core.

Now, we consider the behavior of the set $W_v(x_0, T - t_0)$ along the optimal state trajectory $x^*(t)$. Toward this end, in each current state $x^*(t)$, we define the characteristic function $v(S; x^*(t), T - t)$ using any one of the approaches recalled in the previous section.

The imputation set in the current cooperative game $\Gamma_v(x^*(t), T - t)$ is of the form

$$L_v(x^*(t), T - t) = \Big\{\xi \in R^n \mid \xi_i \geq v(\{i\}; x^*(t), T - t), \quad i = 1, \ldots, n;$$

$$\sum_{i \in N} \xi_i = v(N; x^*(t), T - t)\Big\}.$$

Consider the family of current games

$$\{\Gamma_v(x^*(t), T - t) = \langle N, v(S; x^*(t), T - t)\rangle, \quad t_0 \leq t \leq T\},$$

determined along the optimal state trajectory $x^*(t)$ and their solutions $W_v(x^*(t), T - t) \subset L_v(x^*(t), T - t)$ generated by the same cooperative game solution (or principle of optimality) as the one agreed upon at initial position $(x_0, T - t_0)$, that is, $W_v(x_0, T - t_0)$.

It is obvious that the set $W_v(x^*(T), 0)$ is a solution of the terminal game $\Gamma_v(x^*(T), 0)$ and contains the unique imputation $H(x^*(T)) = \{H_i(x^*(T)), i = 1, \ldots, n\}$, where $H_i(x^*(T))$ is the terminal payoff of player i along the trajectory $x^*(t)$.

5.2 Designing a Time-Consistent Solution

Assume that, at initial state x_0, the players agree upon the imputation $\xi^0 \in W_v(x_0, T - t_0)$, with $\xi^0 = \left(\xi_1^0, \ldots, \xi_n^0\right)$. Denote by $\xi_i(x^*(t))$ player i's share on the time interval $[t_0, t]$, where t is any intermediate date in the planning horizon. By writing $\xi_i(x^*(t))$, we wish to highlight that this payoff is computed along the cooperative state trajectory and, more specifically, that cooperation has prevailed from t_0 till t. If cooperation remains in force for the rest of the game, then player i will receive her remaining due $\eta_i^t = \xi_i^0 - \xi_i(x^*(t))$ during the time interval $[t, T]$. In order for the original agreement (the imputation ξ^0) to be maintained, it is necessary that the vector $\eta^t = (\eta_i^t, \ldots, \eta_n^t)$ belong to the set $W_v(x^*(t), T - t)$, i.e., that it be a solution of the current subgame $\Gamma_v(x^*(t), T - t)$. If such a condition is satisfied at each instant of time $t \in [t_0, T]$ along the trajectory $x^*(t)$, then the imputation ξ^0 is realized. This is the conceptual meaning of the imputation's time consistency (Petrosyan 1977; Petrosjan and Danilov 1982).

Along the trajectory $x^*(t)$ on the time interval $[t, T]$, $t_0 \leq t \leq T$, the grand coalition N obtains the payoff

$$v(N; x^*(t), T - t) = \sum_{i \in N} \left[\int_t^T h_i(x^*(\tau), u_1^*(\tau), \ldots, u_n^*(\tau)) d\tau + H_i(x^*(T)) \right].$$

Then, the difference

$$v(N; x_0, T - t_0) - v(N; x^*(t), T - t) = \int_{t_0}^{t} \sum_{i \in N} h_i(x^*(\tau), u_1^*(\tau), \dots, u_n^*(\tau)) d\tau,$$

is the payoff that the grand coalition N realizes on the time interval $[t_0, t]$. Under our assumption of transferable payoffs, the share of the ith player in the above payoff can be represented as

$$\gamma_i(t) = \int_{t_0}^{t} \beta_i(\tau) \sum_{i=1}^{n} h_i(x^*(\tau), u_1^*(\tau), \dots, u_n^*(\tau)) d\tau = \gamma_i(x^*(t), \beta), \tag{13.6}$$

where $\beta_i(\tau), i \in N$, is the $[t_0, T]$ integrable function satisfying the condition

$$\sum_{i=1}^{n} \beta_i(t) = 1, \quad t_0 \le t \le T. \tag{13.7}$$

Differentiating (13.6) with respect to time, we get

$$\dot{\gamma}_i(t) = \frac{d\gamma_i}{dt}(t) = \beta_i(t) \sum_{i \in N} h_i(x^*(t), u_1^*(\tau), \dots, u_n^*(\tau)).$$

This quantity can be interpreted as the instantaneous gain of player i at time t. Hence, the vector $\beta(t) = (\beta_1(t), \dots, \beta_n(t))$ prescribes a distribution of the total gain among the members of N. The point here is that by properly choosing $\beta(t)$, the players could implement the desirable outcome, i.e., the agreed-upon imputation ξ^0. This is achieved by distributing the players' payoffs over time, so that, at each instant $t \in [t_0, T]$, no player wishes to deviate from cooperation (or no objection is raised) against the original agreement (the imputation ξ^0).

Definition 2. The imputation $\xi^0 \in W_v(x_0, T - t_0)$ is called time consistent in the game $\Gamma_v(x_0, T - t_0)$ if the following conditions are satisfied:

(1) There exists an optimal trajectory $x^*(t)$ along which

$$W_v(x^*(t), T - t) \neq \emptyset, t_0 \le t \le T.$$

(2) There exists an integrable vector function $\beta(t) = (\beta_1(t), \dots, \beta_n(t))$ on $[t_0, T]$ such that, for each $t_0 \le t \le T$, $\sum_{i=1}^{n} \beta_i(t) = 1$ and

$$\xi^0 \in \bigcap_{t_0 \le t \le T} [\gamma(x^*(t), \beta) \oplus W_v(x^*(t), T - t)], \tag{13.8}$$

where

$$\gamma(x^*(t),\beta) = (\gamma_1(x^*(t),\beta),\ldots,\gamma_n(x^*(t),\beta)),$$

it holds that $W_v(x^*(t), T-t)$ is a solution of the current game $\Gamma_v(x^*(t), T-t)$.

Remark 3. In the above definition, the set $[\gamma(x^*(t),\beta) \oplus W_v(x^*(t), T-t)]$ is defined as

$$[\gamma(x^*(t),\beta) \oplus W_v(x^*(t), T-t)] = \{\gamma(x^*(t),\beta) + a : \ a \in W_v(x^*(t), T-t)\}.$$

Definition 3. We say that the cooperative differential game $\Gamma_v(x_0, T-t_0)$ with side payments has a time-consistent solution $W_v(x_0, T-t_0)$ if all imputations $\xi \in W_v(x_0, T-t_0)$ are time consistent.

From the definition of time consistency, at instant $t = T$, we have $\xi^0 \in \gamma(x^*(T),\beta) \oplus W_v(x^*(T), 0)$, where $W_v(x^*(T), 0)$ is a solution of the current game $\Gamma_v(x^*(T), 0)$ played on the trajectory $x^*(t)$, $t \in [t_0, T]$, and has the only imputation $\xi^T = H(x^*(T)) = \{H_i(x^*(T))\}$. The imputation ξ^0 may be represented as

$$\xi^0 = \gamma(x^*(T),\beta) + H(x^*(T)),$$

or

$$\xi^0 = \int_{t_0}^{T} \beta(\tau) \sum_{i\in N} h_i(x^*(\tau), u_1^*(\tau),\ldots,u_n^*(\tau))d\tau + H(x^*(T)).$$

The time-consistent imputation $\xi^0 \in W_v(x_0, T-t_0)$ may be realized as follows: From (13.8), at any instant $t_0 \le t \le T$, we have

$$\xi^0 \in [\gamma(x^*(t),\beta) \oplus W_v(x^*(t), T-t)], \tag{13.9}$$

where

$$\gamma(x^*(t),\beta) = \int_{t_0}^{t} \beta(\tau) \sum_{i\in N} h_i(x^*(\tau), u_1^*(\tau),\ldots,u_n^*(\tau))d\tau \tag{13.10}$$

is the payoff on the time interval $[t_0, t]$, with player i's share in the gain on the same interval being

$$\gamma_i(x^*(t),\beta) = \int_{t_0}^{t} \beta_i(\tau) \sum_{i\in N} h_i(x^*(\tau, u_1^*(\tau),\ldots,u_n^*(\tau)))d\tau. \tag{13.11}$$

When the game proceeds along the optimal trajectory, on each time interval $[t_0, t]$ the players share the total gain

$$\int_{t_0}^{t} \sum_{i \in N} h_i(x^*(\tau), u_1^*(\tau), \ldots, u_n^*(\tau))d\tau,$$

and the following inclusion must be satisfied:

$$\xi^0 - \gamma(x^*(t), \beta) \in W_v(x^*(t), T - t). \tag{13.12}$$

Furthermore, (13.12) implies the existence of a vector $\xi^t \in W_v(x^*(t), T - t)$ such that $\xi^0 = \gamma(x^*(t), \beta) + \xi^t$. Following the choice of $\beta(t)$, the vector of gains to be obtained by the players in the remaining stage of the game is given by

$$\xi^t = \xi^0 - \gamma(x^*(t), \beta) = \int_{t}^{T} \beta(\tau) \sum_{i \in N} h_i(x^*(\tau), u_1^*(\tau), \ldots, u_n^*(\tau))d\tau + H(x^*(T)),$$

and belongs to the set $W_v(x^*(t), T - t)$.

Definition 4. The vector function

$$\alpha_i(\tau) = \beta_i(\tau) \sum_{i \in N} h_i(x^*(\tau), u_1^*(\tau), \ldots, u_n^*(\tau)), \quad i \in N, \tag{13.13}$$

is called an imputation distribution procedure (IDP).

It is clearly seen from the above definition that an IDP redistributes over time to player i the total payoff to which she is entitled in the whole game. The definition of IDP was introduced first in Petrosjan and Danilov (1979); see also Petrosjan and Danilov (1982). Note that for any vector $\beta(\tau)$ satisfying conditions (13.6) and (13.7), at each time instant $t_0 \leq t \leq T$, the players are guided by the imputation $\xi^t \in W_v(x^*(t), T - t)$ and by the same cooperative game solution concept throughout the game.

To gain some additional insight into the construction of a time-consistent solution, let us make the following additional assumption:

Assumption 4. *The vector $\xi^t \in W_v(x^*(t), T - t)$ is continuously differentiable in t.*

Under the above assumption, we can always ensure the time consistency of the imputation $\xi^0 \in W_v(x_0, T - t_0)$ by properly choosing the time function $\beta(t)$. To show this, let $\xi^t \in W_v(x^*(t), T - t)$ be a continuously differentiable function of t.

Construct the difference $\xi^0 - \xi^t = \gamma(t)$ to get

$$\xi^t + \gamma(t) \in W_v(x_0, T - t_0).$$

Let $\beta(t) = (\beta_1(t), \dots, \beta_n(t))$ be the $[t_0, T]$ integrable vector function satisfying conditions (13.6), (13.7). Instead of writing $\gamma(x^*(t), \beta)$, we will for simplicity write $\gamma(t)$. Rewriting (13.6) in vector form, we get

$$\int_{t_0}^{t} \beta(\tau) \sum_{i \in N} h_i(x^*(\tau), u_1^*(\tau), \dots, u_n^*(\tau)) d\tau = \gamma(t).$$

Differentiating with respect to t and rearranging terms, we get the following expression for $\beta(t)$:

$$\begin{aligned} \beta(t) &= \frac{1}{\sum_{i \in N} h_i(x^*(t), u_1^*(t), \dots, u_n^*(t))} \cdot \frac{d\gamma(t)}{dt} \\ &= -\frac{1}{\sum_{i \in N} h_i(x^*(t), u_1^*(t), \dots, u_n^*(t))} \cdot \frac{d\xi^t}{dt}, \end{aligned} \tag{13.14}$$

where the last expression follows from equality

$$\xi^0 = \gamma(t) + \xi^t.$$

We check that $\beta(t)$ satisfies the condition (13.7). Indeed,

$$\begin{aligned} \sum_{i \in N} \beta_i(t) &= -\frac{\sum_{i \in N} \frac{d\xi_i^t}{dt}}{\sum_{i \in N} h_i(x^*(t))} = -\frac{\frac{d}{dt} v(N; x^*(t), T - t)}{\sum_{i \in N} h_i(x^*(t))}, \\ &= -\frac{\frac{d}{dt}\left[\sum_{i \in N}\left(\int_t^T h_i(x^*(\tau)) d\tau + H_i(x^*(T))\right)\right]}{\sum_{i \in N} h_i(x^*(t))} = \frac{\sum_{i \in N} h_i(x^*(t))}{\sum_{i \in N} h_i(x^*(t))} = 1, \end{aligned}$$

since $\sum_{i \in N} \xi_i^t = v(N; x^*(t), T - t)$.

We see that, if the above assumption is satisfied and

$$W_v(x^*(t), T - t) \neq \emptyset, \quad t \in [t_0, T], \tag{13.15}$$

then the solution $W_v(x_0, T - t_0)$ is time consistent.

If the solution $W_v(x_0(t), T - t_0)$ contains many imputations, additional requirements must be added to eliminate some of these imputations, or even better, to select one of them. For instance, Yeung (2006) introduced the following irrational-behavior-proofness condition:

$$v(\{i\}; x_0, T - t_0) \leq \int_{t_0}^{t} h_i(x^*(\tau), u_1^*(\tau), \ldots, u_n^*(\tau))d\tau + v(\{i\}; x^*(t), T-t),\ i \in N, \tag{13.16}$$

where $x^*(\tau)$, $\tau \in [t_0, T]$ is the cooperative trajectory and $v(\{i\}; x^*(t), T - t)$ is the value of characteristic function for a one-player coalition in the subgame $\Gamma(x^*(t), T - t)$. If this condition holds, then it is always better to cooperate even if the grand coalition will break at an intermediate moment $t \in [t_0, T]$. It is easily seen that if, instead of $h_i(x^*(\tau), u_1^*(\tau), \ldots, u_n^*(\tau))$, the time-consistent IDP is used, this condition will always hold.

5.3 An Environmental Management Example

To illustrate the different steps involved in the design of time-consistent imputation distribution procedure, we consider a simple example.

Consider a 3-player differential game of pollution control. Denote by $y_i(t)$ the industrial production of player $i, i = 1, 2, 3$ and $t \in [0, \infty)$. Pollution is an inevitable by-product of production. Let us assume that there exists a monotone increasing relationship between production and pollution emissions, which we denote $e_i(t)$. Consequently, the benefit from production for player i can be expressed as a function of emissions, that is, $B_i(e_i)$.

Pollution accumulates over time. We denote by $x(t)$ the stock of pollution at time t and assume that its evolution over time is governed by the following linear differential equation:

$$\dot{x}(t) = \sum_{i \in N} e_i(t) - \delta x(t), \quad x(t_0) = 0, \tag{13.17}$$

where $\delta > 0$ is the absorption rate of pollution by nature.[6]

Denote by $D_i(x)$ the environmental damage cost of player i caused by the pollution stock. In the environmental economics literature, the typical assumptions are that the benefit function $B_i(e_i)$ is concave increasing, satisfying $B_i(0) = 0$, and the damage cost $D_i(x)$ is convex increasing. To perform some calculations explicitly, we retain the following functional forms:

[6]We have assumed the initial stock to be zero, which is not a severe simplifying assumption. Indeed, if this was not the case, then $x(0) = 0$ can be imposed and everything can be rescaled consequently.

$$B_i(e_i) = \alpha e_i - \frac{1}{2}e_i^2,$$

$$D_i(x) = \varphi_i x,$$

where α and φ_i are positive parameters. To keep the calculations as parsimonious as possible, while still being able to show how to determine a time-consistent solution, we have assumed that the players differ only in terms of their marginal damage cost (parameter φ_i).

Denoting by $r \in (0,1)$ the common discount rate and assuming welfare maximization behavior, then player i's optimization problem can be formulated as follows:

$$\max_{e_i} W_i = \int_{t_0}^{\infty} e^{-rt}\left(B_i(e_i(t) - D_i(x(t)))\right) dt,$$
$$= \int_{t_0}^{\infty} e^{-rt}\left(\alpha e_i - \frac{1}{2}e_i^2 - \varphi_i x\right) dt,$$

subject to the pollution dynamics in (13.17).

To start with, let us make the following assumptions:

Assumption 5. *The players agree on cooperating throughout the planning horizon and on adopting the Shapley value to share the cooperative payoffs.*

Assumption 6. *If cooperation breaks down at a certain time, then a feedback Nash equilibrium will be implemented for the remaining time.*

Remark 4. The differential game defined above is of the linear-state variety, which implies the following:

1. Feedback and open-loop Nash equilibrium strategies coincide;
2. γ-CF and δ-CF values coincide (see Zaccour 2003).

5.3.1 Cooperative Outcomes

To determine the value of the grand coalition, we solve the following joint optimization problem:

$$\max \sum_{i\in N} W_i = \max \int_{t_0}^{\infty} \sum_{i\in N} e^{-rt}\left(\alpha e_i - \frac{1}{2}e_i^2 - \varphi_i x\right) dt.$$

It is easy to verify that the optimal emissions control of player i is given by

$$e_i^* = \frac{\alpha(r+\delta) - (\varphi_1+\varphi_2+\varphi_3)}{r+\delta}, \quad i = 1, 2, 3. \tag{13.18}$$

The fact that the emissions are constant over time is a by-product of the game's linear-state structure. Further, we see that each player internalizes the damage costs of all coalition members, when selecting the optimal emissions level. To obtain the cooperative state trajectory, it suffices to insert e_i^* in the dynamics and to solve the differential equation to get

$$x^*(t) = 3\left(\frac{\alpha(r+\delta) - (\varphi_1 + \varphi_2 + \varphi_3)}{\delta(r+\delta)}\right)\left(1 - e^{-\delta t}\right). \tag{13.19}$$

Substituting for e_i^* and $x^*(t)$ in the objective function, we get the following grand coalition outcome:

$$v^\delta(N; x(0)) = \frac{3}{2r(r+\delta)^2}\left(\alpha(r+\delta) - (\varphi_1 + \varphi_2 + \varphi_3)\right)^2,$$

where the superscript δ is to highlight that here we are using the δ-CF. This is the total gain to be shared between the players if they play cooperatively during the whole (infinite-horizon) game. To simplify the notation, we write $v^\delta(N; x(0))$ instead of $v^\delta(N; x(0), T - t_0)$ as in this infinite-horizon game, any subgame is also of infinite horizon, and the only thing that matters is the value of state at the start of a subgame.

Next, we determine the emissions, pollution stock, and payoffs for all other possible coalitions. Consider a coalition K, with left-out players being $N \backslash K$. For $K = \{i\}$, which is a singleton, its value is given by its Nash outcome in the *three*-player noncooperative game. It can easily be checked that Nash equilibrium emissions are constant (i.e., independent of the pollution stock) and are given by

$$e_i^{nc} = \frac{\alpha(r+\delta) - \varphi_i}{r+\delta}, \quad i = 1, 2, 3,$$

where the superscript nc stands for noncooperation. Here, each player takes only her damage cost when deciding upon emissions. Solving the state equation, we get the following trajectory

$$x^{nc}(t) = \left(\frac{3\alpha(r+\delta) - (\varphi_1 + \varphi_2 + \varphi_3)}{\delta(r+\delta)}\right)\left(1 - e^{-\delta t}\right).$$

Substituting in the payoff functions of the players, we get their equilibrium payoff, which corresponds to $v^\delta(\{i\})$, that is,

$$v^\delta(\{i\}; x(0)) = \frac{1}{2r(r+\delta)^2}\left((\alpha(r+\delta) - \varphi_i)^2 - 2\varphi_i\left(2\alpha(r+\delta) - \left(\varphi_j + \varphi_k\right)\right)\right).$$

We still need to determine the two-player coalition values. According to the γ characteristic function, the value of a two-player coalition is given by its outcome in a noncooperative game between the coalition and the remaining player. As the γ and δ characteristic function values coincide for the linear-state game, then the value of a coalition K is given by

$$v^{\delta}(K; x(0)) = \max \int_{t_0}^{\infty} \sum_{i \in K} e^{-rt} \left(\alpha e_i - \frac{1}{2} e_i^2 - \varphi_i x \right) dt,$$

$$\dot{x} = \sum_{i \in K} e_i + e_j^{nc} - \delta x, \quad x(t_0) = 0, \quad j \in N \backslash K.$$

Solving the above optimization problem yields the following emissions levels:

$$e_i^K = \frac{\alpha(r+\delta) - \sum_{l \in K} \varphi_l}{r+\delta}, \quad i \in K,$$

$$e_j^K = \frac{\alpha(r+\delta) - \varphi_j}{r+\delta}, \quad j \in N \backslash K.$$

Inserting in the state dynamics and solving, we get

$$x^K(t) = \left(\frac{3\alpha(r+\delta) - 2\left(\sum_{l \in K} \varphi_l\right) - \varphi_j}{\delta(r+\delta)} \right) \left(1 - e^{-\delta t}\right).$$

The coalition's payoff is given by

$$v^{\delta}(K; x(0)) = \frac{1}{r(r+\delta)^2} \left(\left(\alpha(r+\delta) - \sum_{l \in K} \varphi_l \right)^2 \right.$$
$$\left. - \left(\left(\alpha(r+\delta) - \varphi_j \right) \sum_{l \in K} \varphi_l \right) \right), \quad j \in N \backslash K.$$

Once we have computed the characteristic function values for all possible coalitions, we can define the set of imputations

$$L_{\delta}(x_0, T - t_0) = \Big\{ (\xi_1(x_0, T - t_0), \ldots, \xi_m(x_0, T - t_0))$$

$$\text{such that } \xi_i(x_0, T - t_0) \geq v^{\delta}(\{i\}; x_0, T - t_0)$$

$$\text{and} \quad \sum_{i \in N} \xi_i(x_0, T - t_0) = v^{\delta}(N; x_0, T - t_0) \Big\},$$

If the players adopt the Shapley value to allocate the total cooperative game, then

$$Sh\left(x_0, T-t_0\right) = \left(Sh_1\left(x_0, T-t_0\right), \ldots, Sh_n\left(x_0, T-t_0\right)\right)$$

is the unique imputation in $W_v(x_0, T-t_0)$, where

$$\begin{aligned} Sh_1\left(x_0, T-t_0\right) &= \frac{1}{12r\left(r+\delta\right)^2}\Big\{6\alpha^2\left(r+\delta\right)^2 - 36\alpha\varphi_1\left(r+\delta\right) \\ &\quad + 12\varphi_1^2 + 3\varphi_2^2 + 3\varphi_3^2 + 4\left(4\varphi_1\left(\varphi_2+\varphi_3\right) + \varphi_2\varphi_3\right)\Big\}, \\ Sh_2\left(x_0, T-t_0\right) &= \frac{1}{12r\left(r+\delta\right)^2}\Big\{6\alpha^2\left(r+\delta\right)^2 - 36\alpha\varphi_2\left(r+\delta\right) \\ &\quad + 12\varphi_2^2 + 3\varphi_3^2 + 3\varphi_1^2 + 4\left(4\varphi_2\left(\varphi_3+\varphi_1\right) + \varphi_3\varphi_1\right)\Big\}, \\ Sh_3\left(x_0, T-t_0\right) &= \frac{1}{12r\left(r+\delta\right)^2}\Big\{6\alpha^2\left(r+\delta\right)^2 - 36\alpha\varphi_3\left(r+\delta\right) \\ &\quad + 12\varphi_3^2 + 3\varphi_1^2 + 3\varphi_2^2 + 4\left(4\varphi_3\left(\varphi_1+\varphi_2\right) + \varphi_1\varphi_2\right)\Big\}, \end{aligned}$$

where $Sh_i\left(x_0, T-t_0\right) = \xi_i\left(x_0, T-t_0\right)$.

The cooperative game considered here is of the negative-externalities variety, to use the terminology of Chander and Tulkens (1997), and therefore, the core is nonempty. Further, it can easily be shown that this game is convex[7] and, consequently, that the Shapley value is in the core. In our case, the core is given by

$$C\left(x_0, T-t_0\right) = \Big\{\xi\left(x_0, T-t_0\right), \text{ such that}$$

$$\begin{aligned} &\frac{1}{2r\left(r+\delta\right)^2}\Big(\left(\alpha\left(r+\delta\right)-\varphi_1\right)^2 - 2\varphi_1\left(2\alpha\left(r+\delta\right) - \left(\varphi_2+\varphi_3\right)\right)\Big) \leq \xi_1\left(x_0, T-t_0\right) \\ &\quad \leq \frac{1}{2r\left(r+\delta\right)^2}\Big(\left(\alpha\left(r+\delta\right)-\varphi_1\right)^2 - 4\varphi_1\left(\alpha\left(r+\delta\right) - \left(\varphi_2+\varphi_3\right)\right) \\ &\qquad + 2\varphi_1^2 + \left(\varphi_2+\varphi_3\right)^2\Big), \end{aligned}$$

[7] A cooperative game is convex if

$$v(K \cup L) + v\left(K \cap L\right) \geq v(K) + v(L), \quad \forall K, L \subseteq I.$$

$$\frac{1}{2r\,(r+\delta)^2}\left((\alpha\,(r+\delta)-\varphi_2)^2-2\varphi_2\,(2\alpha\,(r+\delta)-(\varphi_1+\varphi_3))\right)\le \xi_2\,(x_0,T-t_0)$$
$$\le \frac{1}{2r\,(r+\delta)^2}\Big((\alpha\,(r+\delta)-\varphi_2)^2-4\varphi_2\,(\alpha\,(r+\delta)-(\varphi_1+\varphi_3))$$
$$+\,2\varphi_2^2+(\varphi_1+\varphi_3)^2\Big),$$

$$\frac{1}{2r\,(r+\delta)^2}\left((\alpha\,(r+\delta)-\varphi_3)^2-2\varphi_3\,(2\alpha\,(r+\delta)-(\varphi_1+\varphi_2))\right)\le \xi_2\,(x_0,T-t_0)$$
$$\le \frac{1}{2r\,(r+\delta)^2}\Big((\alpha\,(r+\delta)-\varphi_3)^2-4\varphi_3\,(\alpha\,(r+\delta)-(\varphi_1+\varphi_2))$$
$$+\,2\varphi_3^2+(\varphi_1+\varphi_2)^2\Big)$$
$$\text{and}\ \sum_{i=1}^{3}\xi_i\,(x_0,T-t_0)=\frac{3}{2r\,(r+\delta)^2}\,(\alpha\,(r+\delta)-(\varphi_1+\varphi_2+\varphi_3))^2\Bigg\}.$$

Following a similar procedure, we can compute the Shapley value in the subgame starting at $x^*\,(t)$ and determine the imputations in the core of that subgame, but it is not necessary to provide the details.

5.3.2 Noncooperative Outcomes

If at some intermediate instant of time $\tau, \tau \in (t_0,\infty)$, cooperation breaks down, then the players implement their Nash equilibrium strategies in the noncooperative game on (t_0,∞), with the initial pollution stock value given by

$$x^*\,(\tau)=3\left(\frac{\alpha\,(r+\delta)-(\varphi_1+\varphi_2+\varphi_3)}{\delta\,(r+\delta)}\right)\left(1-e^{-\delta\tau}\right),$$

that is, the value that would result from cooperation on $[t_0,\tau]$. Since the emission strategies are constant, i.e., they are state independent, their values are the same as computed above, namely,

$$e_i^{nc}=\frac{\alpha\,(r+\delta)-\varphi_i}{r+\delta},\quad i=1,2,3.$$

The resulting state trajectory is obtained by solving on $[\tau,\infty)$ the following differential equation:

$$\dot{x}\,(t)=\sum_{i\in N}\left(\frac{\alpha\,(r+\delta)-\varphi_i}{r+\delta}\right)-\delta x(t),$$

with the initial condition

$$x_\tau = x^*(\tau) = 3\left(\frac{\alpha(r+\delta) - (\varphi_1 + \varphi_2 + \varphi_3)}{\delta(r+\delta)}\right)\left(1 - e^{-\delta\tau}\right).$$

We get

$$x^{nc}(t) = x_\tau e^{\delta(\tau - t)} + \left(\frac{3\alpha(r+\delta) - (\varphi_1 + \varphi_2 + \varphi_3)}{\delta(r+\delta)}\right)\left(1 - e^{\delta(\tau - t)}\right).$$

The Nash equilibrium outcome of player i in the noncooperative game starting at $x_\tau = x^*(\tau)$ is then given by

$$W_i^{nc}\left(x^*(\tau)\right) = \int_\tau^\infty e^{-r(\tau - t)}\left(\alpha e_i^{nc} - \frac{1}{2}\left(e_i^{nc}\right)^2 - \varphi_i x^{nc}(t)\right) dt.$$

5.3.3 Time-Consistent Shapley Value

To design a time-consistent decomposition over time of the Shapley value, that is, to determine the value of $\alpha_i(t)$ in (13.13), we make the following observations. First, the two relevant quantities to be compared by player i at any $\tau \in (t_0, \infty)$ along the cooperative trajectory are $W_i^{nc}(x^*(\tau))$ and $Sh_i(x^*(\tau))$, that is, her noncooperative payoff-to-go and Shapley value outcomes in the subgame starting at $x^*(\tau)$. As the Shapley value is an imputation in this subgame, then clearly,

$$Sh_i\left(x^*(\tau)\right) \geq W_i^{nc}\left(x^*(\tau)\right), \quad \forall \tau \in (t_0, \infty).$$

Second, along the trajectory $x^*(t)$ on the time interval $[t, \infty)$, $t_0 \leq t \leq T$, the grand coalition N obtains the payoff

$$v(N; x^*(t)) = \left[\int_t^\infty \sum_{i \in N} e^{-r(\tau - t)}\left(\alpha e_i^* - \frac{1}{2}\left(e_i^*\right)^2 - \varphi_i x^*(\tau)\right) d\tau\right],$$

where e_i^* and x^* are given in (13.18)–(13.19). Then, the difference

$$v(N; x_0) - v(N; x^*(t)) = \int_{t_0}^t \sum_{i \in N} e^{-r\tau}\left(\alpha e_i^* - \frac{1}{2}\left(e_i^*\right)^2 - \varphi_i x^*(\tau)\right) d\tau,$$

is the payoff that the grand coalition N realizes on the time interval $[t_0, t]$. The share of the ith player in the above payoff is given by

$$\gamma_i(t) = \int_{t_0}^{t} \beta_i(\tau) \left(\sum_{i \in N} e^{-r\tau} \left(\alpha e_i^* - \frac{1}{2} \left(e_i^*\right)^2 - \varphi_i x^*(\tau) \right) \right) d\tau = \gamma_i(x^*(t), \beta), \tag{13.20}$$

where $\beta_i(\tau), i \in N$, is the $[t_0, \infty)$ integrable function satisfying the condition

$$\sum_{i=1}^{n} \beta_i(t) = 1, \quad t_0 \leq t \leq T. \tag{13.21}$$

Third, any time-consistent Shapley value distribution procedure must satisfy the two following conditions:

$$Sh_i(x_0, T - t_0) = \int_{t_0}^{\infty} e^{-rt} \alpha_i(t) dt,$$

$$Sh_i(x_0, T - t_0) = \gamma_i(x^*(t), \beta) + e^{-rt} Sh_i\left(x^*(t)\right),$$

$$= \int_{t_0}^{t} e^{-r\tau} \alpha_i(\tau) d\tau + e^{-rt} Sh_i\left(x^*(t)\right).$$

The first condition states that the total discounted IDP outcome to player i must be equal to her Shapley value in the whole game. The second condition ensures time consistency of this allocation, namely, that at any intermediate instant of time t along the cooperative state trajectory $x^*(t)$, the amount already allocated plus the Shapley value in the subgame at that time must be equal to the Shapley value in the whole game.

The final step is to determine a time function $\alpha_i(t), i \in N$ that satisfies the above two conditions. It is easy to verify that the following formula does so:

$$\alpha_i(t) = r \cdot Sh_i\left(x^*(t)\right) - \frac{d}{dt} Sh_i\left(x^*(t)\right).$$

The above IDP allocates at instant of time t to player i a payoff corresponding to the interest payment (interest rate times her payoff-to-go under cooperation given by her Shapley value) minus the variation over time of this payoff-to-go. We note that the above formula holds true independently of the functional forms involved in the problem.

6 Strongly Time-Consistent Solutions

In the previous section, we established that, for a time-consistent imputation $\xi^0 \in W_v(x_0, T - t_0)$, there is (by construction) an integrable vector function $\beta(t)$ and an imputation ξ^t from the solution $W_v(x^*(t), T-t)$ of the current game $\Gamma_v(x^*(t), T-t)$, neither of which, in general, is unique, such that:

$$\xi^0 = \gamma(x^*(t), \beta) + \xi^t,$$

for each $t \in [t_0, T]$, where $\gamma(x^*(t), \beta)$ is the vector of total payoffs to the players up to time t.

In this section, we raise the following question: If at any instant $t \in [t_0, T]$, the players decide along the cooperative state trajectory to select any imputation $(\xi^t)' \in W_v(x^*(t), T-t)$, then will the new imputation $(\xi^0)' = \gamma(x^*(t), \beta) + (\xi^t)'$ be optimal in $\Gamma_v(x(t_0), T-t_0)$? In other words, will $(\xi^0)' \in W_v(x_0, T-t_0)$? Unfortunately, this does not hold true in general. Still, it is interesting from both a theoretical and practical perspective to verify the conditions under which the response is affirmative. In fact, this will be the case if $\xi^0 \in W_v(x_0, T-t_0)$ is a time-consistent imputation and if, for every $\xi^t \in W_v(x^*(t), T-t)$, the condition

$$\gamma(x^*(t), \beta) + \xi^t \in W_v(x_0, T-t_0)$$

is satisfied. By slightly strengthening this requirement, we obtain the concept of strong time consistency.

Definition 5. The imputation $\xi^0 \in W_v(x_0, T-t_0)$ is called strongly time consistent (STC) in the game $\Gamma_v(x_0, T-t_0)$, if the following conditions are satisfied:

(1) The imputation ξ^0 is time consistent;
(2) For any $t_0 \le t_1 \le t_2 \le T$ and $\beta(t)$ corresponding to the imputation ξ^0, we have:

$$\gamma(x^*(t_2), \beta)) \oplus W_v(x^*(t_2), T-t_2) \subset \gamma(x^*(t_1), \beta)) \oplus W_v(x^*(t_1), T-t_1). \tag{13.22}$$

Definition 6. The cooperative differential game $\Gamma_v(x_0, T-t_0)$ with side payments has a strongly time-consistent solution $W_v(x_0, T-t_0)$ if all imputations from $W_v(x_0, T-t_0)$ are strongly time consistent.

To illustrate, let us first consider the simplest case where the players have only terminal payoffs, that is,

$$K_i(x_0, T-t_0; u_1, \ldots, u_n) = H_i(x(T)) \quad i = 1, \ldots, n,$$

which is obtained from (13.2) by setting $h_i \equiv 0$ for all i. The resulting cooperative differential game with terminal payoffs is again denoted by $\Gamma_v(x_0, T-t_0)$, and we write

$$H(x^*(T)) = (H_1(x^*(T)), \cdots, H_n(x^*(T))),$$

for the vector whose components are the payoffs resulting from the implementation of the optimal state trajectory. It is clear that, in the cooperative differential game $\Gamma_v(x_0, T-t_0)$ with terminal payoffs $H_i(x(T))$, $i = 1, \ldots, n$, only the vector $H(x^*) = \{H_i(x^*), \quad i = 1, \ldots, n\}$ may be time consistent.

It follows from the time consistency of the imputation $\xi^0 \in W_v(x_0, T - t_0)$ that

$$\xi^0 \in \bigcap_{t_0 \le t \le T} W_v(x^*(t), T - t).$$

But since the current game $\Gamma_v(x^*(T), 0)$ is of zero duration, then

$$L_v(x^*(T), 0) = W_v(x^*(T), 0) = H(x^*(T)) = H(x^*).$$

Hence,

$$\bigcap_{t_0 \le t \le T} W_v(x^*(t), T - t) = H(x^*(T)),$$

i.e., $\xi^0 = H(x^*(T))$ and there are no other imputations.

Thus, for the existence of a time-consistent solution in the game with terminal payoffs, it is necessary and sufficient that, for all $t_0 \le t \le T$,

$$H(x^*(T)) \in W_v(x^*(t), T - t).$$

Therefore, if, in the game with terminal payoffs, there is a time-consistent imputation, then the players at the initial state x_0 have to agree upon the realization of the vector (imputation) $H(x^*) \in W_v(x_0, T - t_0)$, and, with the motion along the optimal trajectory $x^*(t)$, at each instant $t_0 \le t \le T$, this imputation $H(x^*)$ belongs to the solution of the current game $\Gamma_v(x^*(t), T - t)$.

This shows that, in the game with terminal payoffs, only one imputation from the set $W_v(x_0, T - t_0)$ may be time consistent. This is very demanding since it requires that the imputation $H(x^*(T))$ belongs to the solutions of all subgames along the optimal state trajectory. Consequently, there is no point in such games in distinguishing between time-consistent and strongly time-consistent solutions.

Now, let us turn to the more general case where the payoffs are not collected only at the end of the game. It can easily be seen that the optimality principle $W_v(x_0, T - t_0)$ is strongly time consistent if for any $\bar{\xi} \in W_v(x_0, T - t_0)$ there exists an IDP $\tilde{\alpha}(t)$, $t \in [t_0, T]$ such that

$$\int_{t_0}^{t} \tilde{\alpha}(\tau) d\tau \oplus W_v(x^*(t), T - t) \subset W_v(x_0, T - t_0),$$

for all $t \in [t_0, T]$ (here, $a \oplus A$, where $a \in R^n$, $A \subset R^n$, means the set of vectors $a + b$, $b \in A$) and

$$\tilde{\alpha}(\tau) = \beta(\tau) \sum_{i=1}^{n} h_i(x^*(\tau), u_1^*(\tau), \ldots, u_n^*(\tau)).$$

If the IDP $\tilde{\alpha}(t)$ is implemented, then, on the time interval $[t_0, t]$, player i would collect the amount

$$\int_{t_0}^{t} \widetilde{\alpha}_i(\tau)d\tau, \quad i = 1, \ldots, n.$$

Strong time consistency means that, if the imputation $\xi \in W_v(x_0, T - t_0)$ and an IDP $\tilde{\alpha}(t)$ of ξ are selected, then after cashing the above, any optimal income (in the sense of the current optimality principle $W_v(x^*(t), T - t)$) on the time interval $[t, T]$ in the subgame $\Gamma_v(x^*(t), T - t)$ together with $\int_{t_0}^{t} \widetilde{\alpha}_i(\tau)d\tau$ constitutes an imputation belonging to the optimality principle $W_v(x_0, T - t_0)$ in the original game $\Gamma_v(x_0, T - t_0)$. To get a clearer picture here, think of the optimality principle as a particular solution of a cooperative game, e.g., the core. The condition of strong time consistency requires that the players stick to the same solution concept throughout the game. Of course, if the retained optimality principle (solution of a cooperative game) has a unique imputation, as in, e.g., the Shapley value and the nucleolus, then the two notions coincide. When this is not the case, as for, e.g., the core or stable set, then strong time consistency is a desirable feature. However, the existence of at least one strongly time-consistent imputation is far from being a given.

In the rest of this section, we show that, by making a linear transformation of the characteristic function, we can construct a strongly time-consistent solution. Let $v(S; x^*(t), T - t)$, $S \subset N$ be any characteristic function in the subgame $\Gamma_v(x^*(t), T - t)$ with an initial condition on the cooperative trajectory. Define a new characteristic function $\bar{v}(S; x^*(t), T - t)$ in the game $\Gamma_v(x^*(t), T - t)$ obtained by the following transformation:

$$\bar{v}(S; x_0, T - t_0) = -\int_{t_0}^{T} v(S; x^*(\tau), T - \tau) \frac{v'^*(x(\tau), T - \tau)}{v(N; x^*(\tau), T - \tau)} d\tau,$$

where

$$v'^*(\tau), T - \tau) = \frac{d}{d\tau} v(N; x^*(\tau), T - \tau) = -\left[\sum_{i=1}^{n} h_i(x^*(\tau), u_1^*(\tau), \ldots, u_n^*(\tau))\right].$$

Here $h_i(x^*(\tau), u_1^*(\tau), \ldots, u_n^*(\tau))$ is the instantaneous payoff of player $i \in N$ along the cooperative trajectory at time $\tau \in [t_0, T]$.

Remark 5. We make the following observations regarding this transformation:

1. The characteristic functions $v(\cdot)$ and $\bar{v}(\cdot)$ have the same value for the grand coalition in the whole game, that is,

$$\bar{v}(N; x_0, T - t_0) = v(N; x_0, T - t_0).$$

2. If $v(S; x_0, T - t_0)$ is superadditive, then $\bar{v}(S; x_0, T - t_0)$ is also superadditive.

Similarly, in the subgame $\Gamma_v(x^*(t), T - t)$ we have

$$\bar{v}(S; x^*(t), T - t) = -\int_t^T v(S; x^*(\tau), T - \tau) \frac{v'^* (x(\tau), T - \tau)}{v(N; x^*(\tau), T - \tau)} d\tau.$$

Let $L_v(x^*(t), T - t)$ be the set of imputations defined in the game $\Gamma_v(x^*(t), T - t)$ with characteristic function $v(S; x^*(t), T - t)$. Choose a selector $\xi(t) \in L_v(x^*(t), T - t)$ and define the following quantities:

$$\bar{\xi} = -\int_{t_0}^T \xi(\tau) \frac{v'^* (x(\tau), T - \tau)}{v(N; x^*(\tau), T - \tau)} d\tau,$$

$$\bar{\xi}(t) = -\int_t^T \xi(\tau) \frac{v'^* (x(\tau), T - \tau)}{v(N; x^*(\tau), T - \tau)} d\tau, \quad t \in [t_0, T].$$

Now define the transformed imputation set $\overline{L}_v(x^*(t), T - t)$ in $\Gamma_v(x_0, T - t_0)$ as the set of all $\bar{\xi}(t)$ for all possible measurable selectors $\xi(\tau) \in L_v(x^*(\tau), T - \tau)$, $\tau \in [t, T]$. Let $\overline{W}_v(x^*(\tau), T - \tau)$ in $\Gamma_v(x^*(\tau), T - \tau)$ be the set of all imputations $\bar{\xi}(t)$ such that

$$\bar{\xi}(t) = -\int_t^T \xi(\tau) \frac{v'^* (x(\tau), T - \tau)}{v(N; x^*(\tau), T - \tau)} d\tau,$$

where $\xi(\tau)$ is a measurable selector $\xi(\tau) \in W_v(x^*(\tau), T - \tau)$. Consequently, if $W_v(x^*(\tau), T - \tau)$ is the core in $\Gamma_v(x^*(\tau), T - \tau)$, $\overline{W}_v(x^*(\tau), T - \tau)$, then it is also the core in $\Gamma_v(x^*(\tau), T - \tau)$ but defined for the linearly transformed characteristic function $\bar{v}(S; x_0, T - t_0)$. It can be easily seen that this new "regularized" core is strongly time consistent (Petrosyan 1993, 1995).

7 Cooperative Differential Games with Random Duration

In this section, we deal with the case where the terminal date of the differential game is random. This scenario is meant to represent a situation where a drastic exogenous event, e.g., a natural catastrophe, would force the players to stop playing the game at hand. Note that the probability of occurrence of such an event is independent of the players' actions.

A two-player zero-sum pursuit-evasion differential game with random terminal time was first studied in Petrosjan and Murzov (1966). The authors assumed that the probability distribution of the terminal date is known, and they derived the Bellman-Isaacs equation for this problem. Nonzero-sum differential games with random duration were analyzed in Petrosyan and Shevkoplyas (2000, 2003), and a Hamilton-Jacobi-Bellman (HJB) equation was obtained. The case of nonconstant discounting was considered in Marín-Solano and Shevkoplyas (2011).[8]

Consider an n-player differential game $\Gamma(x_0)$, starting at instant of time t_0 with the dynamics given by

$$\dot{x} = g(x, u_1, \ldots, u_n), \quad x \in R^n,\ u_i \in U \subseteq \text{comp}\, R^l, \tag{13.23}$$

$$x(t_0) = x_0. \tag{13.24}$$

As in Petrosjan and Murzov (1966) and Petrosyan and Shevkoplyas (2003), the terminal time T of the game is random, with a known probability distribution function $F(t), t \in [t_0, \infty)$. The function $h_i(\tau)$ is Riemann integrable on every interval $[t_0, t]$, that is, for every $t \in [t_0, \infty)$ there exists an integral $\int_{t_0}^{t} h_i(\tau)d\tau$.

The expected integral payoff of player i can be represented by the following Lebesgue-Stieltjes integral:

$$K_i(x_0, t_0, u_1, \ldots, u_n) = \int_{t_0}^{\infty} \left[\int_{t_0}^{t} h_i(\tau, x(\tau), u_1, \ldots, u_n) d\tau \right] dF(t), \ \ i = 1, \ldots, n. \tag{13.25}$$

For all admissible player strategies (controls), let the instantaneous payoff be a nonnegative function:

$$h_i(\tau, x(\tau), u_1, \ldots, u_2) \geq 0, \qquad \forall \ \ \tau \in [t_0, \infty). \tag{13.26}$$

Then, the following proposition holds true (see Kostyunin and Shevkoplyas 2011): Let the instantaneous payoff $h_i(t)$ be a bounded, piecewise continuous function of time, and let it satisfy the condition of nonnegativity (13.26). Then, the expectation of the integral payoff of player i (13.25) can be expressed in the following simple form:

$$K_i(t_0, x_0, u_1, \ldots, u_n) = \int_{t_0}^{\infty} h_i(\tau)(1 - F(\tau))d\tau, \quad i = 1, \ldots, n. \tag{13.27}$$

Moreover, integrals in (13.25) and (13.27) exist or do not exist simultaneously.

Let the game $\Gamma(x_0)$ evolve along the trajectory $x(t)$. Then at each time instant ϑ, $\vartheta \in (t_0; \infty)$ players enter a new game (subgame) $\Gamma(x(\vartheta))$ with initial state

[8] For an analysis of an optimal control problem with random duration, see, e.g., Boukas et al. (1990) and Chang (2004).

$x(\vartheta) = x$. Clearly, there is a probability $F(\vartheta)$ that the game $\Gamma(x_0)$ will be terminated before ϑ. Then, the probability of starting the subgame $\Gamma(x(\vartheta))$ is equal to $(1 - F(\vartheta))$. Consequently, the expected total payoff of player i is given by the following formula:

$$K_i(x, \vartheta, u_1, \ldots, u_n) = \int_{\vartheta}^{\infty} \left[\int_{\vartheta}^{t} h_i(\tau, x(\tau), u_1, \ldots, u_n) d\tau \right] dF_{\vartheta}(t), \tag{13.28}$$

where $F_{\vartheta}(t)$, $t \geq \vartheta$ is the conditional probability distribution function of the random terminal time in game $\Gamma(x(\vartheta))$. Restricting ourselves to only stationary processes, the expression of $F_{\vartheta}(t)$ reduces to

$$F_{\vartheta}(t) = \frac{F(t) - F(\vartheta)}{1 - F(\vartheta)}, \quad t \in [\vartheta, \infty). \tag{13.29}$$

Further, let us assume that there exists a density function $f(t) = F'(t)$. The conditional density function is then given by the following formula:

$$f_{\vartheta}(t) = \frac{f(t)}{1 - F(\vartheta)}. \tag{13.30}$$

Using (13.30), the total payoff for player i in the subgame $\Gamma(x(\vartheta))$ can then be expressed as follows:

$$K_i(x, \vartheta, u_1, \ldots, u_n) = \frac{1}{1 - F(\vartheta)} \int_{\vartheta}^{\infty} \left[\int_{\vartheta}^{t} h_i(\tau, x(\tau), u_1, \ldots, u_n) d\tau \right] f(t) dt.$$

Taking stock of the results in Kostyunin and Shevkoplyas (2011), equation (13.28) can be written equivalently as

$$K_i(x, \vartheta, u_1, \ldots, u_n) = \frac{1}{1 - F(\vartheta)} \int_{\vartheta}^{\infty} \Big((1 - F(\tau)) h_i(\tau, x(\tau), u_1, \ldots, u_n) \Big) d\tau.$$

Denote by $W(x, t)$ the value function for the considered problem. The Hamilton-Jacobi-Bellman equation for the problem at hand is as follows (see Shevkoplyas 2009 and Marín-Solano and Shevkoplyas 2011):

$$\frac{f(t)}{1 - F(t)} W = \frac{\partial W}{\partial t} + \max_{u} \left(h_i(x, u, t) + \frac{\partial W}{\partial x} g(x, u) \right). \tag{13.31}$$

Equation (13.31) can be used for finding feedback solutions for both the noncooperative and the cooperative game with corresponding subintegral function $h_i(x, u, t)$.

To illustrate in a simple way the determination of a time-consistent solution, let us assume an exponential distribution for the random terminal time T, i.e.,

$$f(t) = \lambda e^{-\lambda(t-t_0)}; \quad F(t) = 1 - e^{-\lambda(\tau-t_0)}; \quad \frac{f(t)}{1-F(t)} = \lambda, \tag{13.32}$$

where λ is the rate parameter. The integral payoff $K_i(\cdot)$ of player i is equivalent to the integral payoff of that player in the game with an infinite time horizon and constant discount rate λ, that is,

$$K_i(x_0, t_0, u_1, \ldots, u_n) = \int_{t_0}^{\infty} h(\tau)(1 - F(\tau))d\tau = \int_{t_0}^{\infty} h(\tau)e^{-\lambda(\tau-t_0)}d\tau.$$

It can readily be shown that the derived HJB equation (13.31) in the case of exponential distribution of the terminal time is the same as the well-known HJB equation for the problem with constant discounting with rate λ. In particular, one can see that for $\frac{f(t)}{1-F(t)} = \lambda$, the HJB equation (13.31) takes the following form (see Dockner et al. 2000):

$$\lambda W(x,t) = \frac{\partial W(x,t)}{\partial t} + \max_u \left(h_i(x,u,t) + \frac{\partial W(x,t)}{\partial x} g(x,u) \right). \tag{13.33}$$

The conclusion is that a problem with a random duration and an exponential distribution of T is equivalent to a deterministic problem with a constant discounting of payoffs. This fact was noted in Haurie (2005) for a multigenerational game model with a random game duration; see also Wrzaczek et al. (2014).

We remark that the term $\frac{f(t)}{1-F(t)}$ on the left-hand side of equation (13.31) is the well-known hazard function (or failure rate) in reliability theory, which is denoted by

$$\lambda(t) = \frac{f(t)}{1-F(t)}. \tag{13.34}$$

Using the definition of the hazard function (13.34), we get the following new form for the HJB equation in (13.31):

$$\lambda(t)W(x,t) = \frac{\partial W(x,t)}{\partial t} + \max_u \left[h_i(x,u,t) + \lambda(t)S_i(x(t)) + \frac{\partial W(x,t)}{\partial x} g(x,u) \right], \tag{13.35}$$

For exponential distribution (13.32), the hazard function is constant, that is, $\lambda(t) = \lambda$. So, inserting λ instead of $\lambda(t)$ into (13.35), we easily get the standard HJB equation for a deterministic game with a constant discounting rate λ of the utility function in (13.33); see Dockner et al. (2000).

7.1 A Time-Consistent Shapley Value

Suppose that the players agree to cooperate and coordinate their strategies to maximize their joint payoffs. Further, assume that the players adopt the Shapley value, which we denote by $Sh = \{Sh_i\}_{i=1,\ldots,n}$. Let $\alpha(t) = \{\alpha_i(t) \geq 0\}_{i=1,\ldots,n}$ be the corresponding IDP during the game $\Gamma(x_0, t_0)$, that is,

$$Sh_i = \int_{t_0}^{\infty} (1 - F(t))\alpha_i(t)dt, \quad i = 1, \ldots, n. \tag{13.36}$$

If there exists an IDP $\alpha(t) = \{\alpha_i(t) \geq 0\}_{i=1,\ldots,n}$ such that for any $\vartheta \in [t_0, \infty)$ the Shapley value $\bar{Sh}^{\vartheta} = \{\bar{Sh}_i^{\vartheta}\}$ in the subgame $\Gamma(x^*(\vartheta), \vartheta)$, $\vartheta \in [t_0, \infty)$ can be written as

$$\bar{Sh}_i^{\vartheta} = \frac{1}{(1 - F(\vartheta))} \int_{\vartheta}^{\infty} (1 - F(t))\alpha_i(t)dt, \quad i = 1, \ldots, n, \tag{13.37}$$

then the Shapley value in the game $\Gamma(x_0, t_0)$ can be represented in the following form:

$$\bar{Sh}_i = \int_{t_0}^{\vartheta} (1 - F(\tau))\beta_i(\tau)d\tau + (1 - F(\vartheta))\bar{Sh}_i^{\vartheta}, \ \forall\, \vartheta \in [t_0, \infty), \ i = 1, \ldots, n. \tag{13.38}$$

Differentiating (13.38) with respect to ϑ, we obtain the following expression for the IDP:

$$\alpha_i(\vartheta) = \frac{f(\vartheta)}{(1 - F(\vartheta))} \bar{Sh}_i^{\vartheta} - (\bar{Sh}_i^{\vartheta})', \qquad \vartheta \in [t_0, \infty), \quad i = 1, \ldots, n. \tag{13.39}$$

Using the hazard function, (13.39) can be rewritten in a simple form:

$$\alpha_i(\vartheta) = \lambda(\vartheta)\bar{Sh}_i^{\vartheta} - (\bar{Sh}_i^{\vartheta})', \qquad \vartheta \in [t_0, \infty), \quad i = 1, \ldots, n. \tag{13.40}$$

To sum up, the IDP (13.40) allows us to allocate over time, in a time-consistent manner, the respective Shapley values.

For a detailed analysis of a similar setting, see Shevkoplyas (2011). The described model with a random time horizon has been extended in other ways. In Kostyunin et al. (2014), the case of asymmetric players was considered. Namely, it was assumed that the players may leave the game at different instants of time. For this case, the HJB equation was derived and solved for a specific application of resource extraction. Furthermore, in Gromova and Lopez-Barrientos (2015), the problem was considered with random initial times for asymmetric players.

In Gromov and Gromova (2014), a new class of differential games with regime switches was considered. This formulation generalizes the notion of differential games with a random terminal time by considering composite probability distribution functions. It was shown that the optimal solutions for this class of games can be obtained using the methods of hybrid optimal control theory (see also Gromov and Gromova 2016).

8 Concluding Remarks

The sustainability of a cooperative agreement over time is a challenging issue both in theory and in applications of dynamic games in many areas, e.g., economics, environmental agreements, and management science. This chapter reviewed how a time-consistent solution can be constructed in deterministic differential games with transferable payoffs. The only stochastic aspect we dealt with is the case where the terminal time is random. Time consistency in stochastic dynamic games is the subject of Petrosyan et al. (2004), Baranova and Petrosjan (2006), and Parilina (2014, 2015). For a comprehensive coverage of cooperative stochastic differential games, see the books by Yeung and Petrosjan (2005a, 2012). For examples of an implementation of time-consistent solutions in this class of games, see, e.g., Yeung and Petrosjan (2004, 2005b) and Yeung et al. (2007). Recently, some attention has been devoted to the class of dynamic games played over event trees, that is, stochastic games where the transition between states is not affected by players' actions. Reddy et al. (2013) proposed a node-consistent Shapley value, and Parilina and Zaccour (2015b) a node-consistent core for this class of games. Finally, we note a notable surge of applications of dynamic cooperative game theory in areas such as telecommunications, social networks, and power systems, where time consistency may be an interesting topic to study; see, e.g., Bauso and Basar (in print), Bayens et al. (2013), Opathella et al. (2013), Saad et al. (2009, 2012), and Zhang et al. (2015).

References

Angelova V, Bruttel LV, Güth W, Kamecke U (2013) Can subgame perfect equilibrium threats foster cooperation? An experimental test of finite-horizon folk theorems. Econ Inq 51: 1345–1356

Baranova EM, Petrosjan LA (2006) Cooperative stochasyic games in stationary strategies. Game Theory Appl XI:1–7

Başar T (1983) Performance bounds for hierarchical systems under partial dynamic information. J Optim Theory Appl 39:67–87

Başar T (1984) Affine incentive schemes for stochastic systems with dynamic information. SIAM J Control Optim 22(2):199–210

Başar T (1985) Dynamic games and incentives. In: Bagchi A, Jongen HT (eds) Systems and optimization. Lecture notes in control and information sciences, vol 66. Springer, Berlin/ New York, pp 1–13

Başar T (1989) Stochastic incentive problems with partial dynamic information and multiple levels of hierarchy. Eur J Polit Econ V:203–217. (Special issue on "Economic Design")

Başar T, Olsder GJ (1980) Team-optimal closed-loop Stackelberg strategies in hierarchical control problems. Automatica 16(4):409–414
Başar T, Olsder GJ (1995) Dynamic noncooperative game theory. Academic Press, New York
Bauso D, Basar T (2016) Strategic thinking under social influence: scalability, stability and robustness of allocations. Eur J Control 32:1–15. https://doi.org/10.1016/j.ejcon.2016.04.006i
Bayens E, Bitar Y, Khargonekar PP, Poolla K (2013) Coalitional aggregation of wind power. IEEE Trans Power Syst 28(4):3774–3784
Benoit JP, Krishna V (1985) Finitely repeated games. Econometrica 53(4):905–922
Boukas EK, Haurie A, Michel P (1990) An optimal control problem with a random stopping time. J Optim Theory Appl 64(3):471–480
Breton M, Sokri A, Zaccour G (2008) Incentive equilibrium in an overlapping-generations environmental game. Eur J Oper Res 185:687–699
Buratto A, Zaccour G (2009) Coordination of advertising strategies in a fashion licensing contract. J Optim Theory Appl 142:31–53
Cansever DH, Başar T (1985) Optimum/near optimum incentive policies for stochastic decision problems involving parametric uncertainty. Automatica 21(5):575–584
Chander P, Tulkens H (1997) The core of an economy with multilateral environmental externalities. Int J Game Theory 26:379–401
Chang FR (2004) Stochastic optimization in continuous time. Cambridge University Press, Cambridge
Chiarella C, Kemp MC, Long NV, Okuguchi K (1984) On the economics of international fisheries. Int Econ Rev 25:85–92
De Frutos J, Martín-Herrán G (2015) Does flexibility facilitate sustainability of cooperation over time? A case study from environmental economics. J Optim Theory Appl 165:657–677
De Giovanni P, Reddy PV, Zaccour G (2016) Incentive strategies for an optimal recovery program in a closed-loop supply chain. Eur J Oper Res 249:605–617
Dockner EJ, Jorgensen S, Van Long N, Sorger G (2000) Differential games in economics and management science. Cambridge University Press, Cambridge
Dutta PK (1995) A folk theorem for stochastic games. J Econ Theory 66:1–32
Ehtamo H, Hämäläinen RP (1986) On affine incentives for dynamic decision problems. In: Basar T (ed) Dynamic games and applications in economics. Springer, Berlin, pp 47–63
Ehtamo H, Hämäläinen RP (1989) Incentive strategies and equilibria for dynamic games with delayed information. J Optim Theory Appl 63:355–370
Ehtamo H, Hämäläinen RP (1993) A cooperative incentive equilibrium for a resource management problem. J Econ Dyn Control 17:659–678
Engwerda J (2005) Linear-quadratic dynamic optimization and differential games. Wiley, New York
Eswaran M, Lewis T (1986) Collusive behaviour in finite repeated games with bonding. Econ Lett 20:213–216
Flesch J, Kuipers J, Mashiah-Yaakovi A, Shmaya E, Shoenmakers G, Solan E, Vrieze K (2014) Nonexistene of subgame-perfect-equilibrium in perfect information games with infinite horizon. Int J Game Theory 43:945–951
Flesch J, Predtetchinski A (2015) On refinements of subgame perfect ε-equilibrium. Int J Game Theory. https://doi.org/10.1007/s00182-015-0468-8
Friedman JW (1986) Game theory with applications to economics. Oxford University Press, Oxford
Gao L, Jakubowski A, Klompstra MB, Olsder GJ (1989) Time-dependent cooperation in games. In: Başar TS, Bernhard P (eds) Differential games and applications. Springer, Berlin
Gillies DB (1953) Some theorems on N-person games. Ph.D. Thesis, Princeton University
Gromov D, Gromova E (2014) Differential games with random duration: a hybrid systems formulation. Contrib Game Theory Manag 7:104–119
Gromov D, Gromova E (2016) On a class of hybrid differential games. Dyn Games Appl. https://doi.org/10.1007/s13235-016-0185-3

Gromova EV, Lopez-Barrientos JD (2015) A differential game model for the extraction of nonrenewable resources with random initial times. Contrib Game Theory Manag 8:58–63

Haurie A (1976) A note on nonzero-sum differential games with bargaining solution. J Optim Theory Appl 18:31–39

Haurie A (2005) A multigenerational game model to analyze sustainable development. Ann Oper Res 137(1):369–386

Haurie A, Krawczyk JB, Roche M (1994) Monitoring cooperative equilibria in a stochastic differential game. J Optim Theory Appl 81:79–95

Haurie A, Krawczyk JB, Zaccour G (2012) Games and dynamic games. Scientific World, Singapore

Haurie A, Pohjola M (1987) Efficient equilibria in a differential game of capitalism. J Econ Dyn Control 11:65–78

Haurie A, Tolwinski B (1985) Definition and properties of cooperative equilibria in a two-player game of infinite duration. J Optim Theory Appl 46(4):525–534

Haurie A, Zaccour G (1986) A differential game model of power exchange between interconnected utilities. In: Proceedings of the 25th IEEE conference on decision and control, Athens, pp 262–266

Haurie A, Zaccour G (1991) A game programming approach to efficient management of interconnected power networks. In: Hämäläinen RP, Ehtamo H (eds) Differential games – developments in modelling and computation. Springer, Berlin

Jørgensen S, Martín-Herrán G, Zaccour G (2003) Agreeability and time-consistency in linear-state differential games. J Optim Theory Appl 119:49–63

Jørgensen S, Martín-Herrán G, Zaccour G (2005) Sustainability of cooperation overtime in linear-quadratic differential game. Int Game Theory Rev 7:395–406

Jørgensen S, Zaccour G (2001) Time consistent side payments in a dynamic game of downstream pollution. J Econ Dyn Control 25:1973–1987

Jørgensen S, Zaccour G (2002a) Time consistency in cooperative differential game. In: Zaccour G (ed) Decision and control in management science: in honor of professor Alain Haurie. Kluwer, Boston, pp 349–366

Jørgensen S, Zaccour G (2002b) Channel coordination over time: incentive equilibria and credibility. J Econ Dyn Control 27:801–822

Kaitala V, Pohjola M (1990) Economic development and agreeable redistribution in capitalism: efficient game equilibria in a two-class neoclassical growth model. Int Econ Rev 31:421–437

Kaitala V, Pohjola M (1995) Sustainable international agreements on greenhouse warming: a game theory study. Ann Int Soc Dyn Games 2:67–87

Kostyunin S, Palestini A, Shevkoplyas E (2014) On a nonrenewable resource extraction game played by asymmetric firms. J Optim Theory Appl 163(2):660–673

Kostyunin SY, Shevkoplyas EV (2011) On the simplification of the integral payoff in differential games with random duration. Vestnik St Petersburg Univ Ser 10(4):47–56

Kozlovskaya NV, Petrosyan LA, Zenkevich NA (2010) Coalitional solution of a game-theoretic emission reduction model. Int Game Theory Rev 12(3):275–286

Lehrer E, Scarsini M (2013) On the core of dynamic cooperative games. Dyn Games Appl 3: 359–373

Mailath GJ, Postlewaite A, Samuelson L (2005) Contemporaneous perfect epsilon-equilibria. Games Econ Behav 53:126–140

Marín-Solano J, Shevkoplyas EV (2011) Non-constant discounting and differential games with random time horizon. Automatica 47(12):2626–2638

Martín-Herrán G, Rincón-Zapatero JP (2005) Efficient Markov perfect Nash equilibria: theory and application to dynamic fishery games. J Econ Dyn Control 29:1073–1096

Martín-Herrán G, Taboubi S, Zaccour G (2005) A time-consistent open-loop Stackelberg equilibrium of shelf-space allocation. Automatica 41:971–982

Martín-Herrán G, Zaccour G (2005) Credibility of incentive equilibrium strategies in linear-state differential games. J Optim Theory Appl 126:1–23

Martín-Herrán G, Zaccour G (2009) Credible linear-incentive equilibrium strategies in linear-quadratic differential games. Advances in dynamic games and their applications. Ann Int Soci Dyn Games 10:1–31

Opathella C, Venkatesh B (2013) Managing uncertainty of wind energy with wind generators cooperative. IEEE Trans Power Syst 28(3):2918–2928

Ordeshook PC (1986) Game theory and political theory. Cambridge University Press, Cambridge, UK

Osborne MJ, Rubinstein A (1994) A course in game theory. MIT Press, Cambridge, MA

Parilina E (2014) Strategic stability of one-point optimality principles in cooperative stochastic games. Math Game Theory Appl 6(1):56–72 (in Russian)

Parilina EM (2015) Stable cooperation in stochastic games. Autom Remote Control 76(6): 1111–1122

Parilina E, Zaccour G (2015a) Approximated cooperative equilibria for games played over event trees. Oper Res Lett 43:507–513

Parilina E, Zaccour G (2015b) Node-consistent core for games played over event trees. Automatica 55:304–311

Petrosyan LA (1977) Stability of solutions of differential games with many participants. Vestnik Leningrad State Univ Ser 1 4(19):46–52

Petrosyan LA (1993) Strongly time-consistent differential optimality principles. Vestnik St Petersburg Univ Ser 1 4:35–40

Petrosyan LA (1995) Characteristic function of cooperative differential games. Vestnik St Petersburg Univ Ser 1 1:48–52

Petrosyan LA, Baranova EM, Shevkoplyas EV (2004) Multistage cooperative games with random duration, mathematical control theory, differential games. Trudy Inst Mat i Mekh UrO RAN 10(2):116–130 (in Russian)

Petrosjan L, Danilov NN (1979) Stability of solutions in nonzero sum differential games with transferable payoffs. J Leningrad Univ N1:52–59 (in Russian)

Petrosjan L, Danilov NN (1982) Cooperative differential games and their applications. Tomsk University Press, Tomsk

Petrosjan L, Danilov NN (1986) Classification of dynamically stable solutions in cooperative differential games. Isvestia of high school 7:24–35 (in Russian)

Petrosyan LA, Gromova EV (2014) Two-level cooperation in coalitional differential games. Trudy Inst Mat i Mekh UrO RAN 20(3):193–203

Petrosjan LA, Mamkina SI (2006) Dynamic games with coalitional structures. Int Game Theory Rev 8(2):295–307

Petrosjan LA, Murzov NV (1966) Game-theoretic problems of mechanics. Litovsk Mat Sb 6:423–433 (in Russian)

Petrosyan LA, Shevkoplyas EV (2000) Cooperative differential games with random duration Vestnik Sankt-Peterburgskogo Universiteta. Ser 1. Matematika Mekhanika Astronomiya 4: 18–23

Petrosyan LA, Shevkoplyas EV (2003) Cooperative solutions for games with random duration. Game theory and applications, vol IX. Nova Science Publishers, New York, pp 125–139

Petrosjan L, Zaccour G (2003) Time-consistent Shapley value of pollution cost reduction. J Econ Dyn Control 27:381–398

Petrosjan L, Zenkevich NA (1996) Game theory. World Scientific, Singapore

Radner R (1980) Collusive behavior in noncooperative epsilon-equilibria of oligopolies with long but finite lives. J Econ Theory 22:136–154

Reddy PV, Shevkoplyas E, Zaccour G (2013) Time-consistent Shapley value for games played over event trees. Automatica 49(6):1521–1527

Reddy PV, Zaccour G (2016) A friendly computable characteristic function. Math Soc Sci 82: 18–25

Rincón-Zapatero JP, Martín-Herrán G, Martinez J (2000) Identification of efficient Subgame-perfect Nash equilibria in a class of differential games. J Optim Theory Appl 104:235–242

Saad W, Han Z, Debbah M, Hjorungnes A, Başar T (2009) Coalitional game theory for communication networks [A tutorial]. IEEE Signal Process Mag Spec Issue Game Theory 26(5):77–97
Saad W, Han Z, Zheng R, Hjorungnes A, Başar T, Poor HV (2012) Coalitional games in partition form for joint spectrum sensing and access in cognitive radio networks. IEEE J Sel Top Sign Process 6(2):195–209
Shevkoplyas EV (2009) The Hamilton-Jacobi-Bellman equation for differential games with random duration. Upravlenie bolshimi systemami 26.1, M.:IPU RAN, pp 385–408 (in Russian)
Shevkoplyas EV (2011) The Shapley value in cooperative differential games with random duration. In: Breton M, Szajowski K (eds) Advances in dynamic games. Ann Int Soci Dyn Games 11(part4):359–373
Tolwinski B, Haurie A, Leitmann G (1986) Cooperative equilibria in differential games. J Math Anal Appl 119:182–202
Von Neumann J, Morgenstern O (1944) Theory of games and economic behavior. Princeton University Press, Princeton
Wrzaczek S, Shevkoplyas E, Kostyunin S (2014) A differential game of pollution control with overlapping generations. Int Game Theory Rev 16(3):1450005–1450019
Yeung DWK, Petrosjan L (2001) Proportional time-consistent solutions in differential games. In: Proceedings of international conference on logic, game theory and applications, St. Petersburg, pp 254–256
Yeung DWK, Petrosjan L (2004) Subgame consistent cooperative solutions in stochastic differential games. J Optim Theory Appl 120:651–666
Yeung DWK, Petrosjan L (2005a) Cooperative stochastic differential games. Springer, New York
Yeung DWK, Petrosjan L (2005b) Consistent solution of a cooperative stochastic differential game with nontransferable payoffs. J Optim Theory Appl 124:701–724
Yeung DWK, Petrosjan L (2006) Dynamically stable corporate joint ventures. Automatica 42: 365–370
Yeung DWK, Petrosjan L, Yeung PM (2007) Subgame consistent solutions for a class of cooperative stochastic differential games with nontransferable payoffs. Ann Int Soc Dyn Games 9:153–170
Yeung DWK (2006) An irrational-behavior-proof condition in cooperative differential games. Int Game Theory Rev 8(4):739–744
Yeung DWK, Petrosyan LA (2012) Subgame consistent economic optimization. Birkhauser, New York
Zaccour G (2003) Computation of characteristic function values for linear-state differential games. J Optim Theory Appl 117(1):183–194
Zaccour G (2008) Time consistency in cooperative differential games: a tutorial. INFOR 46(1): 81–92
Zhang B, Johari R, Rajagopal R (2015) Competition and coalition formation of renewable power producers. IEEE Trans Power Syst 30(3):1624–1632
Zheng YP, Başar T, Cruz JB (1982) Incentive Stackelberg strategies for deterministic multi-stage decision processes. IEEE Trans Syst Man Cybern SMC-14(1):10–20

Nontransferable Utility Cooperative Dynamic Games

14

David W. K. Yeung and Leon A. Petrosyan

Contents

D. W. K. Yeung (✉)
SRS Consortium for Advanced Study in Dynamic Games, Hong Kong Shue Yan University, Hong Kong, China
e-mail: dwkyeung@hksyu.edu

L. A. Petrosyan
Faculty of Applied Mathematics-Control Processes, St Petersburg State University, St Petersburg, Russia
e-mail: spbuoasis7@petrlink.ru

T. Başar, G. Zaccour (eds.), *Handbook of Dynamic Game Theory*,
https://doi.org/10.1007/978-3-319-44374-4_13

Abstract

Cooperation in an inter-temporal framework under nontransferable utility/payoffs (NTU) presents a highly challenging and extremely intriguing task to game theorists. This chapter provides a coherent analysis on NTU cooperative dynamic games. The formulations of NTU cooperative dynamic games in continuous time and in discrete time are provided. The issues of individual rationality, Pareto optimality, and an individual player's payoff under cooperation are presented. Monitoring and threat strategies preventing the breakup of the cooperative scheme are presented. Maintaining the agreed-upon optimality principle in effect throughout the game horizon plays an important role in the sustainability of cooperative schemes. The notion of time (subgame optimal trajectory) consistency in NTU differential games is expounded. Subgame consistent solutions in NTU cooperative differential games and subgame consistent solutions via variable payoff weights in NTU cooperative dynamic games are provided.

Keywords

Cooperative games · Nontransferable utility · Differential games · Dynamic games · Time consistency · Group optimality · Individual rationality · Subgame consistency · Variable weights · Optimality principle

1 Introduction

Cooperation suggests the possibility of socially optimal and group-efficient solutions to decisions involving strategic actions, including dynamic game problems. However, in some cases utilities/payoffs of the players may not be transferable. It is well known that utilities or payoffs in areas like economics and national security are nontransferable or noncomparable among the participating agents. Game theoretic decisions concerning utility/payoff from consumption, national defense, social development, coalition building, and political establishments often fall into the category of nontransferable payoffs. For instance, national (or social) planning is used to optimize the system-wide efficiency and well-being of multi-operator systems. It tries to maximize the system's benefits by coordinating different operators' strategies and managing externalities under some policy resolutions. The common practice of planning by gaining group optimality and neglecting individual rationality limits the applicability of national planning solutions. Recognizing that decisions are required to satisfy individual rationality in a real world system, cooperative game theory – which accounts for group optimality and individual rationality – can be used to resolve the limitation in the application of national planning. However, in national planning the well-being of the various operators may not be transferable or comparable. In a static framework, various cooperative solutions satisfying group optimality and individual rationality for nontransferable utility (NTU) games had been developed, like the Nash (1950, 1953) bargaining solution, the Kalai and Smorodinsky (1975) bargaining solution, Kalai's (1977) proportional solution, and the core by Edgeworth (1881).

Since human beings live in time, and decisions generally lead to effects over time, it is by no means an exaggeration to claim that life is a dynamic game. Very often, real-life games are of dynamic nature. In nontransferable utility/payoffs (NTU) cooperative dynamic games, transfer payments are not possible, and the cooperative payoffs are generated directly by the agreed-upon cooperative strategies. The identification of the conditions satisfying individual rationality throughout the cooperation period becomes extremely strenuous. This makes the derivation of an acceptable solution in nontransferable utility games much more difficult. To prevent deviant behaviors of players as the game evolves, monitoring and threat strategies are needed to reach a cooperative solution (see H*ä*m*ä*l*ä*inen et al. 1985). On top of the crucial factors of individual rationality and Pareto optimality, maintaining the agreed-upon solution throughout the game horizon plays an important role in the sustainability of the cooperative scheme. A stringent condition for sustainable cooperation is time (subgame optimal trajectory) consistency. A cooperative solution is time (subgame optimal trajectory) consistent if the optimality principle agreed upon at the outset remains in effect in any subgame starting at a later time with a state brought about by prior optimal behavior. Hence, the players do not have any incentive to deviate from the cooperation scheme.

While in cooperative dynamic games with transferrable payoffs the time (subgame optimal trajectory) consistency problem can be resolved by transfer payment (see Yeung and Petrosyan (2004 and 2012)), the NTU games face a much more challenging task. There is only a small literature on cooperative dynamic games with nontransferable payoffs. Leitmann (1974), Dockner and Jorgensen (1984), H*ä*m*ä*l*ä*inen et al. (1985), and Yeung and Petrosyan (2005), Yeung et al. (2007), de-Paz et al. (2013), and Marin-Solano (2014) studied continuous-time cooperative differential games with nontransferable payoffs. Haurie (1976) examined the property of dynamic consistency of direct application of the Nash bargaining solution in NTU cooperative differential games. Sorger (2006) and Yeung and Petrosyan (2015) presented discrete-time cooperative dynamic games with nontransferable payoffs.

This chapter provides a comprehensive analysis on NTU cooperative dynamic games. Sect. 2 provides the basic formulation of NTU cooperative differential games. The issues of individual rationality, Pareto optimal strategies, and an individual player's payoff under cooperation are presented. An illustration of the derivation of the Pareto optimal frontier and the identification of individually rational outcomes are given in Sect. 3. The formulation of a cooperative solution with monitoring and credible threats preventing deviant behaviors of the players is provided in Sect. 4. The notion of subgame consistency in NTU differential games is expounded in Sect. 5. A subgame consistent NTU resource extraction game is provided in the following section. In Sect. 7, discrete-time NTU cooperative dynamic games and the use of variable payoff weights to maintain subgame consistency are presented. In particular, a theorem for the derivation of cooperative strategies leading to a subgame consistent solution using variable weights is provided. Sect. 8 presents an example of a subgame consistent solution with variable weights in a NTU dynamic game in public good provision. Concluding remarks are given in Sect. 9.

2 NTU Cooperative Differential Game Formulation

Consider the general form of an n-person cooperative differential game with initial state x_0 and duration $[0, T]$. Player $i \in \{1, 2, \ldots, n\} \equiv N$ seeks to maximize the objective:

$$\int_0^T g^i[s, x(s), u_1(s), u_2(s), \ldots, u_n(s)]ds + q^i(x(T)), \tag{14.1}$$

where $x(s) \in X \subset R^m$ denotes the state variables of the game at time s, $q^i(x(T))$ is player i's valuation of the state at terminal time T, and $u_i \in U^i$ is the control of player i, for $i \in N$. The payoffs of the players are nontransferable.

The state variable evolves according to the dynamics:

$$\dot{x}(s) = f[s, x(s), u_1(s), u_2(s), \ldots, u_n(s)], x(0) = x_0. \tag{14.2}$$

The functions $f[s, x, u_1, u_2, \ldots, u_n]$, $g^i[s, \cdot, u_1, u_2, \ldots, u_n]$, and $q^i(\cdot)$, for $i \in N$ and $s \in [0, T]$ are differentiable functions.

To analyze the cooperative outcome, we first state the noncooperative equilibrium (if it exists) as a benchmark for negotiation in a cooperative scheme. If a noncooperative equilibrium exists, we let $\{u_i^*(t) = \phi_i^*(t, x) \in U^i$, for $i \in N\}$ denote a set of feedback Nash equilibrium strategies and $\{V^i(t, x), i \in N\}$ denote the Nash equilibrium payoffs of the players at time $t \in [0, T]$ given that state at time t is $x(t) = x$.

While noncooperative outcomes are (in general) not Pareto optimal, the players would consider cooperation to enhance their payoffs. Since payoffs are nontransferable, the cooperative payoff for an individual player is generated directly by the agreed-upon cooperative strategies. We use $\Gamma_c(x_0, T)$ to denote the cooperative differential game (14.1) and (14.2). Let $\{u_1^c(s), u_2^c(s), \ldots, u_n^c(s)\}$ denote a set of cooperative strategies agreed upon by the players, then the state dynamics becomes:

$$\dot{x}(s) = f[s, x(s), u_1^c(s), u_2^c(s), \ldots, u_n^c(s)], x(0) = x_0, \tag{14.3}$$

and we use $\{x^c(s)\}_{s=0}^T$ to denote the resulting cooperative state trajectory derived from (14.3). We also use the terms $x^c(s)$ and x_s^c along the cooperative trajectory interchangeably when there is no ambiguity.

The payoff of player i under cooperation, and from time t on becomes:

$$W^i(t, x) = \int_t^T g^i[s, x^c(s), u_1^c(s), u_2^c(s), \ldots, u_n^c(s)]ds + q^i(x^c(T)), \tag{14.4}$$

for $x^c(t) = x$, $i \in N$ and $t \in [0, T]$.

If at least one player deviates from the cooperative strategies, the game would revert back to a noncooperative game.

2.1 Individual Rationality

An essential factor for successful cooperation is individual rationality, which means that the payoff incurred to a player under cooperation will be no less than his payoff under noncooperation. Failure to guarantee individual rationality leads to the condition where the concerned participants would reject the agreed-upon cooperative scheme and play noncooperatively. In the case of cooperative dynamic games, individual rationality has to be satisfied for every player throughout the game horizon. Hence it is required that:

$$W^i(t, x^c(t)) \geqq V^i(t, x^c(t)), \tag{14.5}$$

for $i \in N$ and $t \in [0, T]$, along the cooperative state trajectory $\{x^c(T)\}_{t=0}^{T}$.

In general there exist cooperative strategies $\{u_1^c(s), u_2^c(s), \ldots, u_n^c(s)\}$ such that:

$$W^i(0, x^0(t)) \geq V^i(0, x^0(t)), \text{for } i \in N. \tag{14.6}$$

However, there is no guarantee that there exist cooperative strategies $\{u_1^c(s), u_2^c(s), \ldots, u_n^c(s)\}$ such that (14.5) is satisfied. At any time t, if player j's noncooperative payoff $V^j(t, x^c(t))$ is greater than his cooperative payoff $W^j(t, x^c(t))$, he has the incentive to play noncooperatively. This raises the issue of dynamic instability in cooperative differential games. Haurie (1976) discussed the problem of dynamic instability in extending the Nash bargaining solution to differential games.

2.2 Group Optimality

Another factor for successful cooperation is group optimality. Group optimality ensures that all potential gains from cooperation are captured. Failure to fulfill group optimality leads to the condition where the participants prefer to deviate from the agreed-upon solution plan in order to extract the unexploited gains. Group rationality requires the players to seek a set of cooperative strategies/controls that yields a Pareto optimal solution.

2.2.1 Pareto Optimal Strategies

Under cooperation, the players negotiate to establish an agreement on how to play the cooperative game that includes the adoption of a set of cooperative strategies and hence yields the cooperative payoffs to individual players. Pareto optimal outcomes for $\Gamma_c(x_0, T)$ can be identified by choosing a vector of payoff weights

$\alpha = (\alpha^1, \alpha^2, \ldots \alpha^n)$, for $\sum_{j=1}^{n} \alpha^j = 1$ and $\alpha^j > 0$, that solves the following control problem of maximizing weighted sum of payoffs (See Leitmann (1974)):

$$\max_{u_1(s), u_2(s), \ldots, u_n} \left\{ \sum_{j=1}^{n} \alpha^j \left(\int_t^T g^j[s, x(s), u_1(s), u_2(s), \ldots, u_n(s)] ds + q^j(x(T)) \right) \right\}, \tag{14.7}$$

subject to the dynamics (14.2).

Theorem 1. *A set of controls* $\left\{\left[\psi_1^\alpha(t,x), \psi_2^\alpha(t,x), \ldots, \psi_n^\alpha(t,x)\right], \textit{for } t \in [0,T]\right\}$ *provides an optimal solution to the control problem* (14.2) *and* (14.7) *if there exists a continuously differentiable function* $W^\alpha(t,x) : [0,T] \times R^m \to R$ *satisfying the following partial differential equation:*

$$\begin{aligned} -W_t^\alpha(t,x) = &\max_{u_1(s), u_2(s), \ldots, u_n(s)} \left\{ \sum_{j=1}^{n} \alpha^j g^j(t, x, u_1, u_2, \ldots, u_n) \right. \\ &\left. + W_x^\alpha(t,x) f(t, x, u_1, u_2, \ldots, u_n) \right\}, \\ &W^\alpha(T,x) = \sum_{j=1}^{n} q^j(x). \end{aligned} \tag{14.8}$$

Proof. The result follows directly from dynamic programming. ■

Substituting $\left[\psi_1^\alpha(t,x), \psi_2^\alpha(t,x), \ldots, \psi_n^\alpha(t,x)\right]$ for $t \in [0,T]$ into (14.2) yields the dynamics of the Pareto optimal trajectory associated with payoff weight α:

$$\dot{x}(s) = f[s, x(s), \psi_1^\alpha(s,x), \psi_2^\alpha(s,x), \ldots, \psi_n^\alpha(s,x)], x(0) = x_0. \tag{14.9}$$

We use $\{x^\alpha(s)\}_{s=0}^T$ to denote the Pareto optimal cooperative state trajectory under the payoff weights α. Again the terms $x^\alpha(s)$ and x_s^α will be used interchangeably if there is no ambiguity.

2.2.2 An Individual Player's Payoff Under Cooperation

In order to verify individual rationality in a cooperative scheme, we have to derive individual players' payoff functions under cooperation. To do this, we first substitute the optimal controls $\left[\psi_1^\alpha(t,x), \psi_2^\alpha(t,x), \ldots, \psi_n^\alpha(t,x)\right]$ into the objective functions (14.1) to derive the players' payoffs with α being chosen as the cooperative weights as:

$$\begin{aligned} W^{(\alpha)i}(t,x) = &\int_t^T g^i[s, x^\alpha(s), \psi_1^\alpha(s, x^\alpha(s)), \psi_2^\alpha(s, x^\alpha(s)), \ldots, \psi_n^\alpha(s, x^\alpha(s))] ds \\ &+ q^i(x^\alpha(T)), \end{aligned} \tag{14.10}$$

for $i \in N$ where $x^\alpha(t) = x$.

The following theorem characterizes individual cooperative payoffs, when the weights are given by α.

Theorem 2. *If there exist continuously differentiable functions* $W^{(\alpha)i}(t,x) : [0,T] \times R^m \longrightarrow R$*, for* $i \in N$*, satisfying*

$$-W_t^{(\alpha)i}(t,x) = g^i[t,x,\psi_1^\alpha(t,x),\psi_2^\alpha(t,x),\ldots,\psi_n^\alpha(t,x)] + W_x^{(\alpha)i}(t,x) f[t,x,\psi_1^\alpha(t,x),\psi_2^\alpha(t,x),\ldots,\psi_n^\alpha(t,x)],$$

$$W^{(\alpha)i}(T,x) = q^i(x), \text{for } i \in N; \tag{14.11}$$

then $W^{(\alpha)i}(t,x)$ *yields player* i*'s cooperative payoff over the interval* $[t,T]$ *with* α *being the cooperative weight and* $x^\alpha(t) = x$*.*

Proof. See Yeung (2004). ■

Theorem 2 is a deterministic version of the theorem in Yeung (2004). To maintain individual rationality throughout the game, the chosen α has to satisfy:

$$W^{(\alpha)i}(T,x) \geq V^i(t,x), \text{for } t \in [0,T] \text{ and all } i \in N; \tag{14.12}$$

with x being picked as a vector on the cooperative state trajectory $\{x^\alpha(s)\}_{s=0}^T$.

In general, there always exist Pareto optimal payoff weights α such that $W^{(\alpha)i}(0,x_0) \geqq V^i(0,x_0)$. However, there is no guarantee that there would exist any Pareto optimal payoff weights α such that individual rationality is satisfied throughout the game horizon as indicated in (14.12). We use Λ to denote the set of α such that (14.12) is satisfied, and Λ may be an empty set. If Λ is nonempty and if the players agree to choose a vector $\alpha \in \Lambda$ in their cooperation, a Pareto optimal cooperative scheme satisfying individual rationality will result.

3 Pareto Optimal and Individually Rational Outcomes: An Illustration

In this section an explicit NTU cooperative game is used to illustrate the derivation of the cooperative state trajectory, individual rationality, and Pareto optimality described in Sect. 2. We adopt a deterministic version of the renewable resource extraction game by Yeung et al. (2007) and consider a two-person nonzero-sum differential game with initial state x_0 and duration $[0,T]$. The state space of the game is $X \subset R^+$. The state dynamics of the game is characterized by the differential equation:

$$\dot{x}(s) = [a - bx(s) - u_1(s) - u_2(s)], x(0) = x_0 \in X, \tag{14.13}$$

where $u_i \in R^+$ is the control of player i, for $i \in \{1, 2\}$, a, b, and σ are positive constants. Equation (14.13) could be interpreted as the stock dynamics of a biomass of a renewable resource like fish or forest. The state $x(s)$ represents the resource size, $u_i(s)$ is the (nonnegative) amount of resource extracted by player i, a is the natural growth of the resource, and b is the rate of degradation.

At initial time 0, the payoff of player $i \in \{1, 2\}$ is:

$$J^i(0, x_0) = \int_0^T [h_i u_i(s) - c_i u_i(s)^2 x(s)^{-1} + k_i x(s)] e^{-rt} ds + e^{-rT} q_i x(T), \tag{14.14}$$

where h_i, c_i, k_i, and q_i are positive parameters and r is the discount rate.

The term $h_i u_i(s)$ reflects player i's satisfaction level obtained from the consumption of the resource extracted, and $c_i u_i(s)^2 x(s)^{-1}$ measures the dissatisfaction created in the extraction process. $k_i x(s)$ is the benefit to player i related to the existing level of the resource. Total utility of player i is the aggregate level of satisfaction. Payoffs in the form of utility are not transferable between the players. There exists a time discount rate r, and utility received at time t has to be discounted by the factor e^{-rt}. At time T, player i will receive a terminal benefit $q_i x(T)$, where q_i is a nonnegative constant.

3.1 Noncooperative Outcome and Pareto Optimal Trajectories

Invoking the standard techniques for solving differential games, a set of feedback strategies $\{u_i^*(t) = \phi_i^*(t, x)$, for $i \in \{1, 2\}$ and $t \in [0, T]\}$, provides a Nash equilibrium solution to the game (14.13)–(14.14) if there exist continuously differentiable functions $V^i(t, x) : [0, T] \times R \longrightarrow R$, $i \in \{1, 2\}$, satisfying the following partial differential equations:

$$\begin{aligned} V^i(t, x) &= e^{-rt} q_i x, \text{for } i \in \{1, 2\}, \text{and } j \neq i, \\ -V_t^i(t, x) &= \max_{u_i} \left\{ [h_i u_i - c_i u_i^2 x^{-1} + k_i x] e^{-rt} \right. \\ &\qquad \left. + V_x^i(t, x)[a - bx - u_i - \phi_i^*(t, x)] \right\}. \end{aligned} \tag{14.15}$$

Performing the indicated maximization in (14.15) yields:

$$\phi_i^*(t, x) = \frac{[h_i - V_x^i(t, x) e^{rt}] x}{2c}, \text{for } i \in \{1, 2\} \text{ and } x \in X. \tag{14.16}$$

Proposition 1. *The value function of the non cooperative payoff of player i in the game* (14.13) *and* (14.14) *is:*

$$V^i(t, x) = e^{-rt}[A_i(t)x + B_i(t)], \textit{for } i \in \{1, 2\} \textit{ and } t \in [\tau, T], \tag{14.17}$$

where $A_i(t)$ and $B_i(t)$, for $i \in \{1,2\}$, satisfy:

$$\dot{A}_i(t) = (r+b)A_i(t) - k_i - \frac{[h_i - A_i(t)]^2}{4c_i} + \frac{A_i(t)[h_j - A_j(t)]}{2c_j},$$
$$\dot{B}_i(t) = rB_i(t) - aA_i(t), \text{ for } i, j \in \{1,2\} \text{ and } i \neq j,$$
$$A_i(T) = q_i,\ B_i(T) = 0.$$

Proof. *Upon substitution of $\phi_i^*(t,x)$ from (14.16) into (14.15) yields a set of partial differential equations. One can readily verify that (14.17) is a solution to the set of equations (14.15).* ■

Consider the case where the players agree to cooperate in order to enhance their payoffs. If the players agree to adopt a weight $\alpha = (\alpha^1, \alpha^2)$, Pareto optimal strategies can be identified by solving the following optimal control problem:

$$\max_{u_1,u_2} \left\{\alpha^1 J^1(t_0,x_0) + \alpha^2 J^2(t_0,x_0)\right\}$$
$$\equiv \max_{u_1,u_2} \left\{ \int_0^T (\alpha^1[h_1u_1(s) - c_1u_1(s)^2x(s)^{-1} + k_1x(s)] \right.$$
$$+\alpha^2[h_2u_2(s) - c_2u_2(s)^2x(s)^{-1} + k_2x(s)])e^{-rt}ds$$
$$\left. +e^{-rt}[\alpha^1q_1x(T) + \alpha^2q_2x(T)]|x(t_0)\right\}, \tag{14.18}$$

subject to dynamics (14.13).

Invoking Theorem 1 in Sect. 2, the optimal solution of the control problem (14.13) and (14.18) can be characterized as.

Corollary 1. *A set of controls $\{ [\psi_1^\alpha(t,x), \psi_2^\alpha(t,x)]$, for $t \in [0,T]\}$ provides an optimal solution to the control problem (14.13) and (14.18), if there exists a continuously differentiable function $W^{(\alpha)}(t,x) : [0,T] \times R \longrightarrow R$ satisfying the partial differential equation:*

$$-W_t^\alpha(t,x) = \max_{u_1,u_2}\{(\alpha^1[h_1u_1 - c_1u_1^2x^{-1} + k_1x]$$
$$+\alpha^2[h_2u_2 - c_2u_2^2x^{-1} + k_2x])e^{-rt}$$
$$+W_x^\alpha(t,x)[a - bx - u_i - u_j],$$
$$W^\alpha(T,x) = e^{-rt}[\alpha^1q_1x(T) + \alpha^2q_2x(T)]. \tag{14.19}$$

Performing the indicated maximization in Corollary 1 yields:

$$\psi_1^\alpha(t,x) = \frac{[\alpha^1h_1 - W_x^\alpha(t,x)e^{rt}]x}{2\alpha^2c_1}, \text{ and}$$

$$\psi_2^\alpha(t,x) = \frac{[\alpha_2^2 h_2 - W_x^\alpha(t,x)e^{rt}]x}{2\alpha^2 c_2}, \tag{14.20}$$

for $t \in [t_0, T]$.

Proposition 2. *The maximized value function of the optimal control problem* (14.13) *and* (14.18) *is:*

$$W^\alpha(t,x) = \exp[-r(t-t_0)][A^\alpha(t)x + B^\alpha(t)], \tag{14.21}$$

for $t \in [0, T]$, *where* $A^\alpha(t)$ *and* $B^\alpha(t)$ *satisfy:*

$$\dot{A}^\alpha(t) = (r+b)A^\alpha(t) - \frac{[\alpha^1 h_1 - A^\alpha(t)]^2}{4\alpha^1 c_1} - \frac{[\alpha^2 h_2 - A^\alpha(t)]^2}{4\alpha^2 c_2} - k_1 - k_2,$$

$$\dot{B}^\alpha(t) = rB^\alpha(t) - A^\alpha(t)a,$$

$$A^\alpha(T) = \alpha^1 q_1 + \alpha^2 q_2 \ \textit{and}\ B^\alpha(T) = 0.$$

Proof. Upon substitution of $\psi_1^\alpha(t,x)$ and $\psi_2^\alpha(t,x)$ from (14.20) into (14.19) yields a partial differential equation. One can readily verify that (14.21) is a solution to equation (14.19). ∎

Substituting the partial derivatives $W_x^\alpha(t,x)$ into $\psi_1^\alpha(t,x)$ and $\psi_2^\alpha(t,x)$ yields the optimal controls of the problem (14.13) and (14.18) as:

$$\psi_1^\alpha(t,x) = \frac{[\alpha^1 h_1 - A^\alpha(t)]x}{2\alpha^1 c_1},$$

and

$$\psi_2^\alpha(t,x) = \frac{[\alpha^2 h_2 - A^\alpha(t)]x}{2\alpha^2 c_2}, \qquad \text{for } t \in [0,T]. \tag{14.22}$$

Substituting these controls into (14.13) yields the dynamics of the Pareto optimal trajectory associated with a weight α as:

$$\dot{x}(s) = a - bx(s) - \frac{[\alpha^1 h_1 - A^\alpha(s)]x(s)}{2\alpha^1 c_1} - \frac{[\alpha^2 h_2 - A^\alpha(s)]x(s)}{2\alpha^2 c_2}, x(0) = x_0 \in X. \tag{14.23}$$

Equation (14.23) is a first-order linear differential equation, whose solution $\{x^\alpha(s)\}_{s=0}^T$ can be obtained explicitly using standard solution techniques.

3.2 An Individual Player's Payoff Under Cooperation

In order to verify individual rationality, we have to derive the players' payoffs under cooperation. Substituting the cooperative controls in (14.22) into the players' payoff functions yields the payoff of player 1 as:

$$W^{(\alpha)1}(t,x)$$
$$= \int_t^T \left[\frac{h_1[\alpha^1 h_1 - A^\alpha(s)]x^\alpha(s)}{2\alpha^1 c_1} - \frac{[\alpha^1 h_1 - A^\alpha(s)]^2 x^\alpha(s)}{4(\alpha^1)^2 c_1} + k_1 x^\alpha(s) \right] e^{-rs} ds$$
$$+ e^{-rT} q_1 x^\alpha(T) | x^\alpha(t),$$

and the payoff of player 2 as:

$$W^{(\alpha)2}(t,x)$$
$$= \int_t^T \left[\frac{h_2[\alpha^2 h_2 - A^\alpha(s)]x^\alpha(s)}{2\alpha^2 c_2} - \frac{[\alpha^2 h_2 - A^\alpha(s)]^2 x^\alpha(s)}{4(\alpha^2)^2 c_2} + k_2 x^\alpha(s) \right] e^{-rs} ds$$
$$+ e^{-rT} q_2 x(T) | x^\alpha(t). \qquad (14.24)$$

Invoking Theorem 2 in Sect. 2, the value function $W^{(\alpha)1}(t,x)$ can be characterized as:

$$-W_t^{(\alpha)1}(t,x) = \left[\frac{h_1[\alpha^1 h_1 - A^\alpha(t)]x}{2\alpha^1 c_1} - \frac{[\alpha^1 h_1 - A^\alpha(t)]^2 x}{4(\alpha^1)^2 c_1} + k_1 x \right] e^{-rt} \qquad (14.25)$$
$$+ W_x^{(\alpha)1}(t,x) \left[a - bx - \frac{[\alpha^1 h_1 - A^\alpha(t)]x}{2\alpha^1 c_1} - \frac{[\alpha^2 h_2 - A^\alpha(t)]x}{2\alpha^2 c_2} \right],$$

for $x = x^\alpha(t)$.

Boundary conditions require:

$$W^{(\alpha)1}(T,x) = e^{-rT} q_1 x. \qquad (14.26)$$

If there exist continuously differentiable functions $W^{(\alpha)1}(t,x) : [0,T] \times R \longrightarrow R$ satisfying (14.25) and (14.26), then player 1's payoff in the cooperative game under the cooperation scheme with weight α is indeed $W^{(\alpha)1}(t,x)$.

Proposition 3. *The function $W^{(\alpha)1}(t,x)$ satisfying (14.25) and (14.26) can be solved as:*

$$W^{(\alpha)1}(t,x) = e^{-rt}[A_1^\alpha(t)x + B_1^\alpha(t)], \qquad (14.27)$$

where $A_1^\alpha(t)$ *and* $B_1^\alpha(t)$ *satisfy:*

$$\dot{A}_1^\alpha(t) = \left[r + b + \frac{[\alpha^1 h_1 - A^\alpha(t)]}{2\alpha^1 c_1} + \frac{[\alpha^2 h_2 - A^\alpha(t)]}{2\alpha^2 c_2} \right] A_1^\alpha(t) - \frac{[\alpha^1 h_1 - A^\alpha(t)][h_1 + A^\alpha(t)]}{4(\alpha^1)^2 c_1} - k_1,$$

$$\dot{B}_1^\alpha(t) = rB_1^\alpha(t) - aA_1^\alpha(t), \ \ A_1^\alpha(T) = q_1 \text{ and } B_1^\alpha(T) = 0.$$

Proof. Upon calculating the derivatives $W_t^{(\alpha)1}(t,x)$ and $W_x^{(\alpha)1}(t,x)$ from (14.27) and then substituting them into (14.25)–(14.26) yield Proposition 3. ■

Following a similar analysis, player 2's cooperative payoff under payoff weights α can be obtained as:

Proposition 4. *The function* $W^{(\alpha)2}(t,x)$ *can be solved as:*

$$W^{(\alpha)2}(t,x) = e^{-rt}[A_2^\alpha(t)x + B_2^\alpha(t)], \tag{14.28}$$

where $A_2^\alpha(t)$ *and* $B_2^\alpha(t)$ *satisfy:*

$$\dot{A}_2^\alpha(t) = \left[r + b + \frac{[\alpha^1 h_1 - A^\alpha(t)]}{2\alpha^1 c_1} + \frac{[\alpha^2 h_2 - A^\alpha(t)]}{2\alpha^2 c_2} \right] \hat{A}_2^{\alpha_1^\tau}(t) - \frac{[\alpha^2 h_2 - A^\alpha(t)][\alpha^2 h_2 + A^\alpha(t)]}{4\alpha^2 c_2} - k_2,$$
$$\dot{B}_2^\alpha(t) = rB_2^\alpha(t) - aA_2^\alpha(t), \ \ A_2^\alpha(T) = q_2 \text{ and } B_2^\alpha(T) = 0.$$

Proof. The proof follows that of Proposition 3. ■

For Pareto optimality and individual rationality to hold simultaneously the set of payoff weights α, they must satisfy $W^{(\alpha)i}(0, x_0) \geq V^i(0, x_0)$.

4 Monitoring and Threat Strategies

For games that are played over time and the payoffs are nontransferable, the derivation of a cooperative solution satisfying individual rationality throughout the cooperation duration becomes extremely difficult. To avoid the breakup of the scheme as the game evolves, monitoring and threats to prevent deviant behaviors of players will be adopted.

4.1 Bargaining Solution in a Cooperative Game

Consider the two-country fishery management game in H*ä*m*ä*l*ä*inen et al. (1985) in which the accumulation dynamics of the fish biomass $x(s) \in X \subseteq R^+$ is:

$$\dot{x}(s) = x(s)[\gamma - \beta \ln x(s) - u_1(s) - u_2(s)],\ x(0) = x_0, \tag{14.29}$$

where γ and β are constants and $u_i(s)$ is the fishing effort of country i.

The payoff to country i is:

$$\int_0^\theta \ln[u_i(s)x(s)]e^{-rt}ds, \text{for } i \in \{1,2\}. \tag{14.30}$$

In particular, the payoffs of the countries are not transferrable. By the transformation of variable $z(s) = \ln x(s)$, (14.29)–(14.30) becomes:

$$\dot{z}(s) = [\gamma - \beta z(s) - u_1(s) - u_2(s)], z(0) = z_0,$$

$$\int_0^\theta [\ln(u_i(s)) + z(s)]e^{-rt}ds, \text{for } i \in \{1,2\}. \tag{14.31}$$

First consider an open-loop Nash equilibrium of the game (14.31). Detailed analysis of open-loop solutions can be found in Başar and Olsder (1999). The corresponding current-value Hamiltonians can be expressed as:

$$H^i(z,\lambda,u_1,u_2) \equiv [\ln(u_i(s)) + z(s)]$$
$$+\lambda_i(s)[\gamma - \beta z(s) - u_1(s) - u_2(s)], \tag{14.32}$$

for $i \in \{1,2\}$, where $\lambda_i(s)$ is the costate variable.

Necessary conditions for an open-loop equilibrium include the optimal controls:

$$u_i(s) = \frac{1}{\lambda_i(s)}, \tag{14.33}$$

and the adjoint equations:

$$\dot{\lambda}_i(s) = -1 + \lambda_i(s)(\beta + r), \text{for } i \in \{1,2\}. \tag{14.34}$$

A constant solution to (14.34) is:

$$\bar{\lambda}_i = \frac{1}{(\beta + r)}$$

which yields the constant equilibrium strategy:

$$\bar{u}_i = \frac{1}{\bar{\lambda}_i} = \beta + r, \text{for } i \in \{1,2\}. \tag{14.35}$$

The constant costate vectors suggest a Bellman value function linear in the state variable, that is, $e^{-rt}V_i(z) = e^{-rt}(\bar{A}_i + \bar{B}_i z)$, for $i \in \{1, 2\}$, where

$$\bar{B}_i = \frac{1}{(\beta + r)},$$

and

$$\bar{A}_i = \frac{1}{r}\left[-2 + \ln(\beta + r) + \frac{\gamma}{\beta + r}\right], \text{for } i \in \{1, 2\}. \tag{14.36}$$

In this case the open-loop equilibrium is also a feedback Nash equilibrium because the players' strategies are independent of the state variable. One can readily verify the linear Bellman function by invoking the corresponding HJB equation for a feedback equilibrium solution as:

$$-\frac{\partial V_t^i(z)e^{-rt}}{\partial t} = \max_{u_i}\left\{[\ln(u_i) + z]e^{-rt} + V_z^i(z)e^{-rt}(\gamma - \beta z - u_i - \bar{u}_j)\right\}, \tag{14.37}$$

for $i \in \{1, 2\}$.

Performing the indicated maximization, one obtains the Bellman value function $e^{-rt}V^i(z) = e^{-rt}(\bar{A}_i + \bar{B}_i z)$.

Taking the Nash equilibrium outcome as the status quo with state z^0 at time t_0, one obtains the non cooperative payoff of player i as $e^{-rt_0}(\bar{A}_i + \bar{B}_i z^0)$, for $i \in \{1, 2\}$.

Now consider a Pareto optimal outcome obtained by solving the control problem with performance criterion:

$$J_\alpha^\theta = \int_{t_0}^{\theta} e^{-rt}[\alpha^1 \ln(u_1(s)x(s)) + \alpha^2 \ln(u_2(s)x(s))]ds, \tag{14.38}$$

and state equation (14.29), where $\alpha^1 > 0$, $\alpha^2 > 0$, and $\alpha^1 + \alpha^2 = 1$.

Again by the variable transformation $z(s) = \ln x(s)$, (14.38) can be expressed as:

$$J_\alpha^\theta \equiv \int_{t_0}^{\theta} e^{-rt}[\alpha^1 \ln(u_1(s) + z(s)) + \alpha^2 \ln(u_2(s) + z(s))]ds. \tag{14.39}$$

The maximum principle yields the adjoint differential equation:

$$\dot{q}(s) = -(\alpha^1 + \alpha^2) + q(s)(\beta + r)$$

and the optimal cooperative controls:

$$u_i(s) = \frac{\alpha^i}{q(s)}, \text{for } i \in \{1, 2\}. \tag{14.40}$$

A constant costate variable solution can be obtained as:

$$\bar{q} = \frac{1}{(\beta + r)}$$

which yields the constant cooperative optimal controls:

$$\bar{u}_i^* = \frac{1}{\bar{q}} = \alpha^i(\beta + r), \text{for } i \in \{1, 2\}. \tag{14.41}$$

The Bellman value function reflecting the payoff for each player under cooperation can be obtained as:
$e^{-rt}V_i^*(z) = e^{-rt}(\bar{A}_i^* + \bar{B}_i^* z)$, where $\bar{B}_i^* = \frac{1}{(\beta+r)}$, and

$$\bar{A}_i^* = \frac{1}{r}\left[-1 + \ln(\beta + r) + \frac{\gamma}{\beta + r} + \ln(\alpha^i)\right], \tag{14.42}$$

for $i \in \{1, 2\}$.

The agreed-upon optimality principle to allocation of the players' payoffs is to follow the Kalai-Smorodinsky bargaining solution (1975). In this case of symmetric players, the bargaining point corresponds to $\alpha^1 = \alpha^2 = 1/2$. Notice that the weighting $\alpha^1 = \alpha^2 = 1/2$ is not affected by the initial state z^0 at which the bargaining occurs. This property is due to the special structure of the model and will not be observed in more complicated systems.

4.2 Threats and Equilibria

If an arbitrator could enforce the agreement at (t_0, z^0) with player i getting the cooperative payoff $y_i^* = e^{-rt_0}(\bar{A}_i^* + \bar{B}_i^* z^0)$, then the two players would use a fishing effort $\bar{u}_i^* = \frac{1}{2}(\beta + r)$, which is half of the noncooperative equilibrium effort. In the absence of an enforcement mechanism, one player may be tempted to deviate at some time from the cooperative agreement. It is assumed that the deviation by a player will be noticed by the other player after a delay of δ. This will be called the cheating period. The non-deviating player will then have the possibility to retaliate by deviating from the cooperative behavior for a length of time of h. This can be regarded as the punishment period. At the end of this period, a new bargaining process will start, and a new cooperative solution will be obtained through an appropriate scheme. Therefore, each player may announce a threat corresponding to the control he will use in a certain punishment period if he detects cheating by the other player.

Consider the situation where country i announces that if country j deviates then it will use a fishing effort $u_i^m \geq \beta + r$ for a period of length h as retaliation. At $(t, z^*(t))$, country j expects an outcome $e^{-rt}V_j^*(z^*(t)) = e^{-rt}(\bar{A}_j^* + \bar{B}_j^* z(t))$ if it continues to cooperate. Thus $V_j^*(z^*(t))$ denotes the infinite-horizon bargaining game from the initial stock level $z^*(t)$ at time t for country j. If country j deviates, the best outcome it can expect is the solution of the following optimal control problem:

$$C_j(t, z^*(t)) = \max_{u_j} \int_t^{t+\delta+h} e^{-rs}[\ln(u_j(s) + z(s))]ds$$
$$+e^{-r(t+\delta+h)} V_j^*(z^*(t+\delta+h)) \tag{14.43}$$

subject to:

$$\dot{z}(s) = \begin{cases} \gamma - \beta z(s) - u_j(s) - \frac{\beta+r}{2}, \text{ if } t \le s \le t+\delta \\ \gamma - \beta z(s) - u_j(s) - u_i^m, \text{if } t+\delta \le s \le t+\delta+h \end{cases}, \tag{14.44}$$

$$z(t) = z^*(t).$$

The threat u_i^m and the punishment period h will be effective at $(t, z^*(t))$ if:

$$C_j(t, z^*(t)) \le e^{-rt} V_j^*(z^*(t)). \tag{14.45}$$

The values δ, h and u_i^m appear as design parameters in the optimal control problem (14.43) and (14.44). For some values of these parameters, the inequality (14.45) will be achieved. Then the threat of country i will prevent cheating at the point $(t, z^*(t))$ by country j. If the inequality (14.45) is not satisfied, then either the threat must be made more powerful or the agreement must be changed.

Any threat, to be effective in practice, must be credible. The credibility of a threat is a complicated matter. The threat of adopting noncooperative strategies after cheating has been detected provides an upper limit for the credible loss due to the implementation of punishment. Any threat which yields a payoff lower than that under a one-sided optimization by the threatening player will not be credible. Another necessary condition for a credible threat is that the payoff with the implementation of the punishment strategy will not be lower than the initially noncooperative solution. As pointed out in H*ä*m*ä*l*ä*inen et al. (1985), the use of this principle in developing general credibility conditions is difficult because the outcome of a noncooperative equilibrium generally depends on the value of the state in the system (with state-dependent control strategies).

5 Notion of Subgame Consistency in NTU Differential Games

In addition to individual rationality and Pareto efficiency, the sustainability of the agreed-upon solution is also of concern to the participating players. Haurie (1976) pointed out that the property of dynamic consistency, which is crucial in maintaining sustainability in cooperation, is absent in the direct application of the Nash bargaining solution in differential games. One of the ways to uphold sustainability of a cooperation scheme is to maintain the condition of time (subgame optimal trajectory) consistency. In particular, a cooperative solution is subgame consistent if the optimality principle agreed upon at the outset remains in effect in any subgame starting at a later time with a state brought about by prior optimal behavior. Hence the players do not have incentives to deviate from the previously adopted optimal behavior along the cooperative path. Subgame consistent solutions for differential games and dynamic games with transferable payoffs under deterministic and stochastic dynamics can be found in Petrosyan and Zenkevich (1996), Petrosyan (1997), Yeung and Petrosyan (2004, 2010).

In nontransferable payoff cooperative differential games, transfer payments among players are not possible, and the cooperative payoffs are generated directly by the agreed-upon cooperative strategies. This makes the derivation of a time-consistent solution in nontransferable games much more difficult. Moreover, in the case of subgame consistent solutions for dynamic games with transferable payoffs, full Pareto optimality can be achieved with the use of transfer payments. In the case of nontransferable payoffs, very often there do not exist cooperative strategies satisfying individual rationality which are also Pareto optimal. Yeung and Petrosyan (2005 and 2016) and Yeung et al. (2007) presented subgame consistent solutions in cooperative stochastic differential games with nontransferable payoffs for a restricted class of optimality principles.

Under cooperation with nontransferable payoffs, the players negotiate to establish an agreement (optimality principle) on how to play the cooperative game and how to distribute the resulting payoff. In particular, the chosen optimality principle has to satisfy group optimality and individual rationality. Subgame consistency requires that the optimality principle agreed upon at the outset remains in effect in any subgame starting at a later time with a state brought about by prior optimal behavior. Hence the players do not have incentives to deviate from the cooperation scheme.

Consider the cooperative differential game $\Gamma_c(x_0, T)$ with game structures (14.1) and (14.2) in which the players agree to an optimality principle. According to the agree-upon optimality principle, the players will:

(i) Adopt cooperative controls $\{u_1^c(s), u_2^c(s), \ldots, u_n^c(s)\}$, for $s \in [0, T]$, and the corresponding state dynamics $\dot{x}(s) = f[s, x(s), u_1^c(s), u_2^c(s), \ldots, u_n^c(s)]$, $x(t_0) = x_0$, with the resulting cooperative state trajectory $\{x^c(s)\}_{s=0}^T$

(ii) Receive an imputation $W^i(t, x) = \int_t^T g^i[s, x^c(s), u_1^c(s), u_2^c(s), \ldots, u_n^c(s)]ds + q^i(x^c(T))$, for $x^c(t) = x$, $i \in N$ and $t \in [0, T]$, where $\dot{x}^c(s) = f[s, x^c(s), u_1^c(s), u_2^c(s), \ldots, u_n^c(s)]$, $x^c(t_0) = x_0$

Now consider the game $\Gamma_c(x_\tau, T - \tau)$ at time τ where $x_\tau = x_\tau^c$ and $\tau \in [0, T]$, and under the same optimality principle, the players will:

(i) Adopt cooperative controls $\{\hat{u}_1^c(s), \hat{u}_2^c(s), \ldots, \hat{u}_n^c(s)\}$, for $s \in [\tau, T]$, and the cooperative trajectory generated is $\dot{\hat{x}}^c(s) = f[s, \hat{x}^c(s), \hat{u}_1^c(s), \hat{u}_2^c(s), \ldots, \hat{u}_n^c(s)]$, $\hat{x}^c(\tau) = x^c(\tau)$

(ii) Receive an imputation $\hat{W}^i(t, x) = \int_t^T g^i[s, \hat{x}^c(s), \hat{u}_1^c(s), \hat{u}_2^c(s), \ldots, \hat{u}_n^c(s)]ds + q^i(\hat{x}^c(T))$, for $x = \hat{x}^c(t)$, $i \in N$ and $t \in [\tau, T]$

A necessary condition is that the cooperative payoff is no less than the noncooperative payoff, that is, $W^i(t, x_t^c) \geq V^i(t, x_t^c)$ and $\hat{W}^i(t, \hat{x}_t^c) \geq V^i(t, \hat{x}_t^c)$.

A formal definition of time (subgame optimal trajectory) consistency can be stated as:

Definition 1. A cooperative solution is time (subgame optimal trajectory) consistent if $\{u_1^c(s), u_2^c(s), \ldots, u_n^c(s)] = \{\hat{u}_1^c(s), \hat{u}_2^c(s), \ldots, \hat{u}_n^c(s)\}$, for $s \in [\tau, T]$ and $\tau \in [0, T]$, which implies that $W^i(t, x_t^c) = \hat{W}^i(t, \hat{x}_t^c)$ for $t \in [\tau, T]$.

Note that if Definition 1 is satisfied, the imputation agreed upon at the outset of the game will be in any subgame $[\tau, T]$.

6 A Subgame Consistent NTU Cooperative Differential Game

Consider the renewable resource extraction game in Sect. 3 with the resource stock dynamics:

$$\dot{x}(s) = [a - bx(s) - u_1(s) - u_2(s)], x(0) = x_0 \in X, \tag{14.46}$$

and the payoff of players:

$$J^i(0, x_0) = \int_0^T [h^i u_i(s) - c^i u_i(s)^2 x(s)^{-1} + k_i x(s)] e^{-rt} ds + e^{-rT} q_i x(T), \tag{14.47}$$

for $i \in \{1, 2\}$.

To derive a subgame consistent solution, we have to invoke the results concerning the noncooperative and cooperative payoffs provided in Sect. 3. Invoking Proposition 1, the noncooperative payoff of player i in the game (14.46) and (14.47) is a linear value function:

$$V^i(t, x) = e^{-rt}[A_i(t)x + B_i(t)], \tag{14.48}$$

for $i \in \{1, 2\}$ and $t \in [\tau, T]$.

Invoking Proposition 3 and Proposition 4, the payoff of player i under cooperation with payoff weight α_1 is given by the value function:

$$W^{(\alpha_1)i}(t,x) = e^{-rt}[A_i^{\alpha_1}(t)x + B_i^{\alpha_1}(t)], \tag{14.49}$$

for $i \in \{1,2\}$ and $t \in [\tau, T]$.

To obtain a subgame consistent solutions to the cooperative game (14.46) and (14.47), we first note that group optimality will be maintained only if the optimality principle selects the same weight α_1 throughout the game interval $[t_0, T]$. For subgame consistency to hold, the chosen α_1 must also maintain individual rationality throughout the game interval. Therefore the payoff weight α_1 must satisfy:

$$\begin{aligned} W^{(\alpha_1)i}(t,x) &= e^{-rt}[A_2^{\alpha_1}(t)x + B_2^{\alpha_1}(t)] \\ &\geq V^i(t,x) = e^{-rt}[A_i(t)x + B_i(t)], \end{aligned} \tag{14.50}$$

for $i \in \{1,2\}$ and $t \in [t_0, T]$.

Definition 2. We define the set $S_\tau^T = \bigcap_{\tau \le t \le T} S_t$, for $\tau \in [t_0, T)$, where S_t represents the set of α_1 satisfying individual rationality at time $t \in [t_0, T)$ and S_τ^T represents the set of α_1 satisfying individual rationality throughout the interval $[\tau, T)$.

In general $S_\tau^T \neq S_t^T$ for $\tau, t \in [t_0, T)$. If $S_{t_0}^T$ is an empty set, there does not exist any weight α_1 such that condition (14.50) is satisfied. To find out typical configurations of the set S_t for $t \in [t_0, T)$ of the game $\Gamma_c(x_0, T - t_0)$, we perform extensive numerical simulations with a wide range of parameter specifications for $a, b, \sigma, h_1, h_2, k_1, k_2, c_1, c_2, q_1, q_2, T, r$, and x_0. We denote the locus of the values of $\underline{\alpha}_1^t$ along $t \in [t_0, T)$ as curve $\underline{\alpha}_1$ and the locus of the values $\bar{\alpha}_1^t$ as curve $\bar{\alpha}_1$. In particular, typical patterns include:

(i) The curves $\underline{\alpha}_1$ and $\bar{\alpha}_1$ are continuous and move in the same direction over the entire game duration: either both increase monotonically or both decrease monotonically (see Fig. 14.1).
(ii) The curve $\underline{\alpha}_1$ declines continuously, and the curve $\bar{\alpha}_1$ rises continuously (see Fig. 14.2).
(iii) The curves $\underline{\alpha}_1$ and $\bar{\alpha}_1$ are continuous. One of these curves would rise/fall to a peak/trough and then fall/rise (see Fig. 14.3).
(iv) The set $S_{t_0}^T$ can be nonempty or empty.

A subgame consistent solution will be reached when the players agree to adopt a weight $\alpha_1 \in S_{t_0}^T \neq \phi$ throughout the game interval $[t_0, T]$.

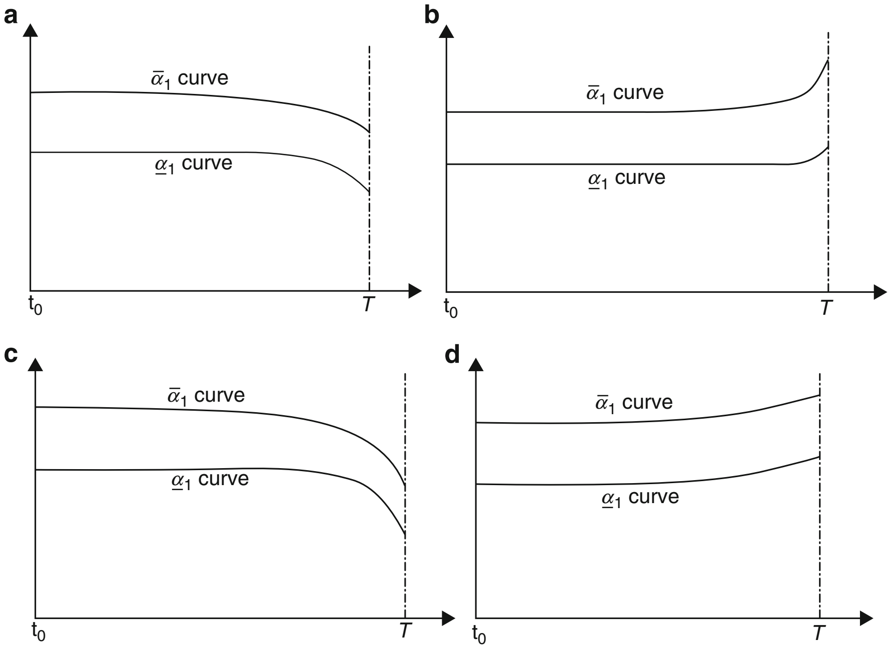
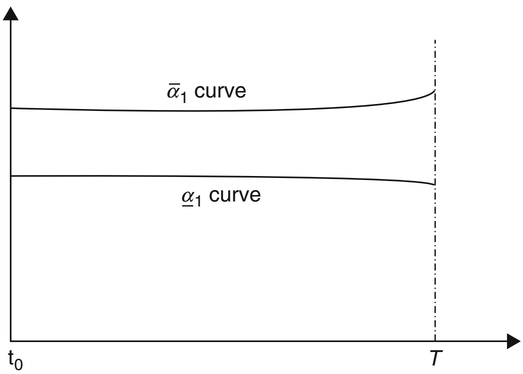


Fig. 14.1 (**a**–**d**) Loci of $\underline{\alpha}_1$ and $\bar{\alpha}_1$: Monotonic and same direction

Fig. 14.2 Loci of $\underline{\alpha}_1$ declines continuously and $\bar{\alpha}_1$ rises continuously



In the case when $S_{t_0}^{T}$ is empty for the game with horizon $[0, T]$, the players may have to shorten the duration of cooperation to a range $[0, \bar{T}_1]$ in which a nonempty $S_{t_0}^{\bar{T}_1}$ exists. To accomplish this, one has to identify the terminal conditions at time $\bar{T}_1$ and examine the subgame in the time interval $[\bar{T}_1, T]$ after the duration of cooperation.

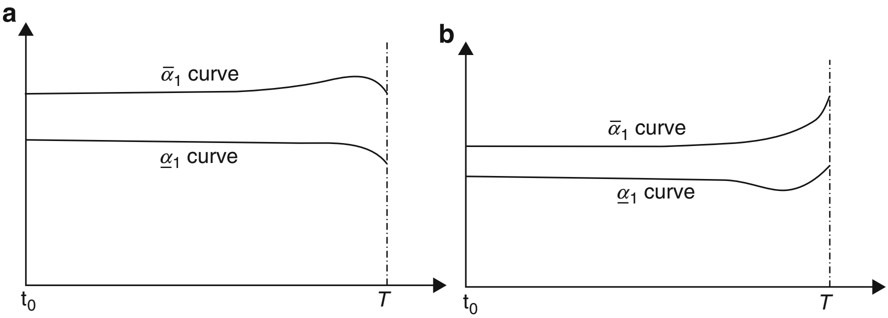


Fig. 14.3 (**a**, **b**) Loci of $\underline{\alpha}_1$ and $\bar{\alpha}_1$: rise/fall to a peak/trough and then fall/rise

7 Discrete-Time Analysis and Variable Weights

In this section, we study discrete-time NTU cooperative dynamic games.

7.1 NTU Cooperative Dynamic Games

Consider the general $T-$stage $n-$person nonzero-sum discrete-time dynamic game with initial state x_1^0. The state space of the game is $X \subset R^m$ and the state dynamics of the game is characterized by the difference equation:

$$x_{k+1} = f_k(x_k, u_k^1, u_k^2, \ldots, u_k^n), \tag{14.51}$$

for $k \in \{1, 2, \ldots, T\} \equiv \kappa$ and $x_1 = x_1^0$, where $u_k^i \in R^{m_i}$ is the control vector of player i at stage k and $x_k \in X$ is the state of the game. The payoff that player i seeks to maximize is:

$$\sum_{k=1}^{T} g_k^i(x_k, u_k^1, u_k^2, \ldots, u_k^n) + q^i(x_{T+1}), \tag{14.52}$$

where $i \in \{1, 2, \ldots, n\} \equiv N$ and $q^i(x_{T+1})$ is the terminal payoff that player i will receive at stage $T + 1$.

The payoffs of the players are not transferable. Let $\{\phi_k^i(x)$, for $k \in \kappa$ and $i \in N\}$ denote a set of strategies that provides a feedback Nash equilibrium solution (if it exists) to the game (14.51) and (14.52) and $\{V^i(k, x)$, for $k \in \kappa$ and $i \in N\}$ denote the value functions yielding the payoff to player i over the stages from k to T.

To improve their payoffs, the players agree to cooperate and act according to an agreed-upon cooperative scheme. Since payoffs are nontransferable, the payoffs of individual players are directly determined by the optimal cooperative strategies adopted. Consider the case in which the players agree to adopt a vector of payoff weights $\alpha = (\alpha^1, \alpha^2, \ldots, \alpha^n)$ in all stages, where $\sum_{j=1}^{n} \alpha^j = 1$. Conditional upon the vector of weights α, the agents' Pareto optimal cooperative strategies can be generated by solving the dynamic programming problem of maximizing the weighted sum of payoffs:

$$\sum_{j=1}^{n} \left[\sum_{k=1}^{T} \alpha^j g_k^j (x_k, u_k^1, u_k^2, \ldots, u_k^n) + \alpha^j q^j (x_{T+1}) \right] \tag{14.53}$$

subject to (14.51).

An optimal solution to the problem (14.51) and (14.53) can be characterized by the following Theorem.

Theorem 3. *A set of strategies* $\{\psi_k^{(\alpha)i}(x)$*, for* $k \in \kappa$ *and* $i \in N\}$ *provides an optimal solution to the problem* (14.51) *and* (14.53) *if there exist functions* $W^{(\alpha)}(k, x)$*, for* $k \in K$*, such that the following recursive relations are satisfied:*

$$W^{(\alpha)}(T+1, x) = \sum_{j=1}^{n} \alpha^j q^j (x_{T+1}), \tag{14.54}$$

$$\begin{aligned} W^{(\alpha)}(k, x) = \max_{u_k^1, u_k^2, \ldots, u_n^n} & \left\{ \sum_{j=1}^{n} \alpha^j g_k^j \left(x_k, u_1^k, u_2^k, \ldots, u_n^k \right) \right. \\ & \left. + W^{(\alpha)}[k+1, f_k(x_k, u_1^k, u_2^k, \ldots, u_n^k)] \right\} \\ = & \sum_{j=1}^{n} \alpha^j g_k^j [x, \psi_k^{(\alpha)1}(x), \psi_k^{(\alpha)2}(x), \ldots, \psi_k^{(\alpha)n}(x)] \\ & + W^{(\alpha)}[k+1, f_k(x, \psi_k^{(\alpha)1}(x), \psi_k^{(\alpha)2}(x), \ldots, \psi_k^{(\alpha)n}(x))]. \end{aligned} \tag{14.55}$$

Proof. The conditions in (14.54)–(14.55) follow directly from dynamic programming. ■

Substituting the optimal control $\{\psi_k^{(\alpha)i}(x)$, for $k \in \kappa$ and $i \in N\}$ into the state dynamics (14.51), one can obtain the dynamics of the cooperative trajectory as:

$$x_{k+1} = f_k(x_k, \psi_k^{(\alpha)1}(x_k), \psi_k^{(\alpha)2}(x_k), \ldots, \psi_k^{(\alpha)n}(x_k)), \tag{14.56}$$

for $k \in \kappa$ and $x_1 = x^0$.

We use $x_k^{(\alpha)} \in X_k^{(\alpha)}$ to denote the value of the state at stage k generated by (14.56). The term $W^{(\alpha)}(k, x)$ yields the weighted cooperative payoff over the stages from k to T.

Given that all players are adopting the cooperative strategies in Sect. 2.1, the payoff of player i under cooperation can be obtained as:

$$W^{(\alpha)i}(t,x) = \left\{ \sum_{k=t}^{T} g_k^i[x_k^{(\alpha)}, \psi_k^{(\alpha)1}(x_k^{(\alpha)}), \psi_k^{(\alpha)2}(x_k^{(\alpha)}), \ldots, \psi_k^{(\alpha)n}(x_k^{(\alpha)})] \right.$$
$$\left. + q^i(x_{T+1}^{(\alpha)}) | x_t^{(\alpha)} = x \right\}, \tag{14.57}$$

for $i \in N$ and $t \in \kappa$.

To allow the derivation of the functions $W^{(\alpha)i}(t,K)$ in a more direct way, we derive a deterministic counterpart of the Yeung (2013) analysis and characterize individual players' payoffs under cooperation as follows.

Theorem 4. *The payoff of player i at stage k can be characterized as the value function $W^{(\alpha)i}(k,x)$ satisfying the following recursive system of equations:*

$$W^{(\alpha)i}(T+1,x) = q^i(x_{T+1}),$$
$$W^{(\alpha)i}(k,x) = g_k^i[x_k^{(\alpha)}, \psi_k^{(\alpha)1}(x), \psi_k^{(\alpha)2}(x), \ldots, \psi_k^{(\alpha)n}(x)] \tag{14.58}$$
$$+ W^{(\alpha)i}[k+1, f_k(x, \psi_k^{(\alpha)1}(x), \psi_k^{(\alpha)2}(x), \ldots, \psi_k^{(\alpha)n}(x))],$$

for $i \in N$ and $k \in \kappa$.

Proof. See Yeung (2013). ■

For individual rationality to be maintained throughout all the stages $k \in \kappa$, it is required that:

$$W^{(\alpha)i}(k, x_k^{(\alpha)}) \geq V^i(k, x_k^{(\alpha)}), \tag{14.59}$$

for $i \in N$ and $k \in \kappa$.

Let the set of α weights that satisfies (14.59) be denoted by Λ. If Λ is not empty, a vector $\hat{\alpha} = (\hat{\alpha}^1, \hat{\alpha}^2, \ldots, \hat{\alpha}^n) \in \Lambda$ agreed upon by all players would yield a cooperative solution that satisfies both individual rationality and Pareto optimality throughout the duration of cooperation.

Remark 1. The pros of the constant payoff weights scheme are that full Pareto efficiency is satisfied in the sense that there does not exist any strategy path which would enhance the payoff of a player without lowering the payoff of at least one of the other players in all stages.

The cons of the constant payoff weights scheme include the inflexibility in accommodating the preferences of the players in a cooperative agreement and the high possibility of the nonexistence of a set of weights that satisfies individual rationality throughout the duration of cooperation .

7.2 Subgame Consistent Cooperation with Variable Weights

As shown in Sect. 6.1, constant weights schemes in NTU cooperative dynamic games often would not be able to guarantee individual rationality throughout the game horizon. To resolve this difficulty, time varying payoff weights can be adopted. Sorger (2006) presented a recursive Nash bargaining solution for a discrete-time NTU cooperative dynamic game by allowing the payoff weights to be re-negotiated. Yeung and Petrosyan (2015) considered the derivation of subgame consistent solutions for NTU cooperative dynamic games with variable payoff weights. A salient property of a subgame consistent solution is that the agreed-upon optimality principle remains in effect for the subgame starting at each stage of the original game, and hence the players do not have any incentive to deviate from the solution plan. Let $\Gamma(t, x_t)$ denote the cooperative game in which the objective of player i is:

$$\sum_{k=t}^{T} g_k^i \left(x_k, u_k^1, u_k^2, \ldots, u_k^n\right) + q^i(x_{T+1}), \text{for } i \in N, \tag{14.60}$$

and the state dynamics is:

$$x_{k+1} = f_k(x_k, u_k^1, u_k^2, \ldots, u_k^n), \tag{14.61}$$

for $k \in \{t, t+1, \ldots, T\}$ and the state at stage t is x_t.

Let the agreed-upon optimality principle be denoted by $P(t, x_t)$. The optimality principle $P(t, x_t)$ may include:

(i) The allotment of the players' payoffs satisfying the condition that the proportion of cooperative payoff to noncooperative payoff for each player is the same
(ii) The allotment of the players' payoffs according to the Nash bargaining solution
(iii) The allotment of the players' payoffs according to the Kalai-Smorodinsky bargaining solution

For subgame consistency to be maintained, the agreed-upon optimality principle $P(t, x_t)$ must be satisfied in the subgame $\Gamma(t, x_t)$ for $t \in \{1, 2, \ldots, T\}$. Hence, when the game proceeds to any stage t, the agreed-upon solution policy remains effective.

A time-invariant weight scheme is often hardly applicable for the derivation of a subgame consistent solution for various reasons. First, the set Λ, for which individual rationality holds throughout the duration of the game, is often empty. Second, the existing set of time-invariant weights which satisfy individual rationality may put some players in relatively favorable positions and some in relatively unfavorable positions. This may not be acceptable to the relatively disfavored players. Finally, the original agreed-upon optimality principle could not be maintained as the game proceeds under a time-invariant payoff weight cooperative scheme. In dynamic games with nontransferable payoffs, it is often very difficult (if not impossible) to reach a cooperative solution using constant payoff weights. In general, to derive a set of subgame consistent strategies under a cooperative solution with optimality principle $P(t, x_t)$, a variable payoff weight scheme has to be adopted. In particular, at each stage $t \in \kappa$ the players would adopt a vector of payoff weights $\hat{\alpha} = (\hat{\alpha}^1, \hat{\alpha}^2, \ldots, \hat{\alpha}^n)$ for $\sum_{j=1}^{n} \hat{\alpha}_t^j = 1$ which leads to satisfaction of the agreed-upon optimality principle. The chosen sets of weights $\hat{\alpha} = (\hat{\alpha}^1, \hat{\alpha}^2, \ldots, \hat{\alpha}^n)$ must lead to the satisfaction of the optimality principle $P(t, x_t)$ in the subgame $\Gamma(t, x_t)$ for $t \in \{1, 2, \ldots, T\}$.

7.2.1 Cooperative Strategies in the Terminal Stages

To derive the optimal cooperative strategies in a subgame consistent solution for cooperative dynamic games with nontransferable payoffs, we invoke the principle of optimality in dynamic programming and begin with the final stages of the cooperative game. Consider first the last stage, that is, stage T, with the state $x_T = x \in X$. The players will select a set of payoff weights $\alpha_T = (\alpha_T^1, \alpha_T^2, \ldots, \alpha_T^n)$ which leads to satisfaction of the optimality principle $P(T, x)$. Optimal cooperative strategies can be generated by solving the following dynamic programming problem of maximizing the weighted sum of their payoffs:

$$\sum_{j=1}^{n} \left[\alpha_T^j g_T^j \left(x_T, u_T^1, u_T^2, \ldots, u_T^n\right) + \alpha_T^j q^i(x_{T+1})\right] \tag{14.62}$$

subject to:

$$x_{T+1} = f_T = \left(x_T, u_T^1, u_T^2, \ldots, u_T^n\right), \qquad x_T = x. \tag{14.63}$$

Invoking Theorem 3, given the payoff weights being α_T, the optimal cooperative strategies $\{u_T^i = \psi_T^{(\alpha_T)i}$, for $i \in N\}$ in stage T are characterized by the conditions:

$$W^{(\alpha)}(T+1, x) = \sum_{j=1}^{n} \alpha_T^j q^j(x_{T+1}),$$

$$W^{(\alpha_T)}(T, x) = \max_{u_T^1, u_T^2, \ldots, u_T^n} \left\{ \sum_{j=1}^{n} \alpha_T^j g_T^j \left(x_T, u_T^1, u_T^2, \ldots, u_T^n\right) \right.$$
$$\left. + W^{(\alpha_T)}[T+1, f_T(x_T, u_T^1, u_T^2, \ldots, u_T^n)] \right\}. \tag{14.64}$$

Given that all players are adopting the cooperative strategies, the payoff of player i under cooperation covering stages T and $T+1$ can be obtained as:

$$W^{(\alpha_T)}(T, x) = c[x, \psi_k^{(\alpha_T)1}(x), \psi_k^{(\alpha_T)2}(x), \ldots, \psi_k^{(\alpha_T)n}(x)] + q^i(x_{T+1}^{(\alpha_T)}), \tag{14.65}$$

for $i \in N$.

Invoking Theorem 4, one can characterize $W^{(\alpha_T)i}(T,x)$ by the following equations:

$$W^{(\alpha_T)i}(T+1,x) = q^i(x),$$

$$\begin{aligned} W^{(\alpha_T)i}(T,x) = {} & g_T^i[x, \psi_T^{(\alpha_T)1}(x), \psi_T^{(\alpha_T)2}(x), \dots, \psi_T^{(\alpha_T)n}(x)] \\ & + W^{(\alpha_T)i}[T+1, f_T(x, \psi_T^{(\alpha_T)1}(x), \psi_T^{(\alpha_T)2}(x), \dots \\ & \dots, \psi_T^{(\alpha_T)n}(x))], \end{aligned} \tag{14.66}$$

for $i \in N$.

For individual rationality to be maintained, it is required that:

$$W^{(\alpha_T)i}(T,x) \geq V^i(T,x), \tag{14.67}$$

for $i \in N$.

We use Λ_T to denote the set of weights α_T that satisfies (14.67). Let $\hat{\alpha}_T = (\hat{\alpha}_T^1, \hat{\alpha}_T^2, \dots, \hat{\alpha}_T^n) \in \Lambda_T$ denote the payoff weights at stage T that leads to the satisfaction of the optimality principle $P(T,x)$.

Now we proceed to cooperative scheme in the second to last stage. Given that the payoff of player i at stage T is $W^{(\hat{\alpha}_T)i}(T,x)$, his payoff at stage $T-1$ can be expressed as:

$$\begin{aligned} & g_{T-1}^i(x_{T-1}, u_{T-1}^1, u_{T-1}^2, \dots, u_{T-1}^n) \\ & + g_T^i[x_T, \psi_T^{(\hat{\alpha}_T)1}(x_T), \psi_T^{(\hat{\alpha}_T)2}(x_T), \dots, \psi_T^{(\hat{\alpha}_T)n}(x_T)] + q^i(x_{T+1}) \\ = {} & g_{T-1}^i(x_{T-1}, u_{T-1}^1, u_{T-1}^2, \dots, u_{T-1}^n) + W^{(\hat{\alpha}_T)i}(T, x_T), \end{aligned} \tag{14.68}$$

for $i \in N$.

At this stage, the players will select payoff weights $\alpha_{T-1} = (\alpha_{T-1}^1, \alpha_{T-1}^2, \dots, \alpha_{T-1}^n)$ which lead to satisfaction of the optimality principle $P(T-1,x)$. The players' optimal cooperative strategies can be generated by solving the following dynamic programming problem of maximizing the weighted sum of payoffs:

$$\sum_{j=1}^{n} \alpha_{T-1}^j \left[g_{T-1}^j \left(x_{T-1}, u_{T-1}^1, u_{T-1}^2, \dots, u_{T-1}^n \right) + W^{(\hat{\alpha}_T)j}(T, x_T) \right] \tag{14.69}$$

subject to:

$$x_T = f_{T-1}(x_{T-1}, u_{T-1}^1, u_{T-1}^2, \dots, u_{T-1}^n), x_{T-1} = x. \tag{14.70}$$

Invoking Theorem 3, given the payoff weights being α_{T-1}, the optimal cooperative strategies $\{u_{T-1}^i = \psi_{T-1}^{(\alpha_{T-1})i}$, for $i \in N\}$ at stage $T-1$ are characterized by the conditions:

$$W^{(\alpha_{T-1})}(T,x) = \sum_{j=1}^{n} \alpha_{T-1}^{j} W^{(\hat{\alpha}_T)j}(T,x),$$

$$W^{(\alpha_{T-1})}(T-1,x) = \max_{u_{T-1}^1, u_{T-1}^2, \ldots, u_{T-1}^n} \left\{ \sum_{j=1}^{n} \alpha_{T-1}^{j} g_{T-1}^{j} \left(x_{T-1}, u_{T-1}^1, u_{T-1}^2, \ldots, u_{T-1}^n \right) + W^{(\alpha_{T-1})}[T, f_{T-1}(x_{T-1}, u_{T-1}^1, u_{T-1}^2, \ldots, u_{T-1}^n)] \right\}. \quad (14.71)$$

Invoking Theorem 4, one can characterize the payoff of player i under cooperation covering the stages $T-1$ to $T+1$ by:

$$W^{(\alpha_{T-1})i}(T,x) = W^{(\hat{\alpha}_T)i}(T,x_T),$$

$$\begin{aligned} W^{(\alpha_{T-1})i}(T-1,x) = {} & g_{T-1}^{j}[x, \psi_{T-1}^{(\alpha_{T-1})1}(x), \psi_{T-1}^{(\alpha_{T-1})2}(x), \ldots, \psi_{T-1}^{(\alpha_{T-1})n}(x)] \\ & + W^{(\alpha_{T-1})i}[T, f_{T-1}(x, \psi_{T-1}^{(\alpha_{T-1})1}(x), \psi_{T-1}^{(\alpha_{T-1})2}(x), \ldots \\ & \ldots, \psi_{T-1}^{(\alpha_{T-1})n}(x))], \end{aligned} \quad (14.72)$$

for $i \in N$.

For individual rationality to be maintained, it is required that:

$$W^{(\alpha_{T-1})i}(T-1,x) \geq V^{i}(T-1,x), \quad (14.73)$$

for $i \in N$.

We use Λ_{T-1} to denote the set of weights α_{T-1} that leads to satisfaction of (14.73). Let the vector $\hat{\alpha}_{T-1} = (\hat{\alpha}_{T-1}^1, \hat{\alpha}_{T-1}^2, \ldots, \hat{\alpha}_{T-1}^n) \in \Lambda_{T-1}$ be the set of payoff weights that leads to satisfaction of the optimality principle $\Gamma(T-1,x)$.

7.2.2 Optimal Cooperative Strategies in Preceding Stages

Now we proceed to characterize the cooperative scheme at stage $k \in \{1, 2, \ldots, T-1\}$. Following the analysis in Sect. 4.1, the payoff of player i at stage $k+1$ is $W^{(\hat{\alpha}_{k+1})i}(k+1,x)$, and his payoff at stage k can be expressed as:

$$\begin{aligned} & g_k^i \left(x_k, u_k^1, u_k^2, \ldots, u_k^n \right) \\ & \quad + \sum_{h=k}^{T} g_h^i[x_h, \psi_h^{(\hat{\alpha}_T)1}(x_h), \psi_h^{(\hat{\alpha}_T)2}(x_h), \ldots, \psi_h^{(\hat{\alpha}_T)n}(x_h)] + q^i(x_{T+1}) \\ & = g_k^i(x_k, u_k^1, u_k^2, \ldots, u_k^n) + W^{(\hat{\alpha}_{k+1})i}(k, x_{k+1}), \end{aligned} \quad (14.74)$$

for $i \in N$.

At this stage, the players will select a set of weights $\alpha_k = (\alpha_k^1, \alpha_k^2, \ldots, \alpha_k^n)$ which leads to satisfaction of the optimality principle $P(k,x)$. The players' optimal cooperative strategies can be generated by solving the following dynamic programming problem of maximizing the weighted sum of payoffs:

$$\sum_{j=1}^{n} \alpha_k^j \left[g_k^j(x_k, u_k^1, u_k^2, \ldots, u_k^n) + W^{(\hat{\alpha}_{k+1})j}(k+1, x_{k+1}) \right], \tag{14.75}$$

subject to:

$$x_{k+1} = f_k(x_k, u_k^1, u_k^2, \ldots, u_k^n), x_k = x. \tag{14.76}$$

Invoking Theorem 3, given the payoff weights being α_k, the optimal cooperative strategies $\{u_k^i = \psi_k^{(\alpha_k)i}$, for $i \in N\}$ at stage k are characterized by the conditions:

$$W^{(\alpha_k)}(k+1, x) = \sum_{j=1}^{n} \alpha_k^j W^{(\hat{\alpha}_{k+1})j}(k+1, x_{k+1}),$$

$$W^{(\alpha_k)}(k, x) = \max_{u_k^1, u_k^2, \ldots, u_k^n} \left\{ \sum_{j=1}^{n} \alpha_k^j g_k^j \left(x_k, u_k^1, u_k^2, \ldots, u_k^n \right) \right.$$
$$\left. + W^{(\alpha_k)}[k+1, f_k(x_k, u_k^1, u_k^2, \ldots, u_k^n)] \right\}. \tag{14.77}$$

The payoff of player i under cooperation can be obtained as:

$$W^{(\alpha_k)i}(k, x) = g_k^i[x, \psi_k^{(\alpha_k)1}(x), \psi_k^{(\alpha_k)2}(x), \ldots, \psi_k^{(\alpha_k)n}(x)]$$
$$+ W^{(\hat{\alpha}_{k+1})}(k+1, x_{k+1}), \tag{14.78}$$

for $i \in N$.

Invoking Theorem 4, one can characterize $W^{(\alpha_k)i}(k, x)$ by the following equations:

$$W^{(\alpha_k)i}(k+1, x) = W^{(\hat{\alpha}_{k+1})i}(k+1, x),$$

$$W^{(\alpha_k)i}(k, x) = g_k^i[x, \psi_k^{(\alpha_k)1}(x), \psi_k^{(\alpha_k)2}(x), \ldots, \psi_k^{(\alpha_k)n}(x)] \tag{14.79}$$
$$+ W^{(\alpha_k)i}[k+1, f_k(x, \psi_k^{(\alpha_k)1}(x), \psi_k^{(\alpha_k)2}(x), \ldots, \psi_k^{(\alpha_k)n}(x))],$$

for $i \in N$.

For individual rationality to be maintained at stage k, it is required that:

$$W^{(\alpha_k)i}(k, x) \geq V^i(k, x), \tag{14.80}$$

for $i \in N$.

We use Λ_k to denote the set of weights α_k that satisfies (14.80). Again, we use $\hat{\alpha}_k = (\hat{\alpha}_k^1, \hat{\alpha}_k^2, \ldots, \hat{\alpha}_k^n) \in \Lambda_k$ to denote the set of payoff weights that leads to the satisfaction of the optimality principle $P(k, x)$, for $k \in \kappa$.

7.3 Subgame Consistent Solution: A Theorem

We provide below a theorem characterizing a subgame consistent solution of the cooperative dynamic game (14.51) and (14.52) with the optimality principle $P(k, x_k)$.

Theorem 5. *A set of payoff weights $\{\hat{\alpha}_k = (\hat{\alpha}_k^1, \hat{\alpha}_k^2, \ldots, \hat{\alpha}_k^n)$, for $k \in \kappa\}$ and a set of strategies $\{\psi_k^{(\hat{\alpha}_k)i}(x)$, for $k \in \kappa$ and $i \in N\}$ provide a subgame consistent solution to the cooperative dynamic game* (14.51) *and* (14.52) *with the optimality principle $P(k, x)$ if there exist functions $W^{(\hat{\alpha}_k)}(k, x)$ and $W^{(\hat{\alpha}_k)i}(k, x)$, for $k \in \kappa$ and $i \in N$, which satisfy the following recursive relations:*

$$W^{(\hat{\alpha}_{T+1})}(T+1, x) = q^i(x_{T+1}),$$

$$W^{(\hat{\alpha}_k)i}(k, x) = \max_{u_k^1, u_k^2, \ldots, u_k^n} \left\{ \sum_{j=1}^n \hat{\alpha}^j g_k^j \left(x_k, u_k^1, u_k^2, \ldots, u_k^n\right) + \sum_{j=1}^n \hat{\alpha}_k^j W^{(\hat{\alpha}_{k+1})j}[k+1, f_k(x, u_k^1, u_k^2, \ldots, u_k^n)] \right\};$$

$$\begin{aligned} W^{(\hat{\alpha}_k)i}(k, x) = {} & g_k^i[x, \psi_k^{(\hat{\alpha}_k)1}(x), \psi_k^{(\hat{\alpha}_k)2}(x), \ldots, \psi_k^{(\hat{\alpha}_k)n}(x)] \\ & + W^{(\hat{\alpha}_{k+1})i}[k+1, f_k(x, \psi_k^{(\hat{\alpha}_k)1}(x), \psi_k^{(\hat{\alpha}_k)2}(x), \ldots, \psi_k^{(\hat{\alpha}_k)n}(x))], \end{aligned} \tag{14.81}$$

for $i \in N$ and $k \in \kappa$; and the optimality principle

$$P(k, x) \text{ in all stages } k \in \kappa. \tag{14.82}$$

Proof. See the exposition from equation (14.62) to equation (14.80) in Sects. 6.2.1 and 6.2.2. ∎

Substituting the optimal control $\{\psi_k^{(\hat{\alpha}_k)i}(x)$, for $i \in N$ and $k \in \kappa\}$ into the state dynamics (14.51), one can obtain the dynamics of the cooperative trajectory as:

$$x_{k+1} = f_k(x_k, \psi_k^{(\hat{\alpha}_k)1}(x_k), \psi_k^{(\hat{\alpha}_k)2}(x_k), \ldots, \psi_k^{(\hat{\alpha}_k)n}(x_k)), \tag{14.83}$$

$x_1 = x_1^0$ and $k \in \kappa$.

The cooperative trajectory $\{x_k^*$ for $k \in \kappa\}$ is the solution generated by (14.83).

Remark 2. The subgame consistent solution presented in Theorem 5 is conditional Pareto efficient in the sense that the solution is a Pareto-efficient outcome satisfying the condition that the agreed-upon optimality principle is maintained at all stages.

The conditional Pareto efficiency of subgame consistent solution is not fully Pareto efficient in the sense that there may exist payoff patterns which are Pareto superior for some strategy paths not satisfying the agreed-upon optimality principle in all stages. However, there do not exist any strategy paths satisfying the agreed-upon optimality principle at every stage that would lead to the payoff for any player i $W^i(t,x) > W^{(\hat{\alpha}_t)i}(t,x)$, while the payoffs of other players remain no less than $W^{(\hat{\alpha}_t)j}(t,x)$, for $i, j \in N$ and $j \neq i$.

Remark 3. A subgame consistent solution is fully Pareto efficient only if the optimality principle $P(t,x)$ requires the choice of a set of time-invariant payoff weights.

In NTU dynamic games, it is often not possible to reach a cooperative solution satisfying full Pareto efficiency and individual rationality because of the absence of side payments. Since the issue of full Pareto efficiency is of less importance than that of reaching a cooperative solution, achieving the latter at the expense of the former is a practical way out.

8 A Public Good Provision Example

To illustrate the solution mechanism with explicit game structures, we provide the derivation of subgame consistent solutions of public goods provision in a two-player cooperative dynamic game with nontransferable payoffs.

8.1 Game Formulation and Noncooperative Outcome

Consider an economic region with two asymmetric agents from the game example in Yeung and Petrosyan (2015). These agents receive benefits from an existing public capital stock x_t at each stage $t \in \{1, 2, \ldots, 4\}$. The accumulation dynamics of the public capital stock is governed by the difference equation:

$$x_{k+1} = x_k + \sum_{j=1}^{2} u_k^j - \delta x_k, x_1 = x_1^0, \tag{14.84}$$

for $t \in \{1, 2, 3\}$, where u_k^j is the physical amount of investment in the public good and δ is the rate of depreciation.

The objective of agent $i \in \{1, 2\}$ is to maximize the payoff:

$$\sum_{k=1}^{3} [a_k^i x_k - c_k^i (u_k^i)^2](1 + r)^{-(k-1)} + (q^i x_4 + m^i)(1 + r)^{-3}, \tag{14.85}$$

subject to the dynamics (14.84), where $a_k^i x_k$ gives the gain that agent i derives from the public capital at stage $t \in \{1, 2, 3\}$, $c_k^i (u_k^i)^2$ is the cost of investing u_k^i in the public capital, r is the discount rate, and $(q^i x_4 + m^i)$ is the terminal payoff of agent i at stage 4.

The payoffs of the agents are not transferable. We first derive the noncooperative outcome of the game. Invoking the standard analysis in dynamic games (see Başar and Olsder 1999 and Yeung and Petrosyan 2012), one can characterize the noncooperative Nash equilibrium for the game (14.84) and (14.85) as follows. A set of strategies $\{u_t^{i*} = \phi_t^i(x)$, for $t \in \{1, 2, 3\}$ and $i \in \{1, 2\}\}$ provides a Nash equilibrium solution to the game (14.84) and (14.85) if there exist functions $V^i(t, x)$, for $i \in \{1, 2\}$ and $t \in \{1, 2, 3\}$, such that the following recursive relations are satisfied:

$$V^i(t, x) = \max_{u_t^i} \Big\{[a_t^i x - c_t^i (u_t^i)^2](1 + r)^{-(t-1)} + V^i[t + 1, x + \phi_t^j(x) + u_t^i - \delta x]\Big\}, \tag{14.86}$$

for $t \in \{1, 2, 3\}$;

$$V^i(4, x) = (q^i x + m^i)(1 + r)^{-3}, \tag{14.87}$$

for $i \in \{1, 2\}$.

Performing the indicated maximization in (14.86) yields:

$$\phi_t^i(x) = \frac{(1 + r)^{t-1}}{2c_t^i} V_{x_{t+1}}^i \left[t + 1, x + \sum_{j=1}^{2} \phi_t^j(x) - \delta x \right], \tag{14.88}$$

for $i \in \{1, 2\}$ and $t \in \{1, 2, 3\}$.

Proposition 5. *The value function which represents the payoff of agent i can be obtained as:*

$$V^i(t, x) = [A_t^i x + C_t^i](1 + r)^{-(t-1)}, \tag{14.89}$$

for $i \in \{1, 2\}$ and $t \in \{1, 2, 3\}$, where

$$A_3^i = a_3^i + q^i (1 - \delta)(1 + r)^{-1},$$

$$C_3^i = -\frac{(q^i)^2(1+r)^{-2}}{4c_3^i} + \left[q^i \sum_{j=1}^{2} \frac{q^j(1+r)^{-1}}{2c_3^j} + m^i \right] (1+r)^{-1};$$

$$A_2^i = a_2^i + A_3^i(1-\delta)(1+r)^{-1},$$

$$C_2^i = -\frac{1}{4c_2^i}\left(A_3^i(1+r)^{-1}\right)^2 + \left[A_3^{(\sigma_3)i} \left(\sum_{j=1}^{2} \frac{A_3^j(1+r)^{-1}}{2c_2^j} + C_3^i \right) \right] (1+r)^{-1};$$

$$A_1^i = a_1^i + A_2^{(\sigma_2)i}(1-\delta)(1+r)^{-1},$$

and

$$C_1^i = -\frac{1}{4c_1^i}\left(A_2^i(1+r)^{-1}\right)^2 + \left[A_2^i \left(\sum_{j=1}^{2} \frac{A_2^j(1+r)^{-1}}{2c_2^j} + C_2^i \right) \right] (1+r)^{-1}; \tag{14.90}$$

for $i \in \{1,2\}$.

Proof. Using (14.88) and (14.89) to evaluate the system (14.86) and (14.87) yields the results in (14.89) and (14.90). ■

8.2 Cooperative Solution

Now consider first the case when the agents agree to cooperate and maintain an optimality principle $P(t, x_t)$ requiring the adoption of the mid values of the maximum and minimum of the payoff weight α_t^i in the set Λ_t, for $i \in \{1,2\}$ and $t \in \{1,2,3\}$.

In view of Theorem 5, to obtain the maximum and minimum values of α_T^i, we first consider deriving the optimal cooperative strategies at stage $T = 3$ by solving the problem:

$$W^{(\alpha_4)i}(4, x) = (q^i x + m^i)(1+r)^{-3},$$

for $i \in \{1,2\}$,

$$W^{(\alpha_3)}(3, x) = \max_{u_3^1, u_3^2} \left\{ \sum_{j=1}^{2} \alpha_3^j [\alpha_3^j x - c_3^j (u_3^j)^2](1+r)^{-2} \right.$$
$$\left. + \sum_{j=1}^{2} \alpha_3^j [q^j (x + \sum_{j=1}^{2} u_3^j - \delta x) + m^j](1+r)^{-3} \right\}. \tag{14.91}$$

Performing the indicated maximization in (14.91) yields:

$$\psi_3^{(\alpha_3)}(x) = \frac{(1+r)^{-1}}{2\alpha_3^i c_3^i} \sum_{j=1}^{n} \alpha_3^j q^j, \tag{14.92}$$

for $i \in \{1, 2\}$.

Proposition 6. *The value function can be obtained as:*

$$W^{(\alpha_3)}(3, x) = [A_3^{(\alpha_3)} x + C_3^{(\alpha_3)}](1+r)^{-2}, \tag{14.93}$$

where

$$A_3^{(\alpha_3)} = \sum_{j=1}^{2} \alpha_3^j [\alpha_3^j + q^j (1-\delta)(1+r)^{-1}],$$

and

$$C_3^{(\alpha_3)} = -\sum_{j=1}^{2} \alpha_3^j \left[\frac{(1+r)^{-2}}{4\alpha_3^j c_3^j} \left(\sum_{l=1}^{2} \alpha_3^l q_3^l \right)^2 \right] + \sum_{j=1}^{2} \alpha_3^j \left[q^j \left(\sum_{j=1}^{2} \frac{(1+r)^{-1}}{2\alpha_3^j c_3^j} \left(\sum_{l=1}^{2} \alpha_3^l q_3^l \right) \right) + m^j \right] (1+r)^{-1}. \tag{14.94}$$

Proof. Substituting the cooperative strategies from (14.92) into (14.91) yields the function $W^{(\alpha_3)}(3, x)$ in (14.93). ■

The payoff of player i under cooperation can be characterized as:

$$W^{(\alpha_3)i}(3, x) = \left[\alpha_3^i x - \frac{(1+r)^{-2}}{4\alpha_3^i c_3^i} \left(\sum_{l=1}^{2} \alpha_3^l q_3^l \right)^2 \right] (1+r)^{-2} + \left[q^i \left(x + \sum_{j=1}^{2} \frac{(1+r)^{-1}}{2\alpha_3^j c_3^j} \left(\sum_{l=1}^{2} \alpha_3^l q_3^l \right) - \delta x \right) + m^i \right] (1+r)^{-3}, \tag{14.95}$$

for $i \in \{1, 2\}$.

Invoking Theorem 4, the payoff functions of the players in the subgame starting at stage 3 can be obtained as follows.

Proposition 7. *The value function $W^{(\alpha_3)i}(3, x)$ in (14.95) can be obtained as:*

$$W^{(\alpha_3)i}(3, x) = [A_3^{(\alpha_3)i} x + C_3^{(\alpha_3)i}](1+r)^{-2}, \tag{14.96}$$

for $i \in \{1, 2\}$, *where*

$$A_3^{(\alpha_3)i} = [\alpha_3^j + q^j(1-\delta)(1+r)^{-1}],$$

and

$$C_3^{(\alpha_3)i} = -\left[\frac{(1+r)^{-2}}{4\alpha_3^j c_3^j}\left(\sum_{l=1}^{2}\alpha_3^l q_3^l\right)^2\right] + \left[q^j\left(\sum_{j=1}^{2}\frac{(1+r)^{-1}}{2\alpha_3^j c_3^j}\left(\sum_{l=1}^{2}\alpha_3^l q_3^l\right)\right) + m^j\right](1+r)^{-1}. \quad (14.97)$$

Proof. The right-hand side of the equation (14.95) is a linear function with coefficients $A_3^{(\alpha_3)i}$ and $C_3^{(\alpha_3)i}$ in (14.97). Hence Proposition 7 follows. ■

To identify the range of α_3 that satisfies individual rationality, we examine the functions which give the excess of agent i's cooperative over his noncooperative payoff:

$$W^{(\alpha_3)i}(3,x) - V^i(3,x) = [C_3^{(\alpha_3)i} - C_3^i](1+r)^{-2}, \quad (14.98)$$

for $i \in \{1, 2\}$.

For individual rationality to be satisfied, it is required that $W^{(\alpha_3)i}(3,x) - V^i(3,x) \geq 0$ for $i \in \{1, 2\}$. Using $\alpha_3^j = 1 - \alpha_3^i$ and upon rearranging terms, $C_3^{(\alpha_3)i}$ can be expressed as:

$$C_3^{(\alpha_3)i} = q^i\left[\frac{(1+r)^{-2}}{2c_3^i}\left(\frac{\alpha_3^i q^i + (1-\alpha_3^i)q^j}{\alpha_3^1}\right) + \frac{(1+r)^{-2}}{2c_3^j}\left(\frac{\alpha_3^i q^i + (1-\alpha_3^i)q^j}{1-\alpha_3^i}\right)\right] + m^i(1+r)^{-1} - \frac{(1+r)^{-2}}{4c_3^i}\left(\frac{\alpha_3^i q^i + (1-\alpha_3^i)q^j}{\alpha_3^i}\right)^2, \quad (14.99)$$

for $i, j \in \{1, 2\}$ and $i \neq j$.

Differentiating $C_3^{(\alpha_3)i}$ with respect to α_3^i yields:

$$\frac{\partial C_3^{(\alpha_3)i}}{\partial \alpha_3^i} = \frac{(1+r)^{-2}}{2c_3^j}\left(\frac{(q^i)^2}{(1-\alpha_3^i)^2}\right) + \frac{(1+r)^{-2}}{2c_3^i}\left(\frac{(1-\alpha_3^i)q^j}{\alpha_3^i}\right)\left(\frac{(q^i)^2}{(\alpha_3^i)^2}\right), \quad (14.100)$$

which is positive for $\alpha_3^i \in (0, 1)$.

One can readily observe that $\lim_{\alpha_3^i->0} C_3^{(\alpha_3)i} -> -\infty$ and $\lim_{\alpha_3^i->1} C_3^{(\alpha_3)i} -> \infty$. Therefore an $\underline{\alpha}_3^i \in (0,1)$ can be obtained such that $W^{(\underline{\alpha}_3^i, 1-\underline{\alpha}_3^i)i}(3,x) = V^i(3,x)$ and yields agent i's minimum payoff weight value satisfying his own individual rationality. Similarly there exists an $\bar{\alpha}_3^i \in (0,1)$ such that $W^{(\bar{\alpha}_3^i, 1-\bar{\alpha}_3^i)j}(3,x) = V^j(3,x)$ and yields agent i's maximum payoff weight value while maintaining agent j's individual rationality.

Since the maximization of the sum of weighted payoffs at stage 3 yields a Pareto optimum, there exist a nonempty set of α_3 satisfying individual rationality for both agents. Given that the agreed-upon optimality principle $P(t, x_t)$ requires the adoption of the mid values of the maximum and minimum of the payoff weight α_t^i in the set Λ_t, for $t \in \{1,2,3\}$, the cooperative weights in stage 3 is $\hat{\alpha}_3 = \left(\frac{\underline{\alpha}_3^i + \bar{\alpha}_3^i}{2}, 1 - \frac{\underline{\alpha}_3^i + \bar{\alpha}_3^i}{2}\right)$.

Now consider stage 2 problem. We derive the optimal cooperative strategies at stage 2 by solving the problem:

$$W^{(\alpha_2)}(2,x) = \max_{u_2^1, u_2^2} \left\{ \sum_{j=1}^{2} \alpha_2^j [\alpha_2^j x - c_2^j (u_2^j)^2](1+r)^{-1} \right.$$
$$\left. + \sum_{j=1}^{2} \alpha_2^j W^{(\hat{\alpha}_3)j}[3, x + \sum_{j=1}^{2} u_2^j - \delta x] \right\}. \quad (14.101)$$

Performing the indicated maximization in (101) yields:

$$\psi_2^{(\alpha_2)i}(x) = \frac{(1+r)^{-1}}{2\alpha_2^i c_2^i} \sum_{j=1}^{n} \alpha_2^j A_3^{(\hat{\alpha}_3)i}, \text{ for } i \in \{1,2\}.$$

Following the analysis at stage 3, one can obtain:

$$W^{(\alpha_2)}(2,x) = [A_2^{(\alpha_2)} x + C_2^{(\alpha_2)}](1+r)^{-1},$$

$$W^{(\alpha_2)i}(2,x) = [A_2^{(\alpha_2)i} x + C_2^{(\alpha_2)i}](1+r)^{-1}, \quad (14.102)$$

for $i \in \{1,2\}$, where $A_2^{(\alpha_2)}$, $C_2^{(\alpha_2)}$, $A_2^{(\alpha_2)i}$, and $C_2^{(\alpha_2)i}$ are functions that depend on α_2.

Similarly, one can readily show that $\frac{\partial C_2^{(\alpha_2)i}}{\partial \alpha_2^i}$ is positive and $\lim_{\alpha_2^i->0} C_2^{(\alpha_2)i} -> -\infty$ and $\lim_{\alpha_2^i->1} C_2^{(\alpha_2)i} -> \infty$. Agent i's minimum payoff weight is $\underline{\alpha}_2^i \in (0,1)$ which leads to $W^{(\underline{\alpha}_2^i, 1-\underline{\alpha}_2^i)i}(2,x) = V^i(2,x)$, and his maximum payoff weight is $\bar{\alpha}_2^i \in (0,1)$, which leads to $W^{(\bar{\alpha}_2^i, 1-\bar{\alpha}_2^i)j}(2,x) = V^j(2,x)$.

Invoking the agreed-upon optimality principle $P(t, x_t)$, the cooperative weights at stage 2 are $\hat{\alpha}_2 = \left(\frac{\underline{\alpha}_2^i + \bar{\alpha}_2^i}{2}, 1 - \frac{\underline{\alpha}_2^i + \bar{\alpha}_2^i}{2}\right)$.

Finally, following the analysis at stages 2 and 3, one can obtain the cooperative weights at stage 1 as $\hat{\alpha}_1 = \left(\frac{\underline{\alpha}_1^i + \bar{\alpha}_1^i}{2}, 1 - \frac{\underline{\alpha}_1^i + \bar{\alpha}_1^i}{2}\right)$.

The use of variable weights allows the derivation of dynamically stable cooperative solutions for a wide range of optimality principles in NTU cooperative dynamic games. It resolves the problem of lack of guarantee for the agreed-upon optimality principle to be maintained throughout the planning duration.

9 Concluding Remarks

The chapter has provided a coherent account of continuous-time and discrete-time NTU dynamic games. The topics covered include the basic formulation of NTU cooperative differential games, Pareto optimality, optimal cooperative trajectory, individual players' payoffs under cooperation, individual rationality, monitoring and threat strategies, subgame consistency, discrete-time NTU cooperative dynamic games, and subgame consistent solution via variable payoff weights. Potential applications of the NTU cooperative game analysis in existing NTU dynamic game studies are prevalent. The range of coverage of these potential applications is rather wide which include Wirl (1994), Rubio and Escriche (2001), and Dockner and Long (1993) in environmental dynamic games; Clark and Munro (1975), Levhari and Mirman (1980), and Clemhout and Wan (1985) in resource games; Karp (1984) in Trade and tariff games; Wirl (1996), Itaya and Shimomura (2001), and Kessing (2007) in public economics; and Cohen and Michel (1988) and Sorger (2002) in macroeconomics. As interactions among people and globalization increased rapidly in the past few decades, it has become more common to realize that dynamic cooperation is needed to improve the outcomes of human interactions like management of global warming, worldwide financial reform, and epidemics control. The complexity of the problems often involves certain degrees of non-transferability of payoffs. NTU cooperative dynamics games would likely emerge as an important paradigm in interactive decision-making. Finally, there is still a number of challenging issues to be tackled, like the optimal variable weight schemes in differential games, irrational behavior proof conditions, and credible threats.

Acknowledgements Financial support by the Russian Science Foundation (grant 17-11-01079) is gratefully acknowledged.

References

Başar T, Olsder GJ (1999) Dynamic noncooperative game theory. SIAM, Philadelphia

Clark CW, Munro GR (1975) The economics of fishing and modern capital theory. J Environ Econ Manag 2:92–106

Clemhout S, Wan HY (1985) Dynamic common-property resources and environmental problems. J Optim Theory Appl 46:471–481

Cohen D, Michel P (1988) How should control theory be used to calculate a time-consistent government policy. Rev Econ Stud 55:263–274 de-Paz A, Marin-Solano J, Navas J (2013) Time-consistent equilibria in common access resource games with asymmetric players under partial cooperation. Environ Model Assess 18(2): 171–184

Dockner E, Jorgensen S (1984) Cooperative and non-cooperative differential games solutions to an investment and pricing problem. J Int Oper Res Soc 35:731–739

Dockner E, Long NV (1993) International pollution control: cooperative versus non-cooperative strategies. J Environ Econ Manag 25:13–29

Edgeworth FY (1881) Mathematical psychics: an essay on the mathematics to the moral sciences. London, Kegan Paul. Reprinted in Diamond, M.A.

H*ä*m*ä*l*ä*inen R, Haurie A, Kaitala V (1985) Equilibria and threats in a fishery management game. Opt Control Appl Methods 6:315–333

Haurie A (1976) A note on nonzero-sum differential games with bargaining solutions. J Optim Theory Appl 18:31–39

Itaya J, Shimomura K (2001) A dynamic conjectural variation model in the private provision of public goods: a differential game approach. J Public Econ 81:153–172

Kalai E, Smorodinsky M (1975) Other solutions to Nash's bargaining problem. Econometrica 43(3):513–518

Kalai E (1977) Proportional solutions to bargaining situations: interpersonal utility comparison. Econometrica 45:1623–1630

Karp LS (1984) Optimality and consistency in a differential game with non-renewable resources. J Econ Dyn Control 8:73–97

Kessing S (2007) Strategic complementarity in the dynamic private provision of a discrete public good. J Public Econ Theory 9:699–710

Leitmann G (1974) Cooperative and non-cooperative many players differential games. Springer, New York

Levhari D, Mirman L (1980) The great fish war: an example of using a dynamic Nash-Cournot solution. Bell J Econ 11:322–334

Marin-Solano J (2014) Time-consistent equilibria in a differential game model with time inconsistent preferences and partial cooperation. In: Haunschmied J, Veliov V, Wrzaczek S (eds) Dynamic games in economics. Springer, New York. Forthcoming

Nash J (1950) The bargaining Problem. Econometrica 18:155–162

Nash J (1953) Two-person cooperative games. Econometrica 21:128–140

Petrosyan LA (1997) Agreeable solutions in differential games. Int J Math Game Theory Algebra 7:165–177

Petrosyan LA, Zenkevich NA (1996) Game theory. World Scientific, Singapore

Rubio S, Escriche L (2001) Strategic pigouvian taxation, stock externalities and polluting non-renewable resources. J Public Econ 79:297–313

Sorger G (2002) On the long-run distribution of capital in the Ramsey model. J Econ Theory 105:226–243

Sorger G (2006) Recursive bargaining over a productive asset. J Econ Dyn Control 30:2637–2659

Wirl F (1994) Pigouvian taxation of energy for flow and stock externalities and strategic non-competitive energy pricing. J Environ Econ Manag 26:1–18

Wirl F (1996) Dynamic voluntary provision of public goods: extension to non-linear strategies. Eur J Polit Econ 12:555–566

Yeung DWK (2004) Nontransferable individual payoff functions under stochastic dynamic cooperation. Int Game Theory Rev 6:281–289

Yeung DWK (2013) Nontransferable individual payoffs in cooperative stochastic dynamic games. Int J Algebra 7(13):597–606

Yeung DWK, Petrosyan LA (2004) Subgame consistent cooperative solutions in stochastic differential games. J Optim Theory Appl 120(3):651–666

Yeung DWK, Petrosyan LA (2005) Subgame consistent solution of a cooperative stochastic differential game with nontransferable payoffs. J Optim Theory Appl 124(3):701–724

Yeung DWK, Petrosyan LA (2010) Subgame consistent solutions for cooperative stochastic dynamic games. J Optim Theory Appl 145(3):579–596

Yeung DWK, Petrosyan LA (2012) Subgame consistent economic optimization: an advanced cooperative dynamic game analysis. Birkhauser, Boston. ISBN:978-0-8176-8261-3, 395pp

Yeung DWK, Petrosyan LA (2015) Subgame consistent cooperative solution for NTU dynamic games via variable weights. Automatica 50:84–89

Yeung DWK, Petrosyan LA (2016) Subgame consistent cooperation – a comprehensive treatise. Springer, Singapore

Yeung DWK, Petrosyan LA, Yeung PM (2007) Subgame consistent solutions for a class of cooperative stochastic differential games with nontransferable payoffs. Ann Int Soci Dyn Games 9:153–170

Part II
Applications of Dynamic Games

Resource Economics

15

Ngo Van Long

Contents

Abstract

This chapter provides a selective survey of dynamic game models of exploitation of natural resources. It covers both renewable resources and exhaustible resources. In relation to earlier surveys (Long, A survey of dynamic games in economics, World Scientific, Singapore, 2010; Long, Dyn Games Appl 1(1):115–148, 2011), the present work includes many references to new developments that appeared after January 2011 and additional suggestions for future research. Moreover, there is a greater emphasis on intuitive explanation.

Keywords

Exhaustible resources · Renewable resources · Overexploitation · Market structure · Dynamic games

N. V. Long (✉)
McGill University, Montreal, QC, Canada
e-mail: ngo.long@mcgill.ca

T. Başar, G. Zaccour (eds.), *Handbook of Dynamic Game Theory*,
https://doi.org/10.1007/978-3-319-44374-4_15

1 Introduction

Natural resources play an important role in economic activities. Many resources are essential inputs in production. Moreover, according to the World Trade Report of the WTO (2010, p. 40), "natural resources represent a significant and growing share of world trade and amounted to some 24 per cent of total merchandise trade in 2008." The importance of natural resources was acknowledged by classical economists. Smith (1776) points out that the desire to possess more natural resources was one of the motives behind the European conquest of the New World and the establishment of colonies around the globe. Throughout human history, many conflicts between nations or between social classes within a nation (e.g., the "elite" versus the "citizens") are attributable to attempts of possession or expropriation of natural resources (Long 1975; van der Ploeg 2010). Many renewable resources are at risk because of overexploitation. For example, in the case of fishery, according to the Food and Agriculture Organization, in 2007, 80 % of stocks are fished at or beyond their maximum sustainable yield (FAO 2009). Recent empirical work by McWhinnie (2009) found that shared fish stocks are indeed more prone to overexploitation, confirming the theoretical prediction that an increase in the number of agents that exploit a resource will reduce the equilibrium stock level.

Some natural resources, such as gold, silver, oil, and natural gas, are nonrenewable. They are sometimes called "exhaustible resources." Other resources, such as fish and forest, are renewable. Water is renewable in regions with adequate rainfall, but certain aquifers can be considered as nonrenewable, because the rate of recharge is very slow.[1] Conflicts often arise because of lack of well-defined property rights in the extraction of resources. In fact, the word "rivals" was derived from the Latin word "rivales" which designated people who drew water from the same stream (rivus).[2] Couttenier and Soubeyran (2014, 2015) found that natural resources are often causes of civil conflicts and documented the empirical relationship between water shortage on civil wars in sub-Saharan Africa.

Economists emphasize an essential feature of natural resource exploitation: the rates of change in their stocks are influenced by human action. In situations where the number of key players is not too large, the appropriate way of analyzing the rivalrous exploitation of natural resources is to formulate a dynamic game. This chapter provides a selective survey of dynamic game models of exploitation of natural resources. In relation to earlier surveys (Long 2010, 2011), the present work

[1]For a review of the game theoretic approach to water resources, see Dinar and Hogarth (2015). For some recent models of differential games involving water transfer between two countries, see Cabo et al. (2014) and in particular Cabo and Tidball (2016) where countries cooperate in the infrastructure investment stage but play a noncooperative game of water transfer in a second stage. Cabo and Tidball (2016) design a time-consistent imputation distribution procedure to ensure cooperation, along the lines of Jørgensen and Zaccour (2001).

[2]Dictionnaire LE ROBERT, Société du Nouveau Littré, Paris: 1979.

includes many references to new developments that appeared since 2011. Moreover, this chapter places a greater emphasis on intuitive explanation.

The next section reviews critical issues and dynamic game models in the exploitation of renewable resources. Section 3 is devoted to exhaustible resources. The final section offers some thoughts on future directions of research.

2 Renewable Resources

Renewable resources are natural resources for which a positive steady-state stock level can be maintained while exploitation can remain at a positive level for ever. Some examples of renewable resources are forests, aquifers, fish stocks, and other animal species. Without careful management, some renewable resources may become extinct. The problem of overexploitation of natural resources is known as the "Tragedy of the commons" (Hardin 1968).[3] There is a large literature on the dynamic tragedy of the commons. While some papers focus on the case where players use open-loop strategies (e.g., Clark and Munro 1975; Kemp and Long 1984; Kaitala et al. 1985; Long and McWhinnie 2012), most papers assume that players use feedback strategies (e.g., Dinar and Zaccour 2013; Long and Sorger 2006). In what follows, we review both approaches.[4]

2.1 The Tragedy of the Commons: Exploitation of Renewable Natural Assets

The standard fishery model is Clark and Munro (1975). There are m fishermen who have access to a common fish stock, denoted by $S(t) \geq 0$. The quantity of fish that fisherman i harvests at time t is $h_i(t) = \gamma L_i(t)S(t)$ where $L_i(t)$ is his effort and γ is called the catchability coefficient. In the absence of human exploitation, the natural rate of reproduction of the fish stock is $dS/dt = G(S(t))$, where it is assumed that $G(S)$ is a hump-shaped and strictly concave function, with $G(0) = 0$ and $G'(0) > 0$. Taking into account the harvests, the transition equation for the stock is

$$\frac{dS}{dt} = G(S(t)) - \sum_{i=1}^{m} \gamma L_i(t)S(t)$$

By assumption, there exists a unique stock level S_M such that $G'(S_M) = 0$. The stock level S_M is called the maximum sustainable yield stock level. The quantity

[3]However, as pointed out by Ostrom (1990), in some societies, thanks to good institutions, the commons are efficiently managed.

[4]See Fudenberg and Tirole (1991) on the comparison of the concepts of open-loop equilibrium and feedback or Markov-perfect equilibrium.

$G(S_M)$ is called the maximum sustainable yield.[5] A common functional form for $G(S)$ is $rS(1 - S/K)$ where r and K are positive parameters. The parameter r is called the intrinsic growth rate, and K is called the carrying capacity, because when the stock S is greater than the carrying capacity level K, the fish population declines.

2.1.1 Open-Loop Games of Fishery

Clark and Munro (1975) propose an open-loop differential game of exploiting a common access fish stock. For simplicity, assume that the market price of fish, P, is exogenous and constant over time. The total effort cost of fisherman i at time t is $cL_i(t)$, where c is a positive parameter, assumed to be small relative to P. Assume that L_i must belong to the interval $\left[0, \overline{L}\right]$ where $\overline{L}$ is player i's maximum possible effort level. Each fisherman i chooses a time path $L_i(t) \in \left[0, \overline{L}\right]$ to maximize the integral of his discounted stream of profit,

$$J_i = \int_0^{\infty} e^{-\rho t} \left[p\gamma L_i(t)S(t) - cL_i(t)\right] dt$$

where $\rho > 0$ is the discount rate, while taking into account the transition equation

$$\dot{S}(t) = G(S) - \gamma L_i(t)S(t) - \gamma \sum_{j \neq i} L_j(t)S(t)$$

If the fishermen were able and willing to cooperate, they would coordinate their efforts, and it is easy to show that this would result in a socially efficient steady-state stock, denoted by S_∞^E, which satisfies the following equation[6]:

$$\rho = G'(S_\infty^E) + \frac{G(S_\infty^E)}{S_\infty^E}\left[\frac{c}{P\gamma S_\infty^E - c}\right] \tag{15.1}$$

The intuition behind this equation is as follows. The left-hand side is the market rate of interest which producers use to discount the future profits. The right-hand side is the net rate of return of leaving a marginal fish in the pool instead of catching it. It is the sum of two terms: the first term, $G'(S_\infty^E)$, is the marginal natural growth rate of the stock (the biological rate of interest), and the second term is the gain that results from the reduction (brought about by a marginal increase in stock level) in the required aggregate effort to achieve the steady-state catch level. In an efficient solution, the two rates of return must be equalized, for otherwise further gains would be achievable by arbitrage.

[5]It has been estimated that about 80 % of fish stocks are exploited at or beyond their maximum sustainable yields. See FAO (2009).

[6]We assume that the upper bound constraint on effort is not binding at the steady state.

What happens if agents do not cooperate? Clark and Munro (1975) focus on the open-loop Nash equilibrium, i.e., each agent i determines her own time path of effort and takes the time path of efforts of other agents as given. Agent i believes that all agents $j \neq i$ are pre-committed to their time paths of effort $L_j(t)$, regardless of what may happen to the time path of the fish stock when i deviates from her plan. Assuming that $\overline{L}$ is sufficiently large, it can be shown that the open-loop Nash equilibrium results in a steady-state stock S_∞^{OL} that satisfies the equation

$$\rho = G'(S_\infty^{OL}) + \frac{1}{m}\frac{G(S_\infty^{OL})}{S_\infty^{OL}}\left[\frac{c}{P\gamma S_\infty^{OL} - c} - (m-1)\right] \tag{15.2}$$

The steady-state stock S_∞^{OL} is socially inefficient. It is equal to the socially efficient stock level S_∞^{E} only if $m = 1$. This inefficiency result is a consequence of a *dynamic overcrowding production externality*: when a fisherman catches more fish today, this will reduce level of tomorrow's stock of fish, which increases tomorrow's effort cost of all fishermen at any intended harvest level.[7]

A weakness of the concept of open-loop Nash equilibrium is that it assumes that players do not use any information acquired during the game and consequently do not respond to deviations that affect the anticipated path of the stock. Commenting on this property, Clemhout and Wan (1991) write: "for resource games at least, the open-loop solution is neither an equally acceptable alternative to the closed loop solution nor a safe approximation to it." For this reason, we now turn to models that focus on closed-loop (or feedback) solution.

2.1.2 Feedback Games of Exploitation of Renewable Natural Assets

The simplest fishery model where agents use feedback strategies is the Great Fish War model of Levhari and Mirman (1980). We present below a slightly modified version of that model. Thanks to the assumed special functional forms (logarithmic utility functions and a net reproduction function that is log-linear), it is possible to derive a closed-form solution of the equilibrium harvesting strategies. However, the essence of the results of Levhari and Mirman (1980) can be preserved under more general functional specifications (e.g., Dutta and Sundaram 1993b).

The model is formulated in discrete time. Consider the case of n identical countries that have common access to a fish stock S_t. Let h_{it} denote country i's harvest in period t. Define the total harvest in period t by $H_t = \sum_{i=1}^{n} h_{it}$. (We will show that $H_t \leq S_t$ in equilibrium.) Assume that the next period's fish stock is given by the difference equation $S_{t+1} = (S_t - H_t)^\kappa$ where $0 < \kappa < 1$. The parameter κ may be regarded as an index of the future availability of the resource. An increase in κ represents a low future availability.

[7] If the amount harvested depends only on the effort level and not on the level of the stock, i.e., $h_i = \gamma L_i$, then in an open-loop equilibrium, there is no dynamic overcrowding production externality. In that case, it is possible that open-loop exploitation is Pareto efficient; see Chiarella et al. (1984).

Harvesting is costless, and the utility of consuming h_{it} is $\sigma \ln h_{it}$, where $\sigma > 0$ is a parameter which is interpreted as the quality of the resource. For the moment, assume $\sigma = 1$. Let β denote the discount factor, where $0 < \beta < 1$. The payoff to country i is $\sum_{t=0}^{\infty} \beta^t \ln h_{it}$. It is simple to verify that if the countries cooperate, the optimal common feedback policy is $h_{it}^C(S_t) = (1 - \beta\kappa) S_t n^{-1}$. The resulting cooperative steady-state stock is $S_\infty = (\beta\kappa)^{\kappa/(1-\kappa)}$.

Turning to the noncooperative game, the Bellman equation for country i is

$$V_i(S) = \max_{h_i} \{\ln(h_i) + \beta V_i((S - H_{-i} - h_i)^\kappa)\}$$

where $H_{-i} = H - h_i$. Levhari and Mirman find that there exists a Markov-perfect Nash equilibrium in which countries use the linear feedback strategy

$$h_{it}^M(S_t) = \frac{1 - \beta\kappa}{n - \beta\kappa(n-1)} S$$

Thus, at each level of the stock, the noncooperative harvest rate exceeds the cooperative one. The resulting steady-state stock level is lower.

The Levhari-Mirman result of overexploitation confirms the general presumption that common access leads to inefficient outcome. The intuition is simple: if each player believes that the unit it chooses not to harvest today will be in part harvested by other players tomorrow, then no player will have a strong incentive to conserve the resource. This result was also found by Clemhout and Wan (1985), who used a continuous-time formulation. The Levhari-Mirman overexploitation result has been extended to the case of a coalitional fish war (Breton and Keoula 2012; Kwon 2006), using the same functional forms for the utility function and the net reproduction function. Kwon (2006) assumes that there are n ex ante identical countries, and m of them form a coalition, so that the number of players is $\eta = n - m + 1$. A coalition is called profitable if the payoff of a coalition member is greater than what it would obtain in the absence of the coalition. The coalitional Great Fish War game is the game involving a coalition and the $n - m$ outsiders. A coalition is said to be stable under Nash conjectures (as defined in d'Aspremont et al. 1983) if two "stability conditions" are satisfied. First is internal stability, which means that if a member drops out of the coalition, assuming that the other $m - 1$ members stay put, it will obtain a lower payoff. Second is external stability, which means that if an outsider joins the coalition, assuming that the existing m members will continue their membership, its payoff will be lower than its status quo payoff. Kwon (2006) shows that the only stable coalition (under the Nash conjectures) is of size two. Thus, when n is large, the overexploitation is not significantly mitigated when such a small stable coalition is formed. Breton and Keoula (2012) investigate a coalitional war model that departs from the Nash conjectures: they replace the Nash conjectures with what they call "rational conjectures," which relies on the farsightedness assumption. As they aptly put it, "the farsightedness assumption in a coalitional game acknowledges the fact that a deviation from a single player will lead to the formation of another coalition structure, as the result of possibly successive moves of her rivals in order to improve their payoff" (p. 298).[8] Breton and Keoula (2012) find that under plausible values of parameters, there is a wide scope for cooperation under the farsightedness assumption. For example, with $n = 20$, $\beta = 0.95$, and $\kappa = 0.82$, a coalition of size $m = 18$ is farsightedly stable (p. 305).

Fesselmeyer and Santugini (2013) extend the Levhari-Mirman fish war model to the case where there are exogenous environmental risks concerning quality and availability.[9] The risks are modeled as follows. Let x_t denote the "state of the environment" at date t. Assume that x_t can take on one of two values in the set $\{1, 2\}$. If $x_t = 1$, then the probability that $x_{t+1} = 2$ is ρ, where $0 < \rho < 1$, and the probability that $x_{t+1} = 1$ is $1 - \rho$. If $x_t = 2$, then $x_{t+1} = 2$ with probability 1. Assume that $x_0 = 1$. Then, since $\rho > 0$, there will be an environmental change at some time in the future. The authors find that if there is the risk that an environmental change (an increase in x_t from 1 to 2) will lead to lower renewability (i.e., $\kappa_2 \geq \kappa_1$), then rivalrous agents tend to reduce their exposure to this risk by harvesting less, as would the social planner; however, the risk worsens the tragedy of the commons in the sense that, at any given stock level, the ratio of Markov-perfect Nash equilibrium exploitation to the socially optimal harvest increases. In contrast, when the only risk is a possible deterioration in the quality of the fish (i.e., $\sigma_2 < \sigma_1$), this tends to mitigate the tragedy of the commons.

Other discrete-time models of dynamic games on renewable natural assets include Amir (1989) and Sundaram (1989), where some existence theorems are provided. Sundaram's model is a generalization of the model of Levhari and Mirman (1980): Sundaram replaces the utility function $\ln h_i$ with a more general strictly concave function $u(h_i)$ with $u'(0) = \infty$, and the transition function $S_{t+1} = (S_t - H_t)^\kappa$ is replaced by $S_{t+1} = f(S_t - H_t)$ where $f(0) = 0$ and $f(.)$ is continuous, strictly increasing, and crosses the 45 degree line exactly once. Assuming that all players are identical, Sundaram (1989) proves that there exists a Markov-perfect Nash equilibrium (MPNE) in which all players use the same strategy. Another result is that along any equilibrium path where players use stationary strategies, the time path of the stock is monotone. Sundaram (1989) also shows that in any symmetric Markov-perfect Nash equilibrium, the MPNE stationary strategy cannot be the same as the cooperative harvesting strategy $h^C(S)$.

Dutta and Sundaram (1993a) provide a further generalization of the model of Levhari and Mirman (1980). They allow the period payoff function to be dependent on the stock S in an additively separable way: $U_i(h_i, S) = u_i(h_i) + w(S)$ where $w(.)$ is continuous and increasing. For example, if the resource stock is a forest, consumers derive not only utility $u_i(h_i)$ from using the harvested timber but also

[8]The farsightedness concept was formalized in Greenberg (1990) and Chwe (1994) and has been applied to the literature on public goods (Ray and Vohra 2001) and international environmental agreements (de Zeeuw 2008; Diamantoudi and Sartzetakis 2015).

[9]For ownership risks, see Bohn and Deacon (2000).

pleasure $w(S)$ when they hike in a large forested area or when a larger forest ensure greater biodiversity than a smaller one. They show that a cooperative equilibrium exists. For a game with only three periods, they construct an example in which there is no Markov-perfect Nash equilibrium, if one player's utility is linear in consumption while his opponent has a strictly concave utility function. When the function $w(S)$ is *strictly convex*, they show by example that the dynamics of the stock along an equilibrium path can be very irregular. One may argue that in some contexts, the strict convexity of $w(S)$ (within a certain range) may be a plausible assumption. For example, within a certain range, as the size of a forest doubles, biodiversity may triple.[10] Finally, they consider the case of an infinite horizon game with a zero rate of discount. In this case, they assume that each player cares only about the long-run average (LRA) payoff, so that the utilities that accrue in the present (or over any finite time interval) do not count. For example, a "player" may be a government of a country in which the majority of voters adhere to Sidgwick's view that it is immoral to discount the welfare of the future generations (Sidgwick 1874). With zero discounting, the LRA criterion is consistent with the axiom that social choice should display "non-dictatorship of the present" (Chichilnisky 1996). Under the LRA criterion, Dutta and Sundaram (1993a) define the "tragedy" of the commons as the situation where the stock converges to a level lower than the golden rule stock level. They show that under the LRA criterion, there exists an MPNE that does not exhibit this "tragedy" property. This result is not very surprising, because, unlike the formulation of Clark and Munro (1975) where harvest depends on the product of effort and stock, $\gamma L_i(t)S(t)$, the model of Dutta and Sundaram assumes that the stock has no effect on the marginal product of labor, and thus, the only incentive to grab excessively comes from the wish to grab earlier than one's rivals, and this incentive may disappear under the LRA criterion, as the present consumption levels do not count.[11] However, whether a tragedy occurs or not, it can be shown that any MPNE is suboptimal from any initial state.[12] It follows that, in a broad sense, the tragedy of the commons is a very robust result.

While the model of Levhari and Mirman (1980) shows that the Markov-perfect equilibrium is generically not Pareto efficient, inefficiency need not hold at every initial stock level. In fact, Dockner and Sorger (1996) provide an example of a fishery model in which there is a continuum of Markov-perfect equilibria, and they show that in the limit, as the discount rate approaches zero, the MPNE stationary

[10] In contrast, in the standard models, where all utility functions are concave, it can be shown that the equilibrium trajectory of the state variable must eventually become monotone. See Dutta and Sundaram (1993b).

[11] Sundaram and Dutta (1993b) extend this "no tragedy" result to the case with mild discounting: as long as the discount rate is low enough, if players use *discontinuous* strategies that threaten to make a drastic increase in consumption when the stock falls below a certain level, they may be able to lock each other into a stock level that is dynamically inefficient and greater than the cooperative steady state.

[12] Except possibly the cooperative steady state (Dutta and Sundaram 1993b, Theorem 3).

steady-state stock converges to the steady-state stock of the (efficient) cooperative solution. This result is of course a local result. It does not imply that the MPNE harvesting rule coincides with the socially optimal one, for all stock levels. A special feature of Dockner and Sorger (1996) model is that they assume a square root utility function. The reproduction function for the stock x is $F(x)$, a strictly concave, hump-shaped function, with $F(0) = F(1) = 0$. There is a constant, exogenous upper bound $\overline{u}$ on harvest rates that is independent of the stock level. They show that the cooperative solution is unique and leads to the steady-state stock $x_1 \in (0, 1)$, where the effect of the stock on the rate of reproduction, $F'(x_1)$, is equal to the discount rate r, a familiar result (Long 1977). In contrast, when two players behave noncooperatively, Dockner and Sorger (1996) show that there is a continuum of symmetric equilibria, which differ from each other in terms of the interval of stock levels such that both players harvest at the maximum rate $\overline{u}$. For each of these equilibria, the harvesting rate when $u < \overline{u}$ is found to be an increasing function of the discount rate r. The intuition behind the multiplicity of equilibria is simple: if one player believes that the other exploits at the maximum rate over a given interval of stock, then she has no incentive to conserve the stock, and thus its best response is to do likewise. Therefore, there is a continuum of intervals of stock with maximum exploitation. The corresponding Markov-perfect exploitation strategy displays a jump discontinuity at the lower bound of each such interval.[13] An interesting property of the model is that as the rate of discount tends to zero, the steady state of the noncooperative common access game coincides with the steady state of the cooperative game.[14] This result is in sharp contrast with Levhari and Mirman (1980), where the tragedy of the commons does not vanish when the discount rate becomes arbitrarily small. This discrepancy can be attributed to two factors. First, there is no exogenous upper bound on harvests in Levhari and Mirman (1980). Second, the discrete-time formulation of the net reproduction function in Levhari and Mirman (1980) is quite different from the continuous-time formulation in Dockner and Sorger (1996), as the discrete-time formulation implies that agents are able to make some short-term commitment to their intended harvest levels.[15] While there may exist many MPNEs, some of which can be discontinuous, it can be shown that there exists a class of utility functions that yield MPNE strategies that are linear in the stock, provided that the resource growth function has parameters that are suitably related to the parameters of the utility function. See Gaudet and Lohoues (2008) and Long (2011). For existence theorems on MPNEs in resource exploitation games, see Amir (1987, 1987).

[13]Dockner and Sorger (1996), Lemma 1.

[14]Dockner and Long (1993) find similar results in a pollution game.

[15]Efficiency can also be ensured if players can resort to trigger strategies, see Cave (1987) and Benhabib and Radner (1992), or if there exist countervailing externalities, as in Martin-Herrán G, Rincón-Zapareto J (2005).

2.2 Renewable Resource Exploitation Under Oligopoly

While most models of renewable resource extraction assume price-taking behavior, there has been a recent increase in interest on the implications of oligopolistic behavior for renewable resources. Most authors rely on specific demand functions in order to derive closed-form solutions (Benchekroun 2008; Dockner et al. 1989; Fujiwara 2011; Jørgensen and Yeung 1996). Jørgensen and Yeung (1996) assume that the demand function is of the form $P = 1/\sqrt{H}$ where H is the aggregate harvest while the cost of harvesting h_i is $ch_i/\sqrt{S}$ where S is the resource stock. Combining with a square root function for the resource growth rate, the model yields MPNE strategies that are linear in the stock. Long (2011) provides a generalization of the model. More recent contributions discuss the role of property rights (Colombo and Labrecciosa 2013a,b), Bertrand rivalry versus Cournot rivalry (Colombo and Labrecciosa 2015), the role of nonlinear strategies (Colombo and Labrecciosa 2015; Lambertini and Mantovani 2014), and the impact of market integration in an international trade framework (Fujiwara 2011).

Benchekroun (2008) assumes the linear demand function $P = A - BH$, with an arbitrary number of firms. To derive closed-form value functions, he approximates the logistic growth function with a tent-shaped function. The slope of the tent at the zero stock level is called the inherent growth rate of the resource. He finds that there exists an MPNE where fishing firms use a piecewise linear strategy: when the stock is small, firms do not harvest at all, until a threshold level of stock is reached. Beyond that threshold, the equilibrium harvesting rate is linear in the stock, until an upper threshold stock level is reached. For stock levels higher than this upper threshold, firms behave as if they had no concern for the stock dynamics. Myopia becomes individually optimal in this range. Benchekroun (2008) obtains a number of interesting results. First, an increase in the inherent growth rate of the resource may result in a lower steady-state stock. This is similar to the voracity effect discussed in Tornell and Lane (1999). Second, reducing the number of oligopolists can lead to higher steady-state output of the industry, in contrast to the results of the model of oligopoly without a resource stock. This result, at first surprising, can be explained by Solow's idea that a monopolist is the conservationist's best friend.

Benchekroun's 2008 model of oligopolistic exploitation of a renewable resource has been modified and extended in several directions, to examine a number of related issues, such as asymmetry among firms (Benchekroun et al. 2014) and mergers (Benchekroun and Gaudet 2015). In Benchekroun et al. (2014), it is found that the simple piecewise linear strategies in Benchekroun (2008) cannot survive a small departure from the symmetric cost assumption. In Benchekroun and Gaudet (2015), the authors show that there exists an interval of asset stock size such that when the common property stock is inside that interval, any merger is profitable, contrary to the standard static model of merger which asserts that any merger involving less than 80 % of the industry will be unprofitable (Salant et al. 1983). Intuitively, the difference is due to the role of the resource stock (an asset) which constraint cumulative output in a resource oligopoly, while in the standard model of Salant et al. (1983), production is not constrained by assets.

2.3 The Effects of Status Concern on the Exploitation of Renewable Resources

While the standard economic theory emphasizes rationality leading to profit maximization and maximization of the utility of consumption, it is well known that there are other psychological factors that are also driving forces behind our actions.[16] Perceptive economists such as Veblen (1899) and noneconomists, such as Kahneman and Tversky (1984), have stressed these factors. Unfortunately, the "Standard Model of Economic Behavior" does not take into account psychological factors such as emulation, envy, status concerns, and so on. Fortunately, in the past two decades, there has been a growing economic literature that examines the implications of relaxing the standard economic assumptions on preferences (see, e.g., Frey and Stutzer 2007).

The utility that an economic agent derives from her consumption, income, or wealth tends to be affected by how these compare to other economic agents' consumption, income, or wealth. The impact of status concern on resource exploitation has recently been investigated in the natural resource literature. Alvarez-Cuadrado and Long (2011) model status concern by assuming that the utility function of the representative individual depends not only on her own level of consumption c_i and effort L_i but also on the average consumption level in the economy, C, such that $u_i = U(c_i, C, L_i)$. This specification captures the intuition that lies behind the growing body of empirical evidence that places interpersonal comparisons as a key determinant of individual well-being. Denote the marginal utility of own consumption, average consumption, and effort by U_1, U_2, and U_L, respectively. The level of utility achieved by the representative individual is increasing in her own consumption but at a decreasing rate, $U_1 > 0$ and $U_{11} < 0$, and decreasing in effort, $U_L < 0$. In addition, it is assumed that the utility function is jointly concave in individual consumption and effort with $U_{1L} \leq 0$, so the marginal utility of consumption decreases with effort. Under this fairly general specification, Alvarez-Cuadrado and Long (2011) show that relative consumption concerns can cause agents to overexploit renewable resources even when these are private properties. Situations where status-conscious agents exploiting a common pool resource behave strategically are analyzed in Long and Wang (2009), Katayama and Long (2010), and Long and McWhinnie (2012).

Long and McWhinnie (2012) consider a finite number of agents playing a Cournot dynamic fishery game, taking into account the effect of the fishing effort of other agents on the evolution of the stock. In other words, they are dealing with a differential game of fishery with status concerns. Long and McWhinnie (2012) show that the overharvesting associated with the standard tragedy of the commons problem becomes intensified by the desire for higher relative performance, leading to a smaller steady-state fish stock and smaller steady-state profit for all the fishermen. This result is quite robust with respect to the way status is modeled.

[16] See e.g. Fudenberg and Levine (2006, 2012).

The authors consider two alternative specifications of relative performance. In the first specification, relative performance is equated to relative after-tax profits. In the second specification, it is relative harvests that matter. The authors examine a tax package (consisting of a tax on relative profit and a tax on effort) and an individual quota as alternative policy tools to implement the socially efficient equilibrium.

The analysis of Long and McWhinnie (2012) relies on two key assumptions: first, each agent takes as given the time paths of resource exploitation of other agents (i.e., the authors restrict attention to open-loop strategies), and second, the agents take the market price of the extracted resource as given (i.e., the goods markets are perfectly competitive). Those assumptions have been relaxed by Benchekroun and Long (2016). Interestingly, they show that when agents use feedback strategies and the transition phase is taken into account, the well-established result that status concern exacerbates the tragedy of the commons must be seriously qualified. More specifically, when agents are concerned about their relative profit, the authors find that there exists an interval of the stock size of the resource for which the extraction policy under status concern is less aggressive than the extraction policy in the absence of status concern.

2.4 Regime-Switching Strategies and Resource Exploitation

Rivalry in the exploitation of common property resources can motivate players to take additional action (other than the choice of the rates of extraction) in order to get an upper hand. The enclosure of the commons is one way of affecting a regime change (Smith 1776, Book 1, Chap. 11), though it may lead to a loss of social welfare when the enclosing costs are high (Long 1994). As Long (1994) points out, the party that encloses the commons is affecting the other parties' production sets (their ability to transform their labor into food). This is itself a kind of externalities that might be more severe than the overcrowding externalities.

Crabbé and Long (1993) study a fishery game where a dominant player deters entry of poachers by creating excessive overcrowding, driving their profits to zero. Tornell (1997) models the game between two infinitely lived agents who fight over the choice over property rights regime: sharing versus exclusive ownership. He shows that a potential equilibrium of the game involves multiple switching between regimes. Thus Tornell's model sheds light on the political instability of some resource-rich economies.

Long et al. (2014) model the choice of switching from one exploitation technology to another when two infinitely lived agents with different costs of technology adoption have common access to a resource stock. They find that the player with low investment cost is the first player to adopt a new harvesting technology. She faces two countervailing incentives: on the one hand, an early switch to a more efficient technology enables her to exploit the resources more cheaply; on the other hand, by inducing the regime change, which tends to result in a faster depletion, she might give her opponent an incentive to hasten the date of his technology adoption, if the opponent investment cost decreases as the stock decreases. As a consequence, in a Markov-perfect equilibrium, the balance of these strategic considerations may make the low-cost player delay technology adoption even if her fixed cost of adoption is zero, contrary to what she would do (namely, immediate adoption) if she were the sole player.

3 Exhaustible Resources

Exhaustible resources (also called nonrenewable resources) are resources for which the rate of change of the *individual* stocks is never positive (even though the *aggregate* stock may increase through discovery of additional stocks). In the simplest formulation, the transition equation for an exhaustible resource stock S is

$$\frac{dS}{dt} = -\sum_{i=1}^{m} E_i(t), \text{ with } S(0) = S_0 > 0$$

where $E_i(t) \geq 0$ denotes the extraction rate of player i. If $m = 1$, the stock is extracted by a single firm. In the case of a shared resource stock, we have $m \geq 2$. There are two different meanings of resource exhaustion. Physical exhaustion means that extractions continue until the stock becomes zero at some finite time T (or possibly asymptotically). In contrast, economic exhaustion means that at some stage, the firm finds it optimal to abandon the stock because the extraction cost becomes too high, even though extraction is still feasible. Depending on the types of questions the researcher is asking, one formulation of exhaustion may be more appropriate than the other.[17] In the case of eventual physical exhaustion, it is most transparent that the opportunity cost of extracting one more unit of the resource this period is the foregone marginal profit next period as that unit would no longer be available for extraction in the next period. Thus, intertemporal arbitrage implies that along an *equilibrium* extraction path, the discounted marginal profits from extraction must be the same between any two adjacent periods. This is known as the Hotelling Rule.[18] Observed extraction paths are not necessarily equilibrium paths because of unanticipated supply shocks or demand shocks. In fact, models of dynamic games involving exhaustible resources were developed after the unanticipated quadrupling in the world price of petroleum between late 1973 and early 1974, "engineered by the newly assertive Organization of Petroleum Exporting Countries (OPEC), an international cartel that includes most large oil producers" (Krugman et al. 2015, p. 572). Not surprisingly, a major emphasis of this literature is on cartel and oligopolies.

[17] See Salo and Tahvonen (2001) for the modeling of economic exhaustion in a duopoly.

[18] See Gaudet (2007) for the theory and empirics related to the Hotelling Rule.

3.1 Exhaustible Resource Extraction Under Different Market Structures

Salant (1976) considers an open-loop game between an exhaustible resource cartel and a competitive fringe, under Nash-Cournot behavior: the cartel takes the time path of extraction of the fringe as given and determines its own time path, knowing that it can influence the market price. Salant finds that the formation of a cartel raises the profits of its members, compared to the case where all firms are price takers. However, nonmembers gain more than cartel members. This result suggests that an exhaustible resource cartel is likely to face defection or cheating by its members. This might well explain the instability of oil prices in the recent history. Ulph and Folie (1980) extend Salant's model to allow for differences in marginal costs. Gilbert (1978) considers instead the case where the cartel is an open-loop Stackelberg leader: it announces to the fringe its time path of future output before the fringe firms make their output decision. However, it can be shown that an open-loop Stackelberg equilibrium is time inconsistent: at a later stage, if the cartel can renege on its preannounced path, it will find it profitable to do so.[19] Benchekroun and Withagen (2012) provide a theoretical justification for the price-taking behavior of the fringe.

To overcome the time-inconsistency problem, Groot et al. (2003) propose a feedback Stackelberg formulation. This formulation assumes that each fringe firm believes its value function to be a linear function of its stock, with a constant slope, which it takes as given. However, this slope is not given: it is in fact influenced by the cartel's extraction policy.

An alternative market structure is oligopoly. Loury (1986) studies a model of oil oligopolists that use open-loop strategies.[20] He finds that under identical and constant extraction costs, smaller firms exhaust their stocks before larger ones and that industry production maximizes a weighted average of profits and consumers' welfare. Benchekroun et al. (2009, 2010) find that under open-loop oligopoly, firms with different costs may produce at the same time, and additional stocks of the resource can result in a lower social welfare. The latter result has a counterpart in the theory of static oligopoly: a small reduction in the marginal cost of higher cost firms may reduce welfare (Lahiri and Ono 1988; Long and Soubeyran 2001). It is also related to Gaudet and Long (1994), who find that a marginal redistribution of resource stocks between two oligopolists to make their reserves more unequal can increase the industry's profit. Models of oligopoly with feedback extraction strategies include Salo and Tahvonen (2001) and Benchekroun and Long (2006). The latter paper shows that a windfall gain (a stock discovery) can be harmful to firms in a nonrenewable resource oligopoly.

[19] For a proof of the time inconsistency of open-loop Stackelberg equilibrium, see, for example, Dockner et al. (2000), or Long (2010).

[20] See also Lewis and Schmalensee (1980).

3.2 Dynamic Resource Games Between Countries

The world markets for gas and oils consist mainly of a small number of large sellers and buyers. For instance, the US Energy Information Administration reports that the major energy exporters concentrate on the Middle East and Russia, whereas the United States, Japan, and China have a substantial share in the imports. These data suggest that bilateral monopoly roughly prevails in the oil market in which both parties exercise market power. What are the implications of market power for welfare of importing and exporting countries and the world?

Kemp and Long (1979) consider an asymmetric two-country world. They assume that the resource-rich economy can only extract the resource, while the resource-poor economy imports the resource as an input in the production of the consumption goods. They study the implications of market power in the resource market by comparing a competitive equilibrium path of extraction and final good production with the outcome under two scenarios where market power is exercised by only one of the countries. If the resource-rich country is aggressive, it will set a time path of oil price so that the marginal revenue from oil exports rises a rate equal to the endogenously determined rate of interest. In the special case where the production function of the final good is Cobb-Douglas, the resource-rich country is not better off relative to the competitive equilibrium.[21] If the resource-poor country is aggressive, it will set a specific tariff path that makes oil producers's price equal to extraction cost, thus effectively appropriating all the resource rents. Kemp and Long (1979) point out that this result will be attenuated if the resource-rich country can also produce the consumption good.[22] Bergstrom (1982) considers a model with many resource-importing countries. He assumes that the international market is integrated so that all importing countries pay the same price for the resource. He shows that if the resource-poor countries can commit to a time-invariant ad valorem tariff rate on oil, they can extract a sizable gain at the expense of resource-rich economies.

Kemp and Long (1980) present a three-country model where there is a dominant resource-poor economy that acts as an open-loop Stackelberg leader in announcing a time path of per-unit tariff rate, while the resource-rich country and the rest of the world are passive. They show that such a time path is time inconsistent, because at a later stage, having been able to induce the resource-rich country to supply more earlier on, the leader will have an incentive to reduce the tariff rate so as to capture a larger share of the world's oil imports.[23]

Karp and Newbery (1992) numerically compute time-consistent tariff policies in a game where several resource-poor economies noncooperatively impose tariffs

[21]This corresponds to the result that, in a closed economy, the market power of an oil monopolist with zero extraction cost disappears when the elasticity of demand is constant. See, e.g., Stiglitz (1976).

[22]This is confirmed in Brander and Djajic (1983), who consider a two-country world in which both countries use oil to produce a consumption good, but only one of them is endowed with oil.

[23]See also Karp (1984) and Maskin and Newbery (1990) for the time-inconsistency issue.

on oil. Assuming that oil producers are price takers and plan their future outputs according to some Markovian price-expectation rule, the authors report their numerical results that is possible for oil-importing countries to be worse off relative to the free trade case. In a different paper, Karp and Newbery (1991) consider two different orders of move in each infinitesimal time period. In their importer-move-first model, they assume that two importing countries noncooperatively choose the quantity to be imported. In the exporter-move-first model, the competitive exporting firms choose how much to export before they know the tariff rates for the period. The authors report their numerical findings that for small values of the initial resource stock, the importer-move-first model yields lower welfare for the importers compared to the exporter-move-first model.

Rubio and Estriche (2001) consider a two-country model where a resource-importing country can tax the polluting fossil fuels imported from the resource-exporting country. Revisiting that model, Liski and Tahvonen (2004) show that there are two incentives for the resource-importing country to intervene in the trade: taxing the imports of fossil fuels serves to improve the importing country's terms of trade, while imposing a carbon tax is the Pigouvian response to climate-change externalities. They show that the gap between the price received by fossil-fuel exporters and the price faced by consumers in the importing country can be decomposed into two components, reflecting the terms-of-trade motive and the Pigouvian motive.

Chou and Long (2009) set up a model with three countries: two resource-importing countries set tariff rates on imported oil, and a resource-exporting country controls the producer's price. It is found that, in a Markov-perfect Nash equilibrium, as the asymmetry between the importing countries increases, the aggregate welfare of the importing countries tends to be higher than under global free trade. The intuition is as follows. With two equally large buyers, the rivalry between them dilutes their market power. In contrast, when one buyer is small and the other is large, the large buyer is practically a monopsonist and can improve its welfare substantially, which means the sum of the welfare levels of both buyers is also larger. Rubio (2011) examines Markov-perfect Nash equilibriums in a dynamic game between a resource-exporting country and n identical noncooperative importing countries that set tariff rates. Rubio (2011) compares the case where the exporting country sets price and the case where it sets quantity. Using a numerical example, he finds that consumers are better off when the seller sets quantity.

Fujiwara and Long (2011) propose a dynamic game model of bilateral monopoly in a resource market where one of the country acts as a global Markovian Stackelberg leader in the sense that the leader announces a stock-dependent (i.e., Markovian) decision rule at the outset of the game, and then the follower chooses its response, also in the form of a stock-dependent decision rule.[24]

[24]For discussions of the concept of global feedback Stackelberg equilibrium, see Basar and Olsder (1982) and Long and Sorger (2010). An alternative notion is the stagewise Stackelberg leadership, which will be explained in more detail in Sect. 3.3 below.

The resource-exporting country posts a price using a Markovian decision rule, $p = p(S)$, where S is the current level of the resource stock. The importing country sets a per-unit tariff rate τ which comes from a decision rule $\tau(S)$. The authors impose a time-consistency requirement which effectively restricts the set of strategies the leader can choose from. They show that the presence of a global Stackelberg leader leaves the follower worse off compared with its payoff in a Markov-perfect Nash equilibrium. Moreover, world welfare is highest in the Markov-perfect Nash equilibrium. These results are in sharp contrast with the results of Tahvonen (1996) and Rubio and Estriche (2001) who, using the concept of stagewise Stackelberg equilibrium, find that when the resource-exporting country is the leader, the stagewise Stackelberg equilibrium coincides with the Markov-perfect Nash equilibrium.[25]

In a companion paper, Fujiwara and Long (2012) consider the case where the resource-exporting country (called Foreign) determines the quantity to sell in each period. There are two resource-importing countries: a strategic, active country, called Home, and a passive country, called ROW (i.e., the rest of the world). The market for the extracted resource is integrated. Therefore Foreign's resource owners receive the same world price whether they export to Home or to ROW. Moreover, Home's consumers must pay a tax τ on top of the world price, while consumers in ROW only pay the world price. Home chooses τ to maximize Home's welfare. Fujiwara and Long (2012) show that, compared with the Markov-perfect Nash equilibrium, both countries are better off if Home is the global Markovian Stackelberg leader. However, if the resource-exporting country is the global Markovian Stackelberg leader, Home is worse off compared to its Markov-perfect Nash equilibrium welfare.

Finally, in managing international trade in fossil fuels, resource-exporting countries should take into account the fact that importing countries cannot be forced to pay a higher price than the cost of alternative energy sources that a backstop technology can provide. Hoel (1978) shows how a fossil-fuel monopolist's market power is constrained by the existence of a backstop technology that competitive firms can use to produce a substitute for the fossil fuels. This result has been generalized to the case with two markets (van der Meijden 2016). Hoel (2011) demonstrates that when different countries have different costs of using a backstop technology, the imposition of a carbon tax by one country may result in a "Green Paradox," i.e., in response to the carbon tax, the near-future extraction of fossil fuels may increase, bringing climate change damages closer to the present. Long and Staehler (2014) find that a technological advance in the backstop technology may result in a similar Green Paradox outcome. For a survey of the literature on the Green Paradox in open economies, see Long (2015b).

[25] In a stagewise Stackelberg equilibrium, no commitment of any significant length is possible. The leader can only commit to the current period decision.

3.3 Fossil Resources and Pollution

Among the most important economic issues of the twenty-first century is the impending risk of substantial damages caused by climate change, which is inherently linked to an important class of exhaustible resources: fossil fuels, such as oil, natural gas, and coal. The publication of the Stern Review (2006) has provided impetus to economic analysis of climate change and policies toward fossil fuels. For a small sample of this large literature, see Heal (2009) and Haurie et al. (2011).

Wirl (1994) considers a dynamic game between a resource-exporting country and an importing country that suffers from the accumulated pollution which arises from the consumption of the resource, $y(t)$. Let $Z(t)$ denote the stock of pollution and $S(t)$ denote the stock of the exhaustible resource. Assume for simplicity that the natural rate of pollution decay is zero. Then $\dot{Z}(t) = y(t) = -\dot{S}(t)$, and hence $S(0) - S(t) = Z(t) - Z(0)$. The stock of pollution gives rise to the damage cost $DZ(t)^2$. The importing country imposes a carbon tax rate τ according to some Markovian rule $\tau(t) = g(Z(t))$. The exporting country follows a pricing rule $p(t) = \phi(S(t)) = \phi(Z(0) - Z(t) + S(0))$. Along the Markov-perfect Nash equilibrium, where $g(.)$ and $\phi(.)$ are noncooperative chosen by the importing country and the exporting country, it is found that the carbon tax rate will rise, and if $S(0)$ is sufficiently large, eventually the consumption of the exhaustible resource tends to zero while the remaining resource stock tends to some positive level $S_L > 0$. This is the case of economic exhaustion, because the equilibrium producer price falls to zero due to rising carbon taxation.

Tahvonen (1996) modifies the model of Wirl (1994) by allowing the exporting country to be a stagewise Stackelberg leader. As explained in Long (2011), if the time horizon is finite and time is discrete, stagewise leadership by the exporter means that in each period, the resource-exporting country moves first by announcing the well-head price p_t for that period. The government of the importing country (the stagewise follower) reacts to that price by imposing a carbon tax τ_t for that period. Working backward, each party's payoff for period $T-1$ can then be expressed as a function of the opening pollution stock, Z_{T-1}. Then, in period $T-2$, the price p_{T-2} is chosen and so on. For tractability, Tahvonen (1996) works with a model involving continuous time and an infinite horizon, which derives its justification by shrinking the length of each period and taking the limit as the time horizon becomes arbitrarily large. He finds that the stagewise Stackelberg equilibrium of this model coincides with the Markov-perfect Nash equilibrium.[26] Liski and Tahvonen (2004) decompose the carbon tax into a Pigouvian component and an optimal tariff component.

Different from the stagewise Stackelberg approach of Tahvonen (1996), Katayama et al. (2014) consider the implication of global Markovian Stackelberg leadership in a model a dynamic game involving a fossil-fuel-exporting cartel and

[26]This result is confirmed by Rubio and Estriche (2001) who modify the model of Tahvonen (1996) by assuming that the per-unit extraction cost in period t is $c \times (S(0) - S(t))$, where c is a positive parameter.

a coalition of importing countries that suffer from accumulated emissions and impose a carbon tax on the fossil fuel. Referring to Fujiwara and Long (2011), who do not consider pollution, Katayama et al. (2014) impose a time-consistency requirement on the Markovian strategy of the global Stackelberg leader. They find that world welfare under the social planner is strictly greater than world welfare under the Markov-perfect Nash equilibrium, which in turn dominates world welfare when the exporting country is the global Stackelberg leader. When the coalition of the importing countries is the global Stackelberg leader, world welfare is lowest compared to the other scenarios. Finally, while the linear-quadratic structure is conducive to analytical solution, there is a need to go beyond that structure. Bearing this in mind, Kagan et al. (2015) take a big step forward in the analysis of resource depletion and climate change, with the help of advanced numerical techniques.

3.4 Extraction of Exhaustible Resources Under Common Access

In the preceding subsections, we have assumed that the property rights of the exhaustible resource stocks are well defined and well enforced. However, there are instances where some exhaustible resources are under common access. For examples, many oil fields are interconnected. Because of seepage, the owner of each oil field in fact can "steal" the oil of his neighbors. Under these conditions, the incentive for each owner to conserve his resource is not strong enough to ensure an efficient outcome. The belief that common access resources are extracted too fast has resulted in various regulations on extraction (McDonald 1971; Watkins 1977).

Khalatbary (1977) presents a model of m oligopolistic firms extracting from m interconnected oil fields. It is assumed that there is an exogenous seepage parameter $\beta > 0$ such that, if $E_i(t)$ denotes extraction from stock $S_i(t)$, the rate of change in $S_i(t)$ is

$$\dot{S}_i(t) = -E_i(t) - \beta S_i(t) + \frac{\beta}{m-1} \sum_{j \neq i} S_j(t)$$

The price of the extracted resource is $P = P\left(\sum E_j\right)$. Khalatbary (1977) assumes that firm i maximizes its integral of the flow of discounted profits, subject to a single transition equation, while taking the time paths of *both* $E_j(t)$ *and* $S_j(t)$ and as given, for all $j \neq i$. He shows that at the open-loop Nash equilibrium, the firms extract at a faster rate than they would if there were no seepage.[27] Kemp and Long (1980, p. 132) point out that firm i should realize that $S_j(t)$ is indirectly dependent on the time path of firm i's extraction, because $\dot{S}_j(t)$ depends on $S_i(t)$ which in turn is affected by firm i's path of extraction from time 0 up to time t. Thus, firm i's

[27] Dasgupta and Heal (1979, Ch. 12) consider the open-loop Nash equilibrium of a similar seepage problem, with just two firms, and reach similar results.

dynamic optimization problem should include m transition equations, not just one, and thus, firm i can influence $S_j(t)$ indirectly.[28] Under this formulation, Kemp and Long (1980) find that the open-loop Nash equilibrium can be efficient.[29]

McMillan and Sinn (1984) propose that each firm conjectures that the extraction of other firms obeys a Markovian rule of the form $\alpha(t) + \gamma S(t)$ where $S(t)$ is the aggregate stock. Their objective is to determine $\alpha(t)$ and γ such that the expectations are fulfilled. They find that there are many equilibria. They obtain the open-loop results of Khalatbary (1977), Dasgupta and Heal (1979), Kemp and Long (1980), Bolle (1980), and Sinn (1984) as special cases: if $\gamma = 0$ and $\alpha(t)$ is the extraction path, one obtains an open-loop Nash equilibrium.

Laurent-Lucchetti and Santugini (2012) combine common property exhaustible resources with uncertainty about expropriation, as in Long (1975). Consider a host country that allows two firms to exploit a common resource stock under a contract that requires each firm to pay the host country a fraction τ of its profit. Under the initial agreement, $\tau = \tau_L$. However, there is uncertainty about how long the agreement will last. The host country can legislate a change in τ to a higher value, τ_H. It can also evict one of the firms. The probability that these changes occur is exogenous. Formulating the problem as a dynamic game between the two firms, in which the risk of expropriation is exogenous and the identity of the firm to be expropriated is unknown ex ante, the authors find that weak property rights have an ambiguous effect on present extraction. Their theoretical finding is consistent with the empirical evidence provided by in Deacon and Bohn (2000).

3.5 Effectiveness of Antibiotics as an Exhaustible Resource

The exhaustible resource model can be modified to study the Markov-perfect equilibrium rate of decrease in the effectiveness of drugs such as antibiotics, when users fail to take into account the externalities of their actions on the payoff of other users. In an editorial on 21 December 2013, titled "The Perils of Antibiotic Use on Farms," the New York Times reported that:

> The rampant use of antibiotics in agriculture has been alarming. The drugs are given not just to treat sick animals, but added in low doses to animal feed or water to speed the growth of cattle, pigs and chickens, thus reducing costs for the producers. Such widespread use of antibiotics in healthy animals has stimulated the emergence of bacterial strains that are resistant to antibiotics and capable of passing their resistance to human pathogens, many of which can no longer be treated by drugs that were once effective against them.
>
> Each year, at least two million Americans fall ill — and 23,000 die — from antibiotic-resistant infections. Doctors are partly to blame because many prescribe antibiotics for conditions like colds that can't be cured with such drugs. The Centers for Disease Control

[28] Sinn (1984) considers a different concept of equilibrium in the seepage model: each firm is committed to achieve a given time path of its stock.

[29] Bolle (1980) obtains a similar result, assuming that there is only one common stock that all m firms have equal access.

> and Prevention estimated in September that up to half of the antibiotics prescribed for humans are not needed or are used inappropriately. It added, however, that overuse of antibiotics on farms contributed to the problem.

This raises the question of how to regulate the use of antibiotics in an economy, given that other economies may have weaker regulations which help their farmers realize more profits in the short run, as compared with the profits in economies with stronger regulations. Cornes et al. (2001) consider two models of dynamic game on the use of antibiotics: a discrete-time model and a continuous-time model. Assume n players share a common pool, namely, the effectiveness of an antibiotic. Their accumulated use of the antibiotic decreases its effectiveness: the more they use the drug, the quicker the bacteria develop their resistance. The discrete-time model yields the result that there are several Markov-perfect Nash equilibria, with different time path of effectiveness. For the continuous-time model, there is a continuum of Markov-perfect Nash equilibria.

There are $n \geq 2$ countries. Let $S(t)$ denote the effectiveness of the antibiotic and $E_i(t)$ denote its rate of use in country i. The rate of decline of effectiveness is described in the following equation:

$$\dot{S}(t) = -\beta \sum\nolimits_{i=1}^{n} E_i(t), \beta > 0, \qquad S(0) = S_0 > 0.$$

Assume that the benefit to country i of using $E_i(t)$ is $B_i(t) = (S(t)E_i(t))^{\alpha}$, where $0 < \alpha < 1$. Call $E_i(t)$ the nominal dose and $S(t)E_i(t)$ the effective dose. If the countries coordinate their policies, the cooperative problem is to maximize the integral of discounted benefits, where $r > 0$ is the discount rate. The optimal cooperative policy rule is linear: $E_i(t) = rS(t)/2\beta n(1-\alpha)$. In the noncooperative scenario, each country uses a feedback strategy $E_i = \phi_i(S)$. Assuming that $\alpha < 1/n$, Cornes et al. (2001) find that there is a Markov-perfect Nash equilibrium where all countries use the linear strategy $E_i(t) = rS(t)/2\beta n(n^{-1} - \alpha)$. Thus, the effectiveness of the antibiotic declines at a faster rate than is socially optimal. Interestingly, in addition to the above linear strategy equilibrium, Cornes et al. (2001) show that there is a continuum of Markov-perfect Nash equilibria where all countries use nonlinear strategies, and S becomes zero at some finite time. Non-uniqueness has also been reported in Clemhout and Wan (1995).

Thus, when countries do not coordinate their policies on the use of biological assets, the result is overexploitation. Another problem of rivalry in a broader biological context is the biological arms race between species, as discussed in Dawkins and Krebs (1979). The lesson is that whatever biological techniques humans may devise in their efforts to exploit and utilize the resources that nature has to offer, we are likely to find ourselves in an arena in which our competitors will fight back. The continuing struggle is as old as life itself and indeed inseparable from it.

Herrmann and Gaudet (2009) also analyze the exploitation of antibiotic effectiveness in terms of a common pool problem. They think of a generic product which takes over once a patent has expired. The authors take into account the interaction between the level of efficacy of the drug and the level of infection in the population. The model is based on an epidemiological model from the biology literature. Unlike Cornes et al. (2001) and Herrmann and Gaudet (2009) do not formulate a differential game model, because they assume that no economic agent takes into account the dynamic effects of their decision.

4 Directions for Future Research

There are a number of issues in resource economics that remain under-explored. The first issue is the spatial dimension. Exhaustible resource stocks are unevenly distributed around the globe, and this fact necessitates the transportation of the extracted resources to consumers. How do resource-exporting firms located at different places compete with each other for customers over time and space? What would be the properties of Markov-perfect equilibrium involving spatially separated resource-extracting oligopolists?[30]

Similarly, renewable resources, such as fish stocks, are also dispersed in space. Harvesting fleets are not stationary: they typically have to travel and fish at many locations. Behringer and Upman (2014) model a fleet that moves along a circle to catch fishes. Their model involves both space and time. However, they do not address the issue of dynamic games (across space and time) among different fleets, and they assume that fish do not move from one pool to another. Modeling dynamic fishing strategies when fish move from one place to another is surely a challenging research topic.[31]

The second topic that deserves exploring is learning about the properties of resources that one exploit, for example, discovering more precise information about the growth function of a resource stock. Mirman and Santugini (2014) have made a useful step in this direction. A third topic is how to provide incentives for cooperation. In this context, we note that de Frutos and Martín-Herrán (2015) provide useful analysis of a generalized concept of incentive equilibrium such that players' behavior (including a Markovian type of punishment) ensures that the system is sufficiently close to the fully cooperative equilibrium outcome. They also give a very clear definition of the concept of incentive equilibrium, an informative historical account of the development and application of this concept, and show how to compute such equilibria numerically. However, since in their example, de Frutos and Martín-Herrán (2015) restrict attention to the linear-quadratic case, much work remains to be done for a general treatment.

The fourth issue is the political economy of resource conservation. The majority of the current electorate may have very little interest in conserving natural resources. Governments may have to balance the need of future generations with the

[30] The case of open-loop Nash equilibrium was addressed by Kolstad (1994) and Keutiben (2014).

[31] This is related to the so-called SLOSS debate in ecology, in which authors disagree as to whether a single large or several small (SLOSS) reserves would be better for conservation.

impatience of the current voters. What would be an appropriate formulation of the dynamic games among generations? One possible way of addressing this problem is to think of the political process as the dual-self problem, as in Fudenberg and Levine (2002, 2012).

Finally, it is high time to depart from the assumption that all players are selfish. Dynamic game models of natural resource exploitation typically rely on that assumption, which clearly leads to the prediction of overexploitation of many resource stocks. However, as Ostrom (1990) points out, in some societies, good social norms are sufficiently developed to avoid the tragedy of the commons. What would be the dynamic evolution of resource stocks if some kinds of social norms are developed to guide the behavior of economic agents?[32] The importance of social norms was recognized by classical economists. Adam Smith (1790) finds that cooperation and mutual help are incorporated in established norms of behavior and that

> upon the tolerable observance of these duties, depend the very existence of human society, which would **crumble into nothing** if mankind were not generally impressed with a reverence for those important rules of conduct. (Smith 1790, Part III, Chap. V, p. 190)

Clearly, Smith's view is that for societies to prosper, there is a need for two invisible hands, not just one. First is the moral invisible hand that encourages the observance of duties; second is the invisible hand of the price system, which guides the allocation of resources. Along the same lines, Roemer (2010, 2015) formulates the concept of Kantian equilibrium, for games in which players are imbued with Kantian ethics (Russell 1945). By definition, this equilibrium is a state of affairs in which players of a common property resource game would not deviate when each finds that if she was to deviate and everyone else would do likewise, she would be worse off. However, Roemer (2010, 2015) restricts attention to static games, as does Long (2016a,b) for the case of mixed strategy Kantian equilibria. Dynamic extension of the concept of Kantian equilibrium has been explored by Long (2015a) and Grafton et al. (2016), who also defined the concept of dynamic Kant-Nash equilibrium to account for the coexistence of Kantian agents and Nashian agents. However, Long (2015a) and Grafton et al. (2016) did not deal with the issue of how the proportion of Kantian agents may change over time, due to learning or social influence. Introducing evolutionary elements into this type of model remains a challenge.[33]

References

Alvarez-Cuadrado F, Long NV (2011) Relative consumption and renewable resource extraction under alternative property rights regimes. Resour Energy Econ 33(4):1028–1053

[32]Myerson and Weibull (2015) formalize the idea that "social conventions usually develop so that people tend to disregard alternatives outside the convention."

[33]For a sample of papers that deal with evolutionary dynamics, see Bala and Long (2005), Breton et al. (2010), and Sethi and Somanathan (1996).

Amir R (1987) Games of resource extraction: existence of Nash equilibrium. Cowles Foundation Discussion Paper No. 825
Amir R (1989) A Lattice-theoretic approach to a class of dynamic games. Comput Math Appl 17:1345–1349
Bala V, Long NV (2005) International trade and cultural diversity with preference selection. Eur J Polit Econ 21(1):143–162
Basar T, Olsder GJ (1982) Dynamic non-cooperative game theory. Academic, London
Behringer S, Upmann T (2014) Optimal harvesting of a spatial renewable resource. J Econ Dyn Control 42:105–120
Breton M, Sbragia L, Zaccour G (2010) A dynamic model of international environmental agreements. Environ Resour Econ 45:25–48
Benchekroun H (2008) Comparative dynamics in a productive asset oligopoly. J Econ Theory 138:237–261
Benchekroun H, Halsema A, Withagen C (2009) On non-renewable resource oligopolies: the asymmetric case. J Econ Dyn Control 33(11):1867–1879
Benchekroun H, Halsema A, Withagen C (2010) When additional resource stocks reduce welfare. J Environ Econ Manag 59(1):109–114
Benchekroun H, Gaudet G, Lohoues H (2014) Some effects of asymmetries in a common pool natural resource oligopoly. Strateg Behav Environ 3:213–235
Benchekroun H, Gaudet G (2015) On the effects of mergers on equilibrium outcomes in a common property renewable asset oligopoly. J Econ Dyn Control 52:209–223
Benchekroun H, Long NV (2006) The curse of windfall gains in a non-renewable resource oligopoly. Aust Econ Pap 45(2):99–105
Benchekroun H, Long NV (2016) Status concern and the exploitation of common pool renewable resources. Ecol Econ 125(C):70–82
Benchekroun H, Withagen C (2012) On price-taking behavior in a non-renewable resource Cartel-Fringe game. Games Econ Bahav 76(2):355–374
Benhabib J, Radner R (1992) The joint exploitation of a productive asset: a game-theoretic approach. Econ Theory 2:155–190
Bergstrom T (1982) On capturing oil rents with a national excise tax. Am Econ Rev 72(1):194–201
Bohn H, Deacon RT (2000) Ownership risk, investment, and the use of natural resources. Am Econ Rev 90(3):526–549
Bolle F (1980) The efficient use of a non-renewable common-pool resource is possible but unlikely. Zeitschrift für Nationalökonomie 40:391–397
Brander J, Djajic S (1983) Rent-extracting tariffs and the management of exhaustible resources. Canad J Econ 16:288–298
Breton M, Keoula MY (2012) Farsightedness in a coalitional great fish war. Environ Resour Econ 51(2):297–315
Breton M, Sbragia L, Zaccour G (2010) A dynamic model of international environmental agreements. Environ Resour Econ 45:25–48
Cabo F, Tidball M (2016) Promotion of cooperation when benefits come in the future: a water transfer case. Resour Energy Econ (to appear)
Cabo F, Erdlenbruch K, Tidball M (2014) Dynamic management of water transfer between two interconnected river basins. Resour Energy Econ 37:17–38
Cave J (1987) Long-term competition in a dynamic game: the Cold Fish War. Rand J Econ 18: 596–610
Chiarella C, Kemp MC, Long NV, Okuguchi K (1984) On the economics of international fisheries. Int Econ Rev 25 (1):85–92
Chichilnisky G (1996) An axiomatic approach to sustainable development. Soc Choice Welf 13:231–257
Chou S, Long NV (2009) Optimal tariffs on exhaustible resources in the presence of Cartel behavior. Asia-Pac J Account Econ 16(3):239–254
Chwe M (1994) Farsighted coalitional stability. J Econ Theory 63:299–325
Clark CW (1990) Mathematical bioeconomics. Wiley, New York

Clark CW, Munro G (1990) The economics of fishing and modern capital theory. J Environ Econ Manag 2:92–106

Clemhout S, Wan H Jr (1985) Dynamic common property resources and environmental problems. J Optim Theory Appl 46:471–481

Clemhout S, Wan H Jr (1991) Environmental problems as a common property resource game. In: Hamalainen RP, Ehtamo HK (eds) Dynamic games and economic analyses. Springer, Berlin, pp 132–154

Clemhout S, Wan H Jr (1995) The non-uniqueness of Markovian strategy equilibrium: the case of continuous time models for non-renewable resources. In: Paper presented at the 5th international symposium on dynamic games and applications, Grimentz

Colombo L, Labrecciosa P (2013a) Oligopoly exploitation of a private property productive asset. J Econ Dyn Control 37(4):838–853

Colombo L, Labrecciosa P (2013b) On the convergence to the Cournot equilibrium in a productive asset oligopoly. J Math Econ 49(6):441–445

Colombo L, Labrecciosa P (2015) On the Markovian efficiency of Bertrand and Cournot equilibria. J Econ Theory 155:332–358

Cornes R, Long NV, Shimomura K (2001) Drugs and pests: intertemporal production externalities. Jpn World Econ 13(3):255–278

Couttenier M, Soubeyran R (2014) Drought and civil war in Sub-Saharan Africa. Econ J 124(575):201–244

Couttenier M, Soubeyran R (2015) A survey of causes of civil conflicts: natural factors and economic conditions. Revue d'Economie Politique 125(6):787–810

Crabbé P, Long NV (1993) Entry deterrence and overexploitation of the fishery. J Econ Dyn Control 17(4):679–704

Dasgupta PS, Heal GM (1979) Economic theory and exhaustible resources. Cambridge University Press, Cambridge

Dawkins R, Krebs J (1979) Arms races between and within species. Proc R Soc Lond Ser B 10:489–511

de Frutos J, Martín-Herrán G (2015) Does flexibility facilitate sustainability of cooperation over time? A case study from environmental economics. J Optim Theory Appl 165(2):657–677

Deacon R, Bohn H (2000) Ownership risk, investment, and the use of natural resources. Am Econ Rev 90(3):526–549

de Zeeuw A (2008) Dynamic effects on the stability of international environmental agreements. J Environ Econ Manag 55:163–174

Diamantoudi F, Sartzetakis E (2015) International environmental agreements: coordinated action under foresight. Econ Theory 59(3):527–546

Dinar A, Hogarth M (2015) Game theory and water resources: critical review of its contributions, progress, and remaining challenges. Found Trends Microecon 11(1–2):1–139

Dinar A, Zaccour G (2013) Strategic behavior and environmental commons. Environ Develop Econ 18(1):1–5

Dockner E, Long NV (1993) International pollution control: cooperative versus non-cooperative strategies. J Environ Econ Manag 24:13–29

Dockner E, Sorger G (1996) Existence and properties of equilibria for a dynamic game on productive assets. J Econ Theory 71:209–227

Dockner E, Feichtinger G, Mehlmann A (1989) Noncooperative solution for a differential game model of fishery. J Econ Dyn Control 13:11–20

Dockner ESJ, Long NV, Sorger G (2000) Differential games in economics and management science. Cambridge University Press, Cambridge

Dutta PK, Sundaram RK (1993a) How different can strategic model be? J Econ Theory 60:42–91

Dutta PK, Sundaram RK (1993b) The tragedies of the commons? Econ Theory 3:413–426

Fesselmeyer E, Santugini M (2013) Strategic exploitation of a common resource under environmental risk. J Econ Dyn Control 37(1):125–136

Food and Agriculture Organization (FAO) (2009) The state of world fisheries and aquaculture 2008. FAO Fisheries and Aquaculture Department, Food and Agriculture Organization of the United Nations, Rome

Frey BS, Stutzer A (2007) Economics and psychology. MIT Press, Cambridge
Fudenberg D, Levine DK (2006) A dual-self model of impulse control. Am Econ Rev 96: 1449–1476
Fudenberg D, Levine DK (2012) Timing and self control. Econometrica 80(1):1–42
Fudenberg D, Tirole J (1991) Game theory. MIT Press, Cambridge
Fujiwara K (2011) Losses from competition in a dynamic game model of renewable resource oligopoly. Resour Energy Econ 33(1):1–11
Fujiwara K, Long NV (2011) Welfare implications of leadership in a resource market under bilateral monopoly. Dyn Games Appl 1(4):479–497
Fujiwara K, Long NV (2012) Optimal tariffs on exhaustible resources: the case of quantity setting. Int Game Theory Rev 14(04):124004-1-21
Gaudet G (2007) Natural resource economics under the rule of hotelling. Canad J Econ 40(4):1033–1059
Gaudet G, Lohoues H (2008) On limits to the use of linear Markov strategies in common property natural resource games. Environ Model Assess 13:567–574
Gaudet G, Long NV (1994) On the effects of the distribution of initial endowments in a non-renewable resource duopoly. J Econ Dyn Control 18:1189–1198
Grafton Q, Kompas T, Long NV (2016) A brave new world? Interaction among Kantians and Nashians, and the dynamic of climate change mitigation. In: Paper presented at the Tinbergen Institute conference on climate change, Amsterdam
Greenberg J (1990) The theory of social situations: an alternative game-theoretic approach. Cambridge University Press, Cambridge
Gilbert R (1978) Dominant firm pricing policy in a market for an exhaustible resource. Bell J Econ 9(2):385–395
Groot F, Withagen C, de Zeeuw A (2003) Strong time-consistency in a Cartel-versus-Fringe model. J Econ Dyn Control 28:287–306
Hardin G (1968) The tragedy of the commons. Science 162:1240–1248
Haurie M, Tavoni BC, van der Zwaan C (2011) Modelling uncertainty and the climate change: recommendations for robust energy policy. Environ Model Assess 17:1–5
Heal G (2009) The economics of climate change: a post-stern perspective. Clim Change 96: 275–297
Herrmann M, Gaudet G (2009) The economic dynamics of antibiotic efficacy under open access. J Environ Econ Manag 57:334–350
Hoel M (1978) Resource extraction, substitute production, and monopoly. J Econ Theory 19(1): 28–37
Hoel M (2011) The supply side of CO_2 with country heterogeneity. Scand J Econ 113(4):846–865
Jørgensen S, Yeung DWK (1996) Stochastic differential game model of common property fishery. J Optim Theory Appl 90(2):381–404
Jørgensen S, Zaccour G (2001) Time consistent side payments in a dynamic game of downstream pollution. J Econ Dyn Control 25:1973–87
Kagan M, van der Ploeg F, Withagen C (2015) Battle for climate and scarcity rents: beyond the linear-quadratic case. Dyn Games Appl 5(4):493–522
Kahneman D, Tversky A (1984) Choices, values and frames. Am Psychol 39:341–350
Kaitala V, Hamalainen RP, Ruusunen J (1985) On the analysis of equilibria and bargaining in a fishery game. In: Feichtinger G (ed) Optimal control theory and economic analysis II. North Holland, Amsterdam, pp 593–606
Karp L (1984) Optimality and consistency in a differential game with non-renewable resources. J Econ Dyn Control 8:73–97
Karp L, Newbery D (1991) Optimal tariff on exhaustible resources. J Int Econ 30:285–99
Karp L, Newbery D (1992) Dynamically consistent oil import tariffs. Canad J Econ 25:1–21
Katayma S, Long NV (2010) A dynamic game of status-seeking with public capital and an exhaustible resource. Opt Control Appl Methods 31:43–53
Katayama S, Long NV, Ohta H (2014) Carbon tax and comparison of trading regimes in fossil fuels. In: Semmler W, Tragler G, Verliov V (eds) Dynamics optimization in environmental economics. Springer, Berlin, pp 287–314

Kemp MC, Long NV (1979) The interaction between resource-poor and resource-rich economies. Aust Econ Pap 18:258–267
Kemp MC, Long NV (1980) Exhaustible resources, optimality, and trade. North Holland, Amsterdam
Kemp MC, Long NV (1984) Essays in the economics of exhaustible resources. North Holland, Amsterdam
Keutiben O (2014) On capturing foreign oil rents. Resour Energy Econ 36(2):542–555
Khalatbary F (1977) Market imperfections and the optimal rate of depletion of natural resources. Economica 44:409–14
Kolstad CD (1994) Hotelling rents in hotelling space: product differentiation in exhaustible resource markets. J Environ Econ Manag 26:163–180
Krugman P, Obstfeld M, Melitz M (2015) International economics: theory and policy, 10th edn. Pearson, Boston
Kwon OS (2006) Partial coordination in the Great Fish War. Environ Resour Econ 33:463–483
Lahiri S, Ono Y (1988) Helping minor firms reduce welfare. Econ J 98:1199–1202
Lambertini L, Mantovani A (2014) Feedback equilibria in a dynamic renewable resource oligopoly: pre-emption, voracity and exhaustion. J Econ Dyn Control 47:115–122
Laurent-Lucchetti J, Santugini M (2012) Ownership risk and the use of common-pool natural resources. J Environ Econ Manag 63:242–259
Levhari D, Mirman L (1980) The great fish war: an example using a dynamic Cournot Nash solution. Bell J Econ 11:322–334
Lewis T, Schmalensee R (1980) On oligopolistic markets for nonrenewable natural resources. Q J Econ 95:471–5
Liski M, Tahvonen O (2004) Can carbon tax eat OPEC's rents? J Environ Econ Manag 47: 1–12
Long NV (1975) Resource extraction under the uncertainty about possible nationalization. J Econ Theory 10(1):42–53
Long NV (1977) Optimal exploitation and replenishment of a natural resource. Essay 4 in Pitchford JD, Turnovsky SJ (eds) Applications of control theory to economic analysis. North-Holland, Amsterdam, pp 81–106
Long NV (1994) On optimal enclosure and optimal timing of enclosure. Econ Rec 70(221): 141–145
Long NV (2010) A survey of dynamic games in economics. World Scientific, Singapore
Long NV (2011) Dynamic games in the economics of natural resources: a survey. Dyn Games Appl 1(1):115–148
Long NV (2015a) Kant-Nash equilibrium in a dynamic game of climate change mitigations. In: Paper presented at the CESifo area conference on climate and energy, Oct 2015 (CESifo), Munich
Long NV (2015b) The green paradox in open economies: lessons from static and dynamic models. Rev Environ Econ Policies 9(2):266–285
Long NV (2016a) Kant-Nash equilibrium in a quantity-setting oligopoly. In: von Mouche P, Quartieri F (eds) Contributions on equilibrium theory for Cournot oligopolies and related games: essays in Honour of Koji Okuguchi. Springer, Berlin, pp 171–192
Long NV (2016b) Mixed-strategy Kant-Nash equilibria in games of voluntary contributions to a public good. In: Buchholz W, Ruebbelke D (eds) The theory of externalities and public goods. Springer
Long NV, McWhinnie S (2012) The tragedy of the commons in a fishery when relative performance matters. Ecol Econ 81:140–154
Long NV, Sorger G (2006) Insecure property rights and growth: the roles of appropriation costs, wealth effects, and heterogeneity. Econ Theory 28(3):513–529
Long NV, Sorger G (2010) A dynamic principal-agent problem as a feedback Stackelberg differential game. Central Eur J Oper Res 18(4):491–509
Long NV, Soubeyran A (2001) Cost manipulation games in oligopoly, with cost of manipulating. Int Econ Rev 42(2):505–533

Long NV, Staehler F (2014) General equilibrium effects of green technological progress. CIRANO working paper 2014s-16. CIRANO, Montreal
Long NV, Wang S (2009) Resource grabbing by status-conscious agents. J Dev Econ 89(1):39–50
Long NV, Prieur F, Puzon K, Tidball M (2014) Markov perfect equilibria in differential games with regime switching strategies. CESifo working paper 4662. CESifo Group, Munich
Loury GC (1986) A theory of oligopoly: Cournot equilibrium in exhaustible resource market with fixed supplies. Int Econ Rev 27(2):285–301
Martin-Herrán G, Rincón-Zapareto J (2005) Efficient Markov-perfect Nash equilibria: theory and application to dynamic fishery games. J Econ Dyn Control 29:1073–96
Maskin E, Newbery D (1990) Disadvantageous oil tariffs and dynamic consistency. Am Econ Rev 80(1):143–156
McDonald SL (1971) Petroleum conservation in the United States: an economic analysis. Johns Hopkins Press, Baltimore
McMillan J, H-W Sinn (1984) Oligopolistic extraction of a common property resource: dynamic equilibria. In: Kemp MC, Long NV (eds) Essays in the economics of exhaustible resources. North Holland, Amsterdam, pp 199–214
McWhinnie SF (2009) The tragedy of the commons in international fisheries: an empirical investigation. J Environ Econ Manag 57:312–333
Mirman L, Santugini M (2014) Strategic exploitation of a common-property resource under rational learning about its reproduction. Dyn Games Appl 4(1):58–72
Myerson R, Weibull J (2015) Tenable strategy blocks and settled equilibria. Econometrica 83(3):943–976
Ostrom E (1990) Governing the commons: the evolution of institutions for collective actions. Cambridge University Press, Cambridge
Ray D, Vohra R (2001) Coalitional power and public goods. J Polit Econ 109(6):1355–1384
Roemer JE (2010) Kantian equilibrium. Scand J Econ 112(1):1–24
Roemer JE (2015) Kantian optimization: a microfoundation for cooperation. J Public Econ 127: 45–57
Rubio SJ (2011) On capturing rent from non-renewable resource international monopoly: prices versus quantities. Dyn Games Appl 1(4):558–580
Rubio SJ, Estriche L (2001) Strategic Pigouvian taxation, stock externalities and polluting non-renewable resources. J Public Econ 79(2):297–313
Russell B (1945) A history of western philosophy. George Allen and Unwin, London
Salant S (1976) Exhaustible resources and industrial structure: a Nash-Cournot approach to the World oil market. J Polit Econ 84:1079–94
Salant S, Switzer S, Reynolds R (1983) Losses from horizontal merger: the effects of an exogenous change in industry structure on Cournot Nash equilibrium. Q J Econ 98(2):185–199
Salo S, Tahvonen O (2001) Oligopoly equilibria in nonrenewable resource markets. J Econ Dyn Control 25:671–702
Sethi R, Somanathan E (1996) The evolution of social norms in common property resource use. Am Econ Rev 86(4):766–88
Sidgwick H (1874) The method of ethics, 7th edn. 1907. Macmillan, London. Reprinted by the Hackett Publishing Company (1981)
Sinn H-W (1984) Common property resources, storage facilities, and ownership structure: a Cournot model of the oil market. Economica 51:235–52
Smith A (1776) An inquiry into the nature and causes of the wealth of nations. In: Cannan E (ed) The modern library, 1937. Random House, New York
Smith A (1790) The theory of moral sentiments. Cambridge University Press, Cambridge, UK. Republished in the series Cambridge texts in the history of philosophy, 2002. Edited by K. Haakonssen
Stiglitz JE (1976) Monopoly power and the rate of extraction of exhaustible resources. Am Econ Rev 66:655–661
Sundaram RK (1989) Perfect equilibria in non-randomized strategies in a class of symmetric games. J Econ Theory 47:153–177

Tahvonen O (1996) Trade with polluting non-renewable resources. J Environ Econ Manag 30:1–17
Tornell A (1997) Economic growth and decline with endogenous property rights. J Econ Growth 2(3):219–257
Tornell A, Lane P (1999) The voracity effect. Am Econ Rev 89:22–46
Ulph A, Folie GM (1980) Exhaustible resources and Cartels: an intertemporal Nash-Cournot model. Canad J Econ 13:645–58
van der Meijden KR, Withagen C (2016) Double limit pricing. In: Paper presented at the 11th Tinbergen Institute conference on climate change, Amsterdam
van der Ploeg R (2010) Rapacious resource depletion, excessive investment and insecure property rights. CESifo working paper no 2981, University of Munich
Veblen TB (1899) The theory of the Leisure class: an economic study of institutions. Modern Library, New York
Watkins GC (1977) Conservation and economic efficiency: Alberta oil proration. J Environ Econ Manag 4:40–56
Wirl F (1994) Pigouvian taxation of energy for flow and stock externalities and strategic non-competitive energy pricing. J Environ Econ Manag 26:1–18
World Trade Organization (2010) World trade report, 2010, Washington, DC

Dynamic Games of International Pollution Control: A Selective Review

16

Aart de Zeeuw

Contents

Abstract

A differential game is the natural framework of analysis for many problems in environmental economics. This chapter focuses on the game of international pollution control and more specifically on the game of climate change with one global stock of pollutants. The chapter has two main themes. First, the different noncooperative Nash equilibria (open loop, feedback, linear, nonlinear) are derived. In order to assess efficiency, the steady states are compared with the steady state of the full-cooperative outcome. The open-loop Nash equilibrium is better than the linear feedback Nash equilibrium, but a nonlinear feedback Nash equilibrium exists that is better than the open-loop Nash equilibrium. Second,

A. de Zeeuw (✉)
The Beijer Institute of Ecological Economics, The Royal Swedish Academy of Sciences, Stockholm, Sweden
e-mail: A.J.deZeeuw@uvt.nl

T. Başar, G. Zaccour (eds.), *Handbook of Dynamic Game Theory*,
https://doi.org/10.1007/978-3-319-44374-4_16

the stability of international environmental agreements (or partial-cooperation Nash equilibria) is investigated, from different angles. The result in the static models that the membership game leads to a small stable coalition is confirmed in a dynamic model with an open-loop Nash equilibrium. The result that in an asymmetric situation transfers exist that sustain full cooperation under the threat that the coalition falls apart in case of deviations is extended to the dynamic context. The result in the static model that farsighted stability leads to a set of stable coalitions does not hold in the dynamic context if detection of a deviation takes time and climate damage is relatively important.

Keywords
Differential games · Multiple Nash equilibria · International pollution control · Climate change · Partial cooperation · International environmental agreements · Stability · Non-cooperative games · Cooperative games · Evolutionary games

JEL classification: C61; C72; Q20

1 Introduction

Differential game theory extends optimal control theory to situations with more than one decision maker. The controls of each decision maker affect the development of the state of the system, given by a set of differential equations. The objectives of the decision makers are integrals of functions that depend on the state of the system, so that the controls of each decision maker affect the objectives of the other decision makers which turns the problem into a game. A differential game is a natural framework of analysis for many problems in environmental economics. Usually these problems extend over time and have externalities in the sense that the actions of one agent affect welfare of the other agents. For example, emissions of all agents accumulate into a pollution stock, and this pollution stock is damaging to all agents.

The problem of international pollution control is a very good example of an application of differential games. Pollution crosses national borders and affects welfare in other countries. However, there is no world government that can correct these externalities in the usual way, with a tax or with some other policy. At the international level, countries play a game and cooperation with the purpose to internalize the externalities is voluntary. This chapter focuses on the problem of international pollution control and more specifically on the problem of climate change. Greenhouse gas emissions from all countries accumulate into a global stock of atmospheric carbon (the state of the system) that leads to climate change which is damaging to all countries.

In this differential game, time enters indirectly through state and controls but directly only through discounting, so that the optimality conditions are stationary in current values. The chapter will start with a short summary of the Pontryagin conditions and the Hamilton-Jacobi-Bellman equations for the basic infinite-horizon continuous-time differential game. The theory provides two remarkable results. First, the resulting Nash equilibria differ. They are called the open-loop and the feedback Nash equilibrium, respectively, referring to the information structure that is implicitly assumed (Başar et al. 1982). Most applications of differential games focus on comparing the open-loop and feedback Nash equilibria (Dockner et al. 2000). Second, even the simple linear-quadratic structure allows for a multiplicity of (nonlinear) feedback Nash equilibria. This is shown for the game of international pollution control with one global stock of pollutants. In case of climate change, this is the stock of atmospheric carbon.

Full cooperation is better for the world as a whole, but incentives to deviate arise, and therefore a noncooperative Nash equilibrium will result in which no country has an incentive to deviate. In case of multiple Nash equilibria, the question arises which one is better. It is shown that the steady-state stock of atmospheric carbon in the open-loop Nash equilibrium is closer to the full-cooperative one than in the linear feedback Nash equilibrium. However, a nonlinear feedback Nash equilibrium exists with the opposite result. In the linear case, the emission policies are negative functions of the stock so that an increase in emissions is partly offset by a future decrease in emissions by the other countries. It follows that with feedback controls, emissions in the Nash equilibrium are higher. However, in the nonlinear case, the emission policies are, in a neighborhood of the steady state, a positive function of the stock so that an increase in emissions is met by even higher future emissions. This threat keeps emissions down. Observations on the state of the system are only beneficial if the countries can coordinate on a nonlinear Nash equilibrium.

The second part of the chapter focuses on international environmental agreements. An agreement is difficult to maintain because of the incentives to deviate. The question is whether these incentives can be suppressed. The first idea is that the countries remaining in the coalition do not take the country that deviates into account anymore so that the deviator loses some benefits of cooperation. This definition of stability leads to small stable agreements. The chapter shows this by analyzing a partial-cooperation open-loop Nash equilibrium between the coalition of countries and the individual outsiders. A second idea is that a deviation triggers more deviations which yields a set of stable agreements. This is called farsighted stability and gives the possibility of a large stable agreement. However, the chapter shows that in a dynamic context, where detection of a deviation takes time and where climate damage is relatively more important than the costs of emission reductions, only the small stable agreement remains. The third idea is that after a deviation, the remaining coalition falls apart completely. It is clear that in the symmetric case, this threat is sufficiently strong to prevent deviations. However, in the asymmetric case, transfers between the coalition members are needed. The chapter shows how these transfers develop over time in a dynamic context. The basic models used to analyze these questions are not all infinite-horizon continuous-time differential games, with the global stock of pollutants as the state of the system. In some of these models, the horizon is finite, the time is discrete, or the state is the level of excess emissions, but all the models are dynamic games with a state transition.

Cooperative game theory not only provides transfers or allocation rules to achieve stability but also to satisfy certain axioms such as fairness. The most common allocation rule is the Shapley value. The question is whether this allocation rule is time consistent in a dynamic context. The chapter shows that an allocation over time exists such that reconsidering the allocation rule at some point in time will not change it. Finally, an evolutionary game approach to the issue of stability is considered. Based on a simple partial-cooperation differential game, with punishments imposed on the outsiders, the chapter shows that replicator dynamics leads to full cooperation for sufficiently high punishments and a sufficiently high initial level of cooperation.

The chapter only focuses on a number of analyses that fit in a coherent story. This also allows to provide some details. A wider range of applications in this area can be found in other surveys on dynamic games and pollution control (e.g., Jorgensen et al. 2010; Calvo and Rubio 2013; Long 2012). Section 2 gives a short summary of the main techniques of differential games. Section 3 analyzes the game of international pollution control. Section 4 introduces partial cooperation and discusses different angles on the stability of international environmental agreements. Section 5 concludes.

2 Differential Games

It was remarkable to find that the equivalent of Bellman's principle of optimality does not generally hold in differential games (Starr and Ho 1969). It follows that the Nash equilibrium in strategies that depend only on time (and the initial state of the system) usually differs from the Nash equilibrium that is found by dynamic programming. The first Nash equilibrium is called the open-loop Nash equilibrium, because information on the state is not used or not available. The second Nash equilibrium is called the feedback Nash or Markov perfect equilibrium: by construction, using dynamic programming, the controls depend on the state of the system, and backward induction yields Markov perfectness. The open-loop Nash equilibrium results from using Pontryagin's maximum principle under the standard assumption that the strategies depend only on time (and the initial state of the system). It is also possible to find a feedback Nash equilibrium with Pontryagin's maximum principle, assuming that the strategies in some way depend on the state of the system (e.g., Long 2006; Tornell and Velasco 1992), but this is usually less convenient than applying dynamic programming. In general, many more Nash equilibria exist, for example, by considering extended information structures (Başar et al. 1982), but most of the applications are restricted to comparing open-loop and feedback Nash equilibria (Dockner et al. 2000). The same applies for applications in environmental economics. Before we discuss some of these applications, we will first summarize the formal model that will be used in the sequel.

2.1 Formal Model

An important class of differential games is given by

$$\max_{u_i(.)} W_i = \int_0^{\infty} e^{-rt} F_i[x(t), u_i(t)]dt, i = 1, 2, \ldots, n, \tag{16.1}$$

subject to

$$\dot{x}(t) = f[x(t), u_1(t), u_2(t), \ldots, u_n(t)], x(0) = x_0, \tag{16.2}$$

where i indexes the n players, x denotes the state of the system, u the controls, r the discount rate, W the total welfare, F the welfare at time t, and f the state transition. Note that the players only interact through the state dynamics. The problem has an infinite horizon, and the welfare function and the state transition do not explicitly depend on time, except for the discount rate. This implies that the problem can be cast into a stationary problem.

In the open-loop Nash equilibrium, the controls only depend on time: $u_i(t)$. This implies that for each player i, an optimal control problem has to be solved using Pontryagin's maximum principle, with the strategies of the other players as exogenous inputs. This results in an optimal control strategy for player i as a function of time and the strategies of the other players. This is in fact the rational reaction or best response of player i. The open-loop Nash equilibrium simply requires consistency of these best responses. Pontryagin's maximum principle yields a necessary condition in terms of a differential equation in the co-state λ_i. If the optimal solution for player i can be characterized by the set of differential equations in x and λ_i, then the open-loop Nash equilibrium can be characterized by the set of differential equations in x and $\lambda_1, \lambda_2, \ldots, \lambda_n$. This is usually the best way to find the open-loop Nash equilibrium. The necessary conditions for player i in terms of the current-value Hamiltonian function

$$H_i(x, u_i, \lambda_i) = F_i(x, u_i) + \lambda_i f[x, u_1(t), .., u_i, .., u_n(t)] \tag{16.3}$$

are that the optimal $u_i^*(t)$ maximizes H_i and that the state x and the co-state λ_i satisfy the set of differential equations

$$\dot{x}(t) = f[x(t), u_1(t), .., u_i^*(t), .., u_n(t)], x(0) = x_0, \tag{16.4}$$

$$\dot{\lambda}_i(t) = r\lambda_i(t) - H_{ix}[x(t), u_i^*(t), \lambda_i(t)], \tag{16.5}$$

with a transversality condition on λ_i. Note that the actions of the other players $u_j(t)$, $j \neq i$ are exogenous to player i. If sufficiency conditions are satisfied and if $u_i^*(t)$ can be explicitly solved from the first-order conditions of optimization, the open-loop Nash equilibrium can be found by solving the set of differential equations in x and $\lambda_1, \lambda_2, \ldots, \lambda_n$ given by

$$\dot{x}(t) = f[x(t), u_1^*(t), u_2^*(t), .., u_n^*(t)], x(0) = x_0, \tag{16.6}$$

$$\dot{\lambda}_i(t) = r\lambda_i(t) - H_{ix}[x(t), u_i^*(t), \lambda_i(t)], i = 1, 2, \ldots, n, \tag{16.7}$$

with transversality conditions on $\lambda_1, \lambda_2, \ldots, \lambda_n$.

In the feedback Nash equilibrium, the controls depend on the current state of the system, and since the problem is basically stationary, they do not depend explicitly on time: $u_i(x)$. The Hamilton-Jacobi-Bellman equations in the current-value functions V_i are given by

$$rV_i(x) = \max_{u_i}\{F_i(x, u_i) + V_i'(x) f[x, u_1(x), .., u_i, .., u_n(x)]\}, i = 1, 2, \ldots, n. \tag{16.8}$$

If sufficiency conditions are satisfied and if $u_i^*(x)$ can be explicitly solved from the first-order conditions of optimization, the feedback Nash equilibrium can be found by solving the set of equations in the current-value functions V_i given by

$$rV_i(x) = F_i(x, u_i^*(x)) + V_i'(x) f[x, u_1^*(x), u_2^*(x), \ldots, u_n^*(x)], i = 1, 2, \ldots, n. \tag{16.9}$$

An important question is whether one or the other Nash equilibrium yields higher welfare to the players. The benchmark for answering this question is the full-cooperative outcome in which the countries maximize their joint welfare $\sum W_i$. This is a standard optimal control problem that can be solved using either Pontryagin's maximum principle or dynamic programming.

We will now turn to typical applications in environmental economics.

3 International Pollution Control

A typical situation in environmental economics that requires a differential game analysis is as follows. The players are countries. Economic activities in these countries generate emissions of some sort. These emissions accumulate into a stock of pollutants that is damaging. If these emissions cross national borders or if these emissions accumulate into a global stock of pollutants, stock externalities between the players occur. An example of the first situation is the acid rain game (Kaitala et al. 1992; Mäler et al. 1998). Certain industries and traffic generate emissions of sulfur dioxide and nitrogen oxides, and the wet and dry depositions of these substances cause acidification of soils, which is a dynamic process. Soils can assimilate this pollution up to what is called the critical load, but the depositions above this critical load lower the pH level of the soils which damages their productivity. Reducing emissions is costly so that a trade-off occurs between the costs of emission reductions and damage of acidification. Winds take these substances across borders so that emissions in one country may lead to depositions in other countries. A differential game results with a transportation matrix in the linear state equation that indicates which fractions of the emissions in each country are transported to the other countries. In steady state, the depositions are equal to the critical loads, but the open-loop Nash equilibrium, the feedback Nash equilibrium, and the full-cooperative outcome differ in their speed of convergence to the steady state and in their level of acidification. General analytical conclusions are hard to derive, but availability of data in Europe on the transportation matrix and the costs of emission reductions allow for some conclusions in terms of the possible gains of cooperation (Mäler et al. 1998).

The second situation in which emissions accumulate into a global stock of pollutants allows for more general analytical conclusions. Climate change is the obvious example. Greenhouse gas emissions accumulate into a stock of atmospheric carbon that causes climate change. The rise in temperature has a number of damaging effects. Melting of ice caps leads to sea level rise which causes flooding or requires expensive protection measures. Precipitation patterns will change which will lead to desertification in some parts of the world and extensive rainfall in other parts. An increase in extreme weather is to be expected with more severe storms and draughts. In general, living conditions will change as well as conditions for agriculture and other production activities. Greenhouse gas emissions E are a by-product of production, and the benefits of production can therefore be modeled as $B(E)$, but the stock of atmospheric carbon S yields costs $D(S)$. This leads to the following differential game:

$$\max_{E_i(.)} W_i = \int_0^\infty e^{-rt}[B(E_i(t)) - D(S(t))]dt, i = 1, 2, \ldots, n, \tag{16.10}$$

with

$$B(E) = \beta E - \frac{1}{2}E^2, D(S) = \frac{1}{2}\gamma S^2, \tag{16.11}$$

subject to

$$\dot{S}(t) = \sum_{i=1}^{n} E_i(t) - \delta S(t), S(0) = S_0, \tag{16.12}$$

where i indexes the n countries, W denotes total welfare, r the discount rate, δ the decay rate of carbon, and β and γ are parameters of the benefit and cost functions. Note that in this model, the countries are assumed to have identical benefit and cost functions. When the costs D are ignored, emission levels E are at the business-as-usual level β yielding maximal benefits, but in the optimum lower emission levels with lower benefits B are traded off against lower costs D. It is standard to assume decreasing marginal benefits and increasing marginal costs, and the quadratic functional forms with the linear state equation (16.12) are convenient for the analysis.

Climatologists, however, predict that an important part of climate change damage will be caused by tipping points in the climate system (Lenton and Ciscar 2013). At some point the system may shift to another domain of attraction with substantial costs to the economy. The tipping point is uncertain, and therefore tipping points are usually labeled as large, abrupt, and persistent changes. The damage function $D(S) = \frac{1}{2}\gamma S^2$ does not reflect this aspect, but there is a way in which this can be handled without losing the linear-quadratic structure. Tipping can be modeled as an uncertain point in time where the damage function switches from $D_1(S) = \frac{1}{2}\gamma_1 S^2$ to $D_2(S) = \frac{1}{2}\gamma_2 S^2$ with $\gamma_2 > \gamma_1$. The problem can now be split into problems before the event and after the event and can be solved backward in time, so that the linear-quadratic structure is preserved. This has been done for the optimal control problem (de Zeeuw and Zemel 2012), but the differential game is still open for further research. In the optimal control solution, the main result is that optimal policy becomes precautionary in the sense that the emissions have to decrease when potential tipping is taken into account.

Climate tipping can also be modeled as a shock to total factor productivity in the Ramsey growth model (van der Ploeg and de Zeeuw 2016). Because the stock of atmospheric carbon increases the probability of tipping, externalities occur between the countries. Furthermore, countries may differ in their vulnerability to climate change (i.e., their shock to total factor productivity) and in their stage of development (i.e., their initial capital stock). Typically, the "South" is more vulnerable to climate change and starts at a lower stage of development than the "North". The open-loop Nash equilibrium of the resulting asymmetric differential game has been derived, but the feedback Nash equilibrium is still open for further research. The main conclusion is that taxes on emissions in the full-cooperative outcome differ in two respects, as compared to the open-loop Nash equilibrium. In the long run, taxes in the North and the South converge to high and similar levels when they cooperate, whereas in the absence of cooperation, the tax in the South is lower and in the North much lower. Furthermore, initially the tax is high in the North and low in the South when they cooperate, so that the South can catch up, whereas in the absence of cooperation, the taxes are both low. In this chapter we will only focus on the open-loop and feedback Nash equilibria of the linear-quadratic differential game (16.10), (16.11), and (16.12) and compare the results (see Dockner and Long 1993; Hoel 1993; Long 1992; van der Ploeg and de Zeeuw 1992).

3.1 Open-Loop Nash Equilibrium

Using Pontryagin's maximum principle, the current-value Hamiltonian functions become

$$H_i(S, E_i, t, \lambda_i) = \beta E_i - \frac{1}{2}E_i^2 - \frac{1}{2}\gamma S^2 + \lambda_i(E_i + \sum_{j \neq i}^{n} E_j(t) - \delta S) \qquad (16.13)$$

and since sufficiency conditions are satisfied, the open-loop Nash equilibrium conditions become

$$E_i(t) = \beta + \lambda_i(t), i = 1, 2, \ldots, n, \tag{16.14}$$

$$\dot{S}(t) = \sum_{i=1}^{n} E_i(t) - \delta S(t), S(0) = S_0, \tag{16.15}$$

$$\dot{\lambda}_i(t) = (r + \delta)\lambda_i(t) + \gamma S(t), i = 1, 2, \ldots, n, \tag{16.16}$$

with transversality conditions on $\lambda_1, \lambda_2, \ldots, \lambda_n$. The symmetric open-loop Nash equilibrium can therefore be characterized by the set of differential equations

$$\dot{S}_{OL}(t) = n(\beta + \lambda_{OL}(t)) - \delta S_{OL}(t), S_{OL}(0) = S_0, \tag{16.17}$$

$$\dot{\lambda}_{OL}(t) = (r + \delta)\lambda_{OL}(t) + \gamma S_{OL}(t), \tag{16.18}$$

with a transversality condition on λ_{OL}, where OL denotes open loop. This yields a standard phase diagram in the state/co-state plane for an optimal control problem, with a stable manifold, and the saddle-point-stable steady state is given by

$$S_{OL}^* = \frac{n\beta(r + \delta)}{\delta(r + \delta) + n\gamma}. \tag{16.19}$$

The negative of the shadow value $-\lambda_{OL}$ can be interpreted as the tax on emissions that is required in each country to implement the open-loop Nash equilibrium. Note that this tax only internalizes the externalities within the countries but not the transboundary externalities. This would require a higher tax that can be found from the full-cooperative outcome of the game.

In the full-cooperative outcome, the countries maximize their joint welfare $\sum W_i$ which is a standard optimal control problem. Using Pontryagin's maximum principle, the current-value Hamiltonian function becomes

$$H(S, E_1, E_2, \ldots, E_n, \lambda) = \sum_{i=1}^{n}(\beta E_i - \frac{1}{2}E_i^2) - \frac{1}{2}\gamma n S^2 + \lambda\left(\sum_{i=1}^{n} E_i - \delta S\right) \tag{16.20}$$

and since sufficiency conditions are satisfied, the optimality conditions become

$$E_i(t) = \beta + \lambda(t), i = 1, 2, \ldots, n, \tag{16.21}$$

$$\dot{S}(t) = \sum_{i=1}^{n} E_i(t) - \delta S(t), S(0) = S_0, \tag{16.22}$$

$$\dot{\lambda}(t) = (r + \delta)\lambda(t) + n\gamma S(t), \tag{16.23}$$

with a transversality condition on λ. The full-cooperative outcome can therefore be characterized by the set of differential equations

$$\dot{S}_C(t) = n(\beta + \lambda_C(t)) - \delta S_C(t), S_C(0) = S_0, \tag{16.24}$$

$$\dot{\lambda}_C(t) = (r + \delta)\lambda_C(t) + n\gamma S_C(t), \tag{16.25}$$

with a transversality condition on λ_C, where C denotes cooperative. This yields a standard phase diagram in the state/co-state plane for an optimal control problem, with a stable manifold, and the saddle-point-stable steady state is given by

$$S_C^* = \frac{n\beta(r + \delta)}{\delta(r + \delta) + n^2\gamma} < S_{OL}^*. \tag{16.26}$$

The negative of the shadow value $-\lambda_C$ can be interpreted again as the tax on emissions that is required in each country to implement the full-cooperative outcome and it is easy to see that this tax is indeed higher than the tax in the open-loop Nash equilibrium, because now the transboundary externalities are internalized as well. The steady state of the full-cooperative outcome S_C^* is lower than the steady state of the open-loop Nash equilibrium S_{OL}^*, as is to be expected. The interesting case arises when we include the steady state of the feedback Nash equilibrium in the next section.

3.2 Feedback Nash Equilibrium

The Hamilton-Jacobi-Bellman equations in the current-value functions $V_i(S)$ become

$$rV_i(S) = \max_{E_i} \left\{ \beta E_i - \frac{1}{2}E_i^2 - \frac{1}{2}\gamma S^2 + V_i'(S)\left(E_i + \sum_{j \neq i}^{n} E_j(S) - \delta S\right)\right\},$$
$$i = 1, 2, \ldots, n, \tag{16.27}$$

with first-order conditions

$$E_i^*(S) = \beta + V_i'(S), i = 1, 2, \ldots, n. \tag{16.28}$$

Since sufficiency conditions are satisfied, the symmetric feedback Nash equilibrium can be found by solving the following differential equation in $V = V_i$, $i = 1, 2, .., n$:

$$rV(S) = \beta(\beta + V'(S)) - \frac{1}{2}(\beta + V'(S))^2 - \frac{1}{2}\gamma S^2 + V'(S)[n(\beta + V'(S)) - \delta S]. \tag{16.29}$$

The linear-quadratic structure of the problem suggests that the value function is quadratic (but we come back to this issue below). Therefore, the usual way to solve this differential equation is to assume that the current-value function V has the general quadratic form

$$V(S) = \sigma_0 - \sigma_1 S - \frac{1}{2}\sigma_2 S^2, \sigma_2 > 0, \tag{16.30}$$

so that a quadratic equation in the state S results. Since this equation has to hold for all S, the coefficients of S^2 and S on the left-hand side and the right-hand side have to be equal. It follows that

$$\sigma_2 = \frac{-(r+2\delta) + \sqrt{(r+2\delta)^2 + 4(2n-1)\gamma}}{2(2n-1)}, \tag{16.31}$$

$$\sigma_1 = \frac{n\beta\sigma_2}{(r+\delta) + (2n-1)\sigma_2}. \tag{16.32}$$

The feedback Nash equilibrium becomes

$$E_i^*(S) = \beta - \sigma_1 - \sigma_2 S, i = 1, 2, \ldots, n, \tag{16.33}$$

and the controlled state transition becomes

$$\dot{S}(t) = n(\beta - \sigma_1 - \sigma_2 S(t)) - \delta S(t), \tag{16.34}$$

which is stable and yields the steady state

$$S_{FB}^* = \frac{n(\beta - \sigma_1)}{\delta + n\sigma_2}, \tag{16.35}$$

where FB denotes feedback.

It is tedious but straightforward to show that

$$S_C^* < S_{OL}^* < S_{FB}^*. \tag{16.36}$$

This is interesting because it implies that in the feedback Nash equilibrium, the countries approach a higher steady-state stock of atmospheric carbon than in the open-loop Nash equilibrium. It turns out that for the feedback information structure, the noncooperative outcome is worse in this respect than for the open-loop information structure. The intuition is as follows. A country argues that when it will increase its emissions, this will increase the stock of atmospheric carbon and this will induce the other countries to lower their emissions, so that part of the increase in emissions will be offset by the other countries. Each country argues the same way so that in the Nash equilibrium emissions will be higher than in case the stock of atmospheric carbon is not observed. Since the feedback model (in which the countries observe the state of the system and are not committed to future actions) is the more realistic model, the "tragedy of the commons" (Hardin 1968) is more severe than one would think when the open-loop model is used. To put it differently, the possible gains of cooperation prove to be higher when the more realistic feedback model is used as the noncooperative model. However, this is not the whole story. Based on a paper by Tsutsui and Mino (1990) that extends one of the first differential game applications in economics by Fershtman and Kamien (1987) on dynamic duopolistic competition, Dockner and Long (1993) show that also nonlinear feedback Nash equilibria (with non-quadratic current-value functions) for this problem exist which lead to the opposite result. This will be the topic of the next section.

3.3 Nonlinear Feedback Nash Equilibria

Recall that the symmetric feedback Nash equilibrium (16.28) is given by

$$E_i^*(S) = \beta + V'(S) := h(S), i = 1, 2, \ldots, n, \tag{16.37}$$

where h denotes the feedback equilibrium control. It follows that the Hamilton-Jacobi-Bellman equation (16.27) can be written as

$$rV(S) = \beta h(S) - \frac{1}{2}(h(S))^2 - \frac{1}{2}\gamma S^2 + [h(S) - \beta][nh(S) - \delta S]. \tag{16.38}$$

Assuming that h is differentiable, differentiating this equation with respect to S and again substituting $V'(S) = h(S) - \beta$ yields an ordinary differential equation in the feedback equilibrium control h that is given by

$$[(2n-1)h(S) + (1-n)\beta - \delta S]h'(S) = (r+\delta)h(S) + \gamma S - \beta(r+\delta). \tag{16.39}$$

The linear feedback Nash equilibrium (16.33) with steady state (16.35) is a solution of this differential equation, but it is not the only one. The differential equation may have multiple solutions because the boundary condition is not specified. The steady-state condition

$$h(S^*) = \frac{\delta S^*}{n} \tag{16.40}$$

can serve as a boundary condition, but the steady state S^* is not predetermined and can take different values. One can also say that the multiplicity of nonlinear feedback Nash equilibria results from the indeterminacy of the steady state in differential games. The solutions of the differential equation (16.39) in the feedback equilibrium control h must lead to a stable system where the state S converges to the steady state S^*. Dockner and Long (1993) show that for $n = 2$, the set of stable solutions is represented by a set of hyperbolas in the (S, h)-plane that cut the steady-state line $\frac{1}{2}\delta S$ in the interval

$$\frac{2\beta(2r+\delta)}{\delta(2r+\delta)+4\gamma} \leq S^* < \frac{2\beta}{\delta}. \tag{16.41}$$

Rubio and Casino (2002) indicate that this result has to be modified in the sense that it does not hold for all initial states S_0 but it holds for intervals of initial states S_0 above the steady state S^*. The upper edge of the interval (16.41) represents the situation where both countries ignore the costs of climate change and choose $E = \beta$. The endpoint to the left of the interval (16.41) is the lowest steady state that can be achieved with a feedback Nash equilibrium. We will call this the "best feedback Nash equilibrium" and denote the steady state by S^*_{BFB}. In this equilibrium the hyperbola $h(S)$ is tangent to the steady-state line, so that $h(S^*_{BFB}) = \frac{1}{2}\delta S^*_{BFB}$ and $h'(S^*_{BFB}) = \frac{1}{2}\delta$.

Using (16.19) and (16.26), it is easy to show that

$$S^*_C = \frac{2\beta(r+\delta)}{\delta(r+\delta)+4\gamma} < S^*_{BFB} = \frac{2\beta(2r+\delta)}{\delta(2r+\delta)+4\gamma} < S^*_{OL} = \frac{2\beta(r+\delta)}{\delta(r+\delta)+2\gamma}. \tag{16.42}$$

This means that if we allow for nonlinear equilibria in this linear-quadratic framework, the feedback Nash equilibrium can be better (in terms of steady states) than the open-loop Nash equilibrium, opposite to what we found for the linear case in the previous section. The reason is that close to S^*_{BFB}, the feedback equilibrium controls h are increasing because $h'(S^*_{BFB}) = \frac{1}{2}\delta$. This means that now each country argues that when it will increase its emissions, this will increase the stock of atmospheric carbon and this will induce the other countries to increase their emissions as well. By coordinating on the best feedback Nash equilibrium, the countries build in some sort of threat which keeps emissions and thus the stock of atmospheric carbon down. Moreover, if the discount rate r converges to zero, the steady state of the best feedback Nash equilibrium S^*_{BFB} converges to the steady state S^*_C of the cooperative outcome. This result can be interpreted as a folk theorem in this differential game, although a general folk theorem for differential games is not available (Dockner and Long 1993; Rowat 2007).

The last result does not necessarily mean that welfare in the best feedback Nash equilibrium also converges to welfare in the cooperative outcome, if the discount rate r converges to zero. The analysis does not show how the paths toward the steady states in both outcomes compare. Kossioris et al. (2008) use the same approach to derive the best feedback Nash equilibrium for a differential game with a nonlinear state transition (representing the eutrophication of a lake). In this case the ordinary differential equation in the feedback equilibrium control h becomes more complicated (i.e., an Abel differential equation of the second kind). It can only be solved with numerical methods. They show that in this case, the steady state of the best feedback Nash equilibrium not always converges to the steady state of the cooperative outcome, if the discount rate r converges to zero. Moreover, welfare in the best feedback Nash equilibrium is generally worse than in the cooperative outcome. Finally, this approach has only been developed for one-dimensional systems, and we have to wait and see how it works out in higher dimensions. All these issues are open for further research.

4 International Environmental Agreements

The game of international pollution control with a global stock of pollutants is a so-called prisoners' dilemma. The full-cooperative outcome yields higher welfare than the noncooperative outcome, but each country has an incentive to deviate and to free ride on the efforts of the other countries. In that sense, the full-cooperative outcome is not stable. However, if the group of remaining countries adjust their emission levels when a country deviates, the deviating country also loses welfare because the externalities between this country and the group are not taken into account anymore. The interesting question arises whether some level of cooperation can be stable in the sense that the free-rider benefits are outweighed by these losses. In the context of cartel theory, d'Aspremont et al. (1983) have introduced the concept of cartel stability. Carraro et al. (1993) and Barrett (1994) have developed it further in the context of international environmental agreements. An agreement is stable if it is both internally and externally stable. Internal stability means that a member of the agreement does not have an incentive to leave the agreement, and external stability means that an outsider does not have an incentive to join the agreement. This can be modeled as a two-stage game. In stage one the countries choose to be a member or not, and in stage two, the coalition and the outsiders decide noncooperatively on their levels of emissions. The basic result of this approach is that the size of the stable coalition is small. Rubio and Casino (2005) show this for the dynamic model with a global stock of pollutants, using an extension of the differential game in the previous sections. This will be the topic of the next section.

4.1 Stable Partial Cooperation

In the first stage of the game, the coalition of size $k \leqq n$ is formed, and in the second stage of the game, this coalition and the $n-k$ outsiders play the differential game (16.10), (16.11), and (16.12) where the objective of the coalition is given by

$$\max_{E_1(.),\ldots,E_k(.)} \sum_{i=1}^{k} W_i . \tag{16.43}$$

The first stage of the game leads to two types of players in the second stage of the game: members of the coalition and outsiders. It is easy to see that the open-loop partial-cooperation Nash equilibrium in the second stage

$$E_i(t) = \beta + \lambda^m(t), i = 1, \ldots, k; E_i(t) = \beta + \lambda^o(t), i = k+1, \ldots, n, \quad (16.44)$$

where m denotes member of the coalition and o denotes outsider, can be characterized by the set of differential equations

$$\dot{S}(t) = k(\beta + \lambda^m(t)) + (n-k)(\beta + \lambda^o(t)) - \delta S(t), S(0) = S_0, \quad (16.45)$$

$$\dot{\lambda}^m(t) = (r+\delta)\lambda^m(t) + k\gamma S(t), \quad (16.46)$$

$$\dot{\lambda}^o(t) = (r+\delta)\lambda^o(t) + \gamma S(t), \quad (16.47)$$

with transversality condition on λ^m and λ^o. This yields a standard phase diagram in the state/co-state plane for an optimal control problem, with a stable manifold, and the saddle-point-stable steady state is given by

$$S^* = \frac{n\beta(r+\delta)}{\delta(r+\delta) + (k^2 + n - k)\gamma}. \quad (16.48)$$

For $k = 1$ the steady state S^* is equal to the steady state S^*_{OL} of the open-loop Nash equilibrium, given by (16.19), and for $k = n$ it is equal to the steady state S^*_C of the full-cooperative outcome, given by (16.26).

Since there are two types of players, there are also two welfare levels, W^m for a member of the coalition and W^o for an outsider. An outsider is better off than a member of the coalition, i.e., $W^o > W^m$, because outsiders emit more and are confronted with the same global level of pollution. However, if a country in stage one considers to stay out of the coalition, it should not compare these welfare levels but the welfare level of a member of the coalition of size k, i.e., $W^m(k)$, with the welfare level of an outsider to a coalition of size $k-1$, i.e., $W^o(k-1)$. This is in fact the concept of internal stability. In the same way, the concept of external stability can be formalized, so that the conditions of internal and external stability are given by

$$W^o(k-1) \leqq W^m(k), k \geqq 1, \quad (16.49)$$

$$W^m(k+1) \leqq W^o(k), k \leqq n-1. \quad (16.50)$$

The question is for which size k these conditions hold. It is not possible to check these conditions analytically, but it can be shown numerically that the size k^* of the stable coalition is equal to 2, regardless of the total number of countries n (Rubio and Casino 2005). This confirms the results in the static literature with a flow pollutant. It shows that the free-rider incentive is stronger than the incentive to cooperate.

In this model it is assumed that the membership decision in stage one is taken once and for all. Membership is fixed in the differential game in stage two. The question is what happens if membership can change over time. Rubio and Ulph (2007) construct an infinite-horizon discrete-time model, with a stock of atmospheric carbon, in which the membership decision is taken at each point in time. A difficulty is that countries do not know whether they will be members of the coalition or outsiders in the next periods, so that they do not know their future value functions. Therefore, it is assumed that each country takes the average future value function into account which allows to set up the dynamic-programming equations for each type of country in the current period. Then the two-stage game can be solved in each time period, for each level of the stock. Rubio and Ulph (2007) show numerically that on the path towards the steady state, an increasing stock of atmospheric carbon is accompanied by a decreasing size of the stable coalition. We will now switch to alternative stability concepts for international environmental agreements.

4.2 An Alternative Stability Concept

The approach with internal/external stability was challenged by Chander and Tulkens (1995) who define a stable coalition differently. They assume that if a country or a group of countries deviates, the group of remaining countries falls apart and plays as individual countries noncooperatively against the deviator. This can be seen as a threat that may prevent deviations. A coalition is stable if deviations of individual countries or groups of countries are prevented in this way. This is the idea of the core in cooperative game theory, and in fact the γ-core concept is applied here. In the symmetric case, it is obvious that the grand coalition is stable, because full-cooperative welfare is higher than noncooperative welfare. However, in the asymmetric case, transfers between the coalition members are needed in order to achieve stability of the grand coalition in this respect.

In a dynamic game, with a stock of atmospheric carbon, the question is how these transfers develop over time. In order to investigate this, Germain et al. (2003) use the asymmetric finite-horizon discrete-time version of the differential game (16.10), (16.11), and (16.12) which is given by

$$\min_{E_i(.)} \sum_{t=1}^{T} d^t [C_i(E_i(t)) + D_i(S(t))], i = 1, 2, \ldots, n, \tag{16.51}$$

subject to

$$S(t) = (1-\delta)S(t-1) + \sum_{i=1}^{n} E_i(t), S(0) = S_0, \tag{16.52}$$

where $C = B(\beta) - B(E)$ denotes the costs of emission reductions, T the time horizon, and d the discount factor.

In the final period T, the level of the stock $S(T-1)$ is given, and a static game is played. The grand coalition minimizes the sum of the total costs over all n countries. Suppose that the minimum is realized by $E_i^C(T)$ leading to $S^C(T)$. Furthermore, suppose that the Nash equilibrium is unique and is given by $E_i^N(T)$

leading to $S^N(T)$. It follows that the full-cooperative and Nash-equilibrium total costs, TC_i^C and TC_i^N, for each country i are given by

$$TC_i^C(S(T-1)) = C_i(E_i^C(T)) + D_i(S^C(T)), \tag{16.53}$$

$$TC_i^N(S(T-1)) = C_i(E_i^N(T)) + D_i(S^N(T)). \tag{16.54}$$

It is clear that the sum of the total costs in the full-cooperative outcome are lower than in the Nash equilibrium which creates the gain of cooperation G that is given by

$$G(S(T-1)) = \sum_{i=1}^{n} TC_i^N(S(T-1)) - \sum_{i=1}^{n} TC_i^C(S(T-1)). \tag{16.55}$$

Suppose that the grand coalition chooses budget-neutral transfers θ_i of the form

$$\theta_i(S(T-1)) = TC_i^N(S(T-1)) - TC_i^C(S(T-1)) - \mu_i G(S(T-1)), \tag{16.56}$$

with $0 < \mu_i < 1$, $\sum_{i=1}^{n} \mu_i = 1$, so that country i ends up with total costs

$$T\tilde{C}_i^C(S(T-1)) = TC_i^N(S(T-1)) - \mu_i G(S(T-1)), 0 < \mu_i < 1, \sum_{i=1}^{n} \mu_i = 1. \tag{16.57}$$

This implies that the grand coalition allocates the costs in such a way that the total noncooperative costs of each country are decreased with a share in the gain of cooperation. This immediately guarantees that full cooperation is individually rational for each country. Chander and Tulkens (1995) show that in case of linear damages D, a proper choice of the μ_i also yields coalitional rationality so that the grand coalition is in the γ-core of the game. Each μ_i is simply equal to the marginal damage of country i divided by the sum of the marginal damages. Chander and Tulkens (1997) show that this result, under reasonable conditions, can be generalized to convex costs and damages.

In the penultimate period $T-1$, the level of the stock $S(T-2)$ is given, and a similar static game is played with objectives given by

$$\min_{E_i(T-1)} [C_i(E_i(T-1)) + D_i(S(T-1)) + dT\tilde{C}_i^C(S(T-1))], i = 1, 2, \ldots, n. \tag{16.58}$$

The same steps can be followed as in the final period T with the damage function D_i replaced by $D_i + dT\tilde{C}_i^C$. Again the full-cooperative outcome, the Nash equilibrium, the gain of cooperation, the transfers, and the resulting total costs for each country in period $T-1$ can be determined, as functions of the stock $S(T-2)$. Backward induction (or dynamic programming) unravels the solution for all time periods down to $t = 1$ with $S(0) = S_0$. Note that the resulting trajectory of the stock of atmospheric carbon is the full-cooperative trajectory. Note also that convexity of the function $D_i + dT\tilde{C}_i^C$ is not guaranteed by convexity of the cost and damage functions, so that second-order conditions and coalitional rationality have to checked. Germain et al. (2003) develop a numerical algorithm that calculates the transfers over time that are needed for stability of the grand coalition. This algorithm is applied to a climate change model with three regions and 30 periods. In that model, transfers are needed from the industrialized countries to China and at the end also a bit to the rest of the world. Below we will consider a third idea for stability of international environmental agreements, but first we will step aside and consider a dynamic allocation rule with another purpose than stability.

4.3 A Time-Consistent Allocation Rule

The analysis in the previous section focuses on the question of how the grand coalition should allocate the costs with the purpose to prevent deviations by any group of countries. However, cooperative game theory also provides allocation rules with other purposes, namely, to satisfy some set of axioms such as fairness. The Shapley value is the most simple and intuitive allocation rule (Shapley 1953). In a dynamic context, with a stock of atmospheric carbon, the question arises how the total costs should be allocated over time. Petrosjan and Zaccour (2003) argue that the allocation over time should be time consistent and show how this can be achieved, in case the Shapley value is the basis for allocation of costs. In this way the initial agreement is still valid if it is reconsidered at some point in time. The asymmetric infinite-horizon continuous-time version of the differential game (16.10) is used, with total costs as objective. This is given by

$$\min_{E_i(.)} \int_0^\infty e^{-rt}[C_i(E_i(t)) + D_i(S(t))]dt, i = 1, 2, \ldots, n, \tag{16.59}$$

subject to (16.12).

The feedback Nash equilibrium can be found with the Hamilton-Jacobi-Bellman equations in the current-value functions $V_i^N(S)$ given by

$$rV_i^N(S) = \min_{E_i}\{C_i(E_i)+D_i(S)+V_i^{N\prime}(S)(E_i+\sum_{j\neq i}^{n} E_j(S)-\delta S)\}, i = 1, 2, \ldots, n, \tag{16.60}$$

and the full-cooperative outcome can be found with the Hamilton-Jacobi-Bellman equation in the current-value function $V^C(S)$ given by

$$rV^C(S) = \min_{E_1,\ldots,E_n} \left\{ \sum_{i=1}^{n}[C_i(E_i) + D_i(S)] + V^{C\prime}(S) \left(\sum_{i=1}^{n} E_i - \delta S \right) \right\}. \tag{16.61}$$

This yields values $V_i(S)$ for individual countries and a value $V^C(S)$ for the grand coalition. Petrosjan and Zaccour (2003) suggest to derive values for a group of countries K by solving the Hamilton-Jacobi-Bellman equations in the current-value functions $V(K, S)$ given by

$$rV(K,S) = \min_{E_i, i \in K} \left\{ \sum_{i \in K} [C_i(E_i) + D_i(S)] + V'(K,S) \left(\sum_{i \in K} E_i + \sum_{i \notin K} E_i^N(S) - \delta S \right) \right\}, \tag{16.62}$$

where $E_i^N(S)$ denotes the emissions in the feedback Nash equilibrium. Note that this is not the feedback partial-cooperation Nash equilibrium because the emissions of the outsiders are fixed at the levels in the feedback Nash equilibrium. This simplification is made because the computational burden would otherwise be very high. This yields values $V(K, S)$ for groups of countries. Note that $V(\{i\}, S) = V_i(S)$ and $V(K, S) = V^C(S)$ if K is the grand coalition. The Shapley value $\phi(S) := [\phi_1(S), \dots, \phi_n(S)]$ is now given by

$$\phi_i(S) := \sum_{i \in K} \frac{(n-k)!(k-1)!}{n!} [V(K,S) - V(K \backslash \{i\}, S)], i = 1, 2, \dots, n, \tag{16.63}$$

where k is the size of the group of countries K.

The question is how the initial Shapley value $\phi(S_0)$ should be allocated over time. An allocation $\psi(t) := [\psi_1(t), \dots, \psi_n(t)]$ over time needs to satisfy

$$\phi_i(S_0) = \int_0^\infty e^{-rt} \psi_i(t) dt, i = 1, 2, \dots, n. \tag{16.64}$$

The allocation $\psi(t)$ yields the required time consistency if for all points in time t the following condition holds:

$$\phi_i(S_0) = \int_0^t e^{-rs} \psi_i(s) ds + e^{-rt} \phi_i(S^C(t)), i = 1, 2, \dots, n, \tag{16.65}$$

where S^C denotes the stock of atmospheric carbon in the full-cooperative outcome. It is easy to show that this condition holds for

$$\psi_i(t) = r\phi_i(S^C(t)) - \frac{d}{dt} \phi_i(S^C(t)), i = 1, 2, \dots, n. \tag{16.66}$$

If the stock of atmospheric carbon S^C is constant so that the Shapley value $\phi(S^C)$ is constant, the allocation (16.66) over time simply requires to pay the interest $r\phi(S^C)$

at each point in time. However, if the stock S^C changes over time, the allocation (16.66) has to be adjusted with the time derivative of the Shapley value $\phi(S^C(t))$ at the current stock of atmospheric carbon. In the next section, we return to the issue of stability.

4.4 Farsightedness

A third approach is provided by the concept of farsightedness (Chwe 1994) which is in a way connecting the two stability concepts above. Internal stability is "too weak," because it assumes no further deviations, but the γ-core approach is "too strong," because it assumes that the coalition falls apart completely. Farsightedness allows that deviations trigger further deviations, but this process comes to an end when a new stable coalition is reached. A set of farsighted stable coalitions can be built, starting at the small-size coalition that is internally and externally stable. The largest stable coalition in this set is usually close in size to the grand coalition. Deviations are prevented because a member of a farsighted stable coalition is better off than an outsider to the next-in-size farsighted stable coalition. This approach is used in a number of papers on international environmental agreements (e.g., Diamantoudi and Sartzetakis 2015; Osmani and Tol 2009).

An interesting issue in a dynamic context is that detection of a deviation may take time and that during that time the situation may have changed. In such a case, it may happen that a deviation is not deterred whereas it would have been deterred in a static context. Using farsighted stability, de Zeeuw (2008) investigates this in a model that is slightly different from the model above. The countries set a target for the total level of emissions, for example, because this level does not cause further temperature rise and prevents climate tipping. Each country can contribute to reducing excess emissions from the current total level E_0. Excess emissions E are costly to all countries, but reduction of emissions A_i (for "abatement") is costly as well. This yields an infinite-horizon discrete-time differential game which is given by

$$\min_{A_i(.)} \sum_{t=1}^{\infty} d^t \left[\frac{1}{2} pE^2(t) + \frac{1}{2} A_i^2(t) \right], i = 1, 2, \ldots, n, \tag{16.67}$$

subject to

$$E(t) = E(t-1) - \sum_{i=1}^{n} A_i(t), E(0) = E_0, \tag{16.68}$$

where p denotes the relative weight of the two cost components and d the discount factor. The level of the excess emissions E is the state of the system. The initial level E_0 has to be brought down to zero. In the full-cooperative outcome, this target is reached faster and with lower costs than in the Nash equilibrium.

Suppose again that the size of the coalition is equal to $k \leqq n$. Dynamic programming yields the feedback partial-cooperation Nash equilibrium. If the value functions are denoted by $V^m(E) := \frac{1}{2}c^m E^2$ and $V^o(E) := \frac{1}{2}c^o E^2$ for a member of the coalition and for an outsider, respectively, it is tedious but straightforward to show that the stationary feedback partial-cooperation Nash equilibrium becomes

$$A^m(E) = \frac{dkc^m E}{1 + d(k^2c^m + (n-k)c^o)}, A^o(E) = \frac{dc^o E}{1 + d(k^2c^m + (n-k)c^o)}, \tag{16.69}$$

where the cost parameters c^m and c^o of the value functions have to satisfy the set of equations

$$c^m = p + \frac{dc^m(1 + dk^2c^m)}{[1 + d(k^2c^m + (n-k)c^o)]^2}, c^o = p + \frac{dc^o(1 + dc^o)}{[1 + d(k^2c^m + (n-k)c^o)]^2}. \tag{16.70}$$

Internal stability requires that $c^o(k-1) \leqq c^m(k)$ and it is easy to show that this can again only hold for $k = 2$. However, farsighted stability requires that $c^o(k-l) \leqq c^m(k)$ for some $l > 1$ such that $k - l$ is a stable coalition (e.g., $k - l = 2$). In this way a set of farsighted stable coalitions can be build.

In a dynamic context, however, it is reasonable to assume that detection of a deviation takes time. This gives rise to the concept of dynamic farsighted stability. The idea is that the deviator becomes an outsider to the smaller stable coalition in the next period. This implies that the deviator has free-rider benefits for one period without losing the cooperative benefits. The question is whether large stable coalitions can be sustained in this case. It is tedious but straightforward to derive the value function $V^d(E) := \frac{1}{2}c^d E^2$ of the deviator. The cost parameter c^d is given by

$$c^d = p + \frac{dc^{o+}}{1 + dc^{o+}} \frac{(1 + dkc^m)^2}{[1 + d(k^2c^m + (n-k)c^o)]^2}, \tag{16.71}$$

where c^{o+} is the cost parameter of the value function of an outsider to the smaller stable coalition in the next period. Deviations are deterred if $c^m < c^d$ but the analysis is complicated. The cost parameters c^m and c^o and the corresponding stable set have to be simultaneously solved, and this can only be done numerically. In de Zeeuw (2008) the whole spectrum of results is presented. The main message is that large stable coalitions can only be sustained if the weighing parameter p is very small. The intuition is that a large p implies relatively large costs of the excess emissions E, so that the emissions and the costs are quickly reduced. In such a case, the threat of triggering a smaller stable coalition, at the low level of excess emissions in the next period, is not sufficiently strong to deter deviations. In the last section, we briefly consider evolutionary aspects of international environmental agreements.

4.5 Evolutionary Games

Evolutionary game theory provides a completely different perspective on the stable size of an international environmental agreement. The idea is that countries behave cooperatively (as members of the coalition) or noncooperatively (as deviators) but that they change their behavior when they observe that the other type of behavior is more successful. If this dynamic behavioral adjustment process converges, the resulting size of the coalition is called the evolutionary stable size.

In order to investigate this, Breton et al. (2010) use a discrete-time version of the differential game (16.10) which is given by

$$\max_{E_i()} W_i = \sum_{t=1}^{\infty} d^t [B(E_i(t)) - D(S(t))], i = 1, 2, \ldots, n, \tag{16.72}$$

with $B(E) = \beta E - \frac{1}{2}E^2$ and linear damage function $D(S) = \gamma S$, subject to (16.52). The linearity assumption simplifies the analysis, because emissions become independent of the stock and the emissions of the other countries. A coalition of size k jointly maximizes $\sum_{i=1}^{k} W_i$. The idea driving the result is that each member of the coalition inflicts a punishment on each outsider (a trade restriction or a carbon tax on exports from the outsider) for irresponsible behavior. The punishment αS is proportional to the level of the stock. Each outsider thus incurs a punishment $k\alpha S$, so that the cost function of an outsider becomes $c^o S$ with $c^o = \gamma + k\alpha$. Each punishment is costly for the punisher as well, so that the cost function of a member of the coalition becomes $c^m S$ with $c^m = \gamma + (n - k)\eta\alpha$, where η denotes how costly. In this way the spectrum of possible stable coalitions is enlarged.

First the feedback partial-cooperation Nash equilibrium of this dynamic game in discrete time is derived. The dynamic-programming equations in the value functions $V^m(S)$ and $V^o(S)$ are given by

$$V^m(S) = \max_{E_1,\ldots,E_k} \left\{ \sum_{i=1}^{k} [\beta E_i - \frac{1}{2}E_i^2 - c^m S^+] + dV^m(S^+) \right\}, \tag{16.73}$$

$$V^o(S) = \max_{E_i} \{\beta E_i - \frac{1}{2}E_i^2 - c^o S^+ + dV^o(S^+)\}, i = k + 1, \ldots, n, \tag{16.74}$$

$$S^+ = (1 - \delta)S + \sum_{i=1}^{n} E_i. \tag{16.75}$$

It is easy to show, using linear value functions $V^m(S) = \sigma_0^m - \sigma_1^m S$ and $V^o(S) = \sigma_0^o - \sigma_1^o S$, that the feedback partial-cooperation Nash equilibrium does not depend on the stock S and is given by

$$E^m = \beta - \frac{kc^m}{1 - d(1-\delta)}, E^o = \beta - \frac{c^o}{1 - d(1-\delta)}. \tag{16.76}$$

Solving the parameters of the value functions, the welfare levels of an *individual* member of the coalition and an outsider become

$$\frac{\beta E^m - \frac{1}{2}E^{m2}}{1-d} - \frac{c^m}{1-d(1-\delta)}\left[(1-\delta)S + \frac{kE^m + (n-k)E^o}{1-d}\right], \tag{16.77}$$

$$\frac{\beta E^o - \frac{1}{2}E^{o2}}{1-d} - \frac{c^o}{1-d(1-\delta)}\left[(1-\delta)S + \frac{kE^m + (n-k)E^o}{1-d}\right]. \tag{16.78}$$

It is common in evolutionary game theory to denote the numbers for each type of behavior as fractions of the total population. The fraction of coalition members is given by $q := k/n$, so that the fraction of outsiders becomes $1 - q$. The level of cooperation is indicated by q, and k and $n - k$ are replaced by qn and $(1-q)n$, respectively. Note that c^m and c^o also depend on k and $n - k$ and therefore on q. The welfare levels (16.77)–(16.78) are denoted by $W^m(S,q)$ and $W^o(S,q)$, respectively.

The basic idea of evolutionary game theory is that the fraction q of coalition members increases if they are more successful than deviators, and vice versa. This is modeled with the so-called replicator dynamics given by

$$q^+ = \frac{W^m(S,q)}{qW^m(S,q) + (1-q)W^o(S,q-1/n)}q. \tag{16.79}$$

The denominator of (16.79) is the weighted average of the welfare of a coalition member and that of a country deviating from the agreement. The weights are the current fractions of coalition members and outsiders. If the welfare of a coalition member is higher than this average, the fraction of coalition members increases, and vice versa. This gives rise to a dynamic system consisting of (16.79) and (16.52). The steady state of this system is given by $W^m(S,q) = W^o(S,q-1/n)$ and $qnE^m + (1-q)nE^o = \delta S$.

The interesting question is what the evolutionary stable fraction of coalition members is and how this depends on the punishment parameters α and η. This requires a numerical analysis. Breton et al. (2010) first vary the parameter α and fix all other parameters. It is shown that below a certain value of α, there are no steady states, and therefore no cooperation arises. However, slightly increasing α yields two steady states. The higher one is the stable steady state and this yields a substantial level of cooperation. The lower one indicates the minimal initial level of cooperation that is needed to reach the stable steady state. Further increasing α leads to full cooperation and extends the area of initial levels of cooperation from where it can be reached. Decreasing the cost of punishment η improves the situation in the same way.

Ochea and de Zeeuw (2015) have a similar result but formulate a different evolutionary game that is not based on a differential game. Cooperating countries are conditional cooperators in the sense that they only continue cooperating if sufficiently many other countries are cooperating ("tit-for-tat" strategies). The other countries are free riders. Replicator dynamics is used as well. It is shown that full cooperation arises if the conditional thresholds and the initial set of cooperators are sufficiently high. Apparently, the cooperators have to be tough and have to start off with a sufficiently high number in order to reach full cooperation.

5 Conclusion

International pollution control with a global stock externality (e.g., climate change) is a typical example of a differential game. The countries are the players and emissions, as by-products of economic activities, accumulate into the global stock of pollutants that is damaging to all countries. In the first part, this chapter compares the steady states of the open-loop, linear feedback and nonlinear feedback Nash equilibria of this symmetric linear-quadratic differential game with the full-cooperative steady state. It is shown that the linear feedback Nash equilibrium performs worse than the open-loop Nash equilibrium in this respect, but a nonlinear feedback Nash equilibrium exists that performs better.

The second part of this chapter focuses on partial cooperation, representing an international environmental agreement. Different stability concepts are considered that may sustain partial or even full cooperation. Internal and external stability allows for only a small stable coalition, as in the static context. Farsighted stability allows for large stable coalitions in the static context, but this breaks down in the dynamic context if detection of a deviation takes time and the costs of pollution are relatively high. On the other hand, the threat of all cooperation falling apart, as is usually assumed in cooperative game theory, prevents deviations and can sustain the grand coalition. In the asymmetric case, transfers between the coalition members are needed for stability. In the dynamic context, these transfers have to change over time, as is shown in this chapter. Cooperative game theory also suggests transfers that satisfy some set of axioms such as fairness. It is shown how transfers yielding the Shapley value have to be allocated over time in order to achieve time consistency of these transfers. Finally, stability in the evolutionary sense is investigated. It is shown that the grand coalition is evolutionary stable if the coalition members start out with a sufficiently high number and inflict a sufficiently high punishment on the outsiders.

This chapter has presented only a selection of the analyses that can be found in the literature on this topic. It has attempted to fit everything into a coherent story and to give examples of the different ways in which differential games are used. The selection was small enough to be able to provide some detail in each of the analyses but large enough to cover the main angles of approaching the problems. The subject is very important, the tool is very powerful, a lot of work is still ahead of us, and therefore it is to be expected that this topic will continue to attract attention.

References

Barrett S (1994) Self-enforcing international environmental agreements. Oxf Econ Pap 46(4(1)):878–894
Başar T, Olsder G-J (1982) Dynamic noncooperative game theory. Academic Press, London
Breton M, Sbragia L, Zaccour G (2010) A dynamic model for international environmental agreements. Environ Resour Econ 45(1):23–48
Calvo E, Rubio SJ (2013) Dynamic models of international environmental agreements: a differential game approach. Int Rev Environ Resour Econ 6(4):289–339
Carraro C, Siniscalco D (1993) Strategies for the international protection of the environment. J Public Econ 52(3):309–328
Chander P, Tulkens H (1995) A core-theoretic solution for the design of cooperative agreements on transfrontier pollution. Int Tax Public Financ 2(2):279–293
Chander P, Tulkens H (1997) The core of an economy with multilateral environmental externalities. Int J Game Theory 26(3):379–401
Chwe MS-Y (1994) Farsighted coalitional stability. J Econ Theory 63(2):299–325
d'Aspremont C, Jacquemin A, Gabszewicz JJ, Weymark JA (1983) On the stability of collusive price leadership. Can J Econ 16(1):17–25
de Zeeuw A (2008) Dynamic effects on the stability of international environmental agreements. J Environ Econ Manag 55(2):163–174
de Zeeuw A, Zemel A (2012) Regime shifts and uncertainty in pollution control. J Econ Dyn Control 36(7):939–950
Diamantoudi E, Sartzetakis ES (2015) International environmental agreements: coordinated action under foresight. Econ Theory 59(3):527–546
Dockner E, Long NV (1993) International pollution control: cooperative versus noncooperative strategies. J Environ Econ Manag 25(1):13–29
Dockner E, Jorgensen S, Long NV, Sorger G (2000) Differential games in economics and management science. Cambridge University Press, Cambridge
Fershtman C, Kamien MI (1987) Dynamic duopolistic competition. Econometrica 55(5): 1151–1164
Germain M, Toint P, Tulkens H, de Zeeuw A (2003) Transfers to sustain dynamic core-theoretic cooperation in international stock pollutant control. J Econ Dyn Control 28(1):79–99
Hardin G (1968) The tragedy of the commons. Science 162:1243–1248
Hoel M (1993) Intertemporal properties of an international carbon tax. Resour Energy Econ 15(1):51–70
Jorgensen S, Martin-Herran G, Zaccour G (2010) Dynamic games in the economics and management of pollution. Environ Model Assess 15(6):433–467
Kaitala V, Pohjola M, Tahvonen O (1992) Transboundary air pollution and soil acidification: a dynamic analysis of an acid rain game between Finland and the USSR. Environ Res Econ 2(2):161–181
Kossioris G, Plexousakis M, Xepapadeas A, de Zeeuw A, Mäler K-G (2008) Feedback Nash equilibria for non-linear differential games in pollution control. J Econ Dyn Control 32(4): 1312–1331
Lenton TM, Ciscar J-C (2013) Integrating tipping points into climate impact assessmens. Clim Chang 117(3):585–597
Long NV (1992) Pollution control: a differential game approach. Ann Oper Res 37(1):283–296
Long NV (2012) Applications of dynamic games to global and transboundary environmental issues: a review of the literature. Strateg Behav Env 2(1):1–59
Long NV, Sorger G (2006) Insecure property rights and growth: the role of appropriation costs, wealth effects, and heterogeneity. Econ Theory 28(3):513–529
Mäler K-G, de Zeeuw A (1998) The acid rain differential game. Environ Res Econ 12(2):167–184
Ochea MI, de Zeeuw A (2015) Evolution of reciprocity in asymmetric international environmental negotiations. Environ Res Econ 62(4):837–854

Osmani D, Tol R (2009) Toward farsightedly stable international environmental agreements. J Public Econ Theory 11(3):455–492
Petrosjan L, Zaccour G (2003) Time-consistent Shapley value allocation of pollution cost reduction. J Econ Dyn Control 27(3):381–398
Rowat C (2007) Non-linear strategies in a linear quadratic differential game. J Econ Dyn Control 31(10):3179–3202
Rubio SJ, Casino B (2002) A note on cooperative versus non-cooperative strategies in international pollution control. Res Energy Econ 24(3):251–261
Rubio SJ, Casino B (2005) Self-enforcing international environmental agreements with a stock pollutant. Span Econ Rev 7(2):89–109
Rubio SJ, Ulph A (2007) An infinite-horizon model of dynamic membership of international environmental agreements. J Environ Econ Manag 54(3):296–310
Shapley LS (1953) A value for n-person games. In: Kuhn HW, Tucker AW (eds) Contributions to the theory of games II. Princeton University Press, Princeton, pp 307–317
Starr AW, Ho YC (1969) Further properties of nonzero-sum differential games. J Optim Theory Appl 3(4):207–219
Tornell A, Velasco A (1992) The tragedy of the commons and economic growth: why does capital flow from poor to rich countries? J Polit Econ 100(6):1208–1231
Tsutsui S, Mino K (1990) Nonlinear strategies in dynamic duopolistic competition with sticky prices. J Econ Theory 52(1):131–161
van der Ploeg F, de Zeeuw AJ (1992) International aspects of pollution control. Environ Res Econ 2(2):117–139
van der Ploeg F, de Zeeuw A (2016, forthcoming) Non-cooperative and cooperative responses to climate catastrophes in the global economy: a North-South perspective. Environ Res Econ

17 Dynamic Games in Macroeconomics

Łukasz Balbus, Kevin Reffett, and Łukasz Woźny

Contents

Ł. Balbus (✉)
Faculty of Mathematics, Computer Sciences and Econometrics, University of Zielona Góra, Zielona Góra, Poland
e-mail: l.balbus@wmie.uz.zgora.pl

K. Reffett
Department of Economics, Arizona State University, Tempe, AZ, USA
e-mail: kevin.reffett@asu.edu

Ł. Woźny
Department of Quantitative Economics, Warsaw School of Economics, Warsaw, Poland
e-mail: lukasz.wozny@sgh.waw.pl

T. Başar, G. Zaccour (eds.), *Handbook of Dynamic Game Theory*,
https://doi.org/10.1007/978-3-319-44374-4_18

Abstract
In this chapter, we survey how the methods of dynamic and stochastic games have been applied in macroeconomic research. In our discussion of methods for constructing dynamic equilibria in such models, we focus on strategic dynamic programming, which has found extensive application for solving macroeconomic models. We first start by presenting some prototypes of dynamic and stochastic games that have arisen in macroeconomics and their main challenges related to both their theoretical and numerical analysis. Then, we discuss the strategic dynamic programming method with states, which is useful for proving existence of sequential or subgame perfect equilibrium of a dynamic game. We then discuss how these methods have been applied to some canonical examples in macroeconomics, varying from sequential equilibria of dynamic nonoptimal economies to time-consistent policies or policy games. We conclude with a brief discussion and survey of alternative methods that are useful for some macroeconomic problems.

Keywords
Strategic dynamic programming · Sequential equilibria · Markov equilibria · Perfect public equilibria · Non-optimal economies · Time-consistency problems · Policy games · Numerical methods · Approximating sets · Computing correspondences

1 Introduction

The seminal work of Kydland and Prescott (1977) on time-consistent policy design initiated a new and vast literature applying the methods of dynamic and stochastic games in macroeconomics and has become an important landmark in modern macroeconomics.[1] In their paper, the authors describe a very simple optimal policy design problem in the context of a dynamic general equilibrium model, where government policymakers are tasked with choosing an optimal mixture of policy instruments to maximize a common social objective function. In this simple model, they show that the consistent policy of the policymaker is not optimal because it does not take account of the effect of his future policy instrument on economic agents' present decision. In fact, Kydland and Prescott (1977) make the point that a policy problem cannot be dealt with just optimal control theory since there a policymaker is interacting with economic agents having rational expectations. In other words, as the successive generations of policymakers cannot commit to the future announced plans of the current generation, they argue that one cannot assume that optimal plans

[1] Of course, there was prior work in economics using the language of dynamic games that was related to macroeconomic models (e.g., Phelps and Pollak 1968; Pollak 1968; Strotz 1955) but the paper of Kydland and Prescott changed the entire direction of the conversation on macroeconomic policy design.

that are designed by any current generation of government policymakers will ever be followed if they are not required to be additionally dynamically consistent. This observation gave rise to a new and very important question of how to construct *credible* government policies, as well as raising the question of whether discretion vs. rules were more important to the design of optimal policy, and the study of dynamic macroeconomic models with strategically interacting agents and limited commitment begun (and has continued for the last four decades).

In subsequent work, Kydland and Prescott (1980) proposed a new set of recursive methods for constructing time-consistent optimal policies in decentralized dynamic equilibrium models with capital and labor. Their methods actually were an integration of new dynamic optimization techniques under additional constraints (i.e., constraints that were added to guarantee decision-makers would look forward or backward in a manner that the resulting optimal decisions for future policy were time consistent). Their methods in this paper introduced the idea of using set-valued operators to construct time-consistent sequential equilibrium solutions defined recursively on an *expanded* set of endogenous state variables that could be used to provide the needed dynamic incentives for them to choose time-consistent solutions.

Their methods, although not explicitly game theoretic, provided an important preamble to the introduction of more general, powerful, and systematic game theoretic approaches that are now central to much work in macroeconomics. These new methods are referred in the literature as "strategic dynamic programming methods" and are built upon the seminal work of Abreu et al. (1986, 1990) (APS) for solving for the equilibrium value set of very general classes of repeated games. As in the original Kydland-Prescott approach (e.g., Kydland and Prescott 1980), they introduce new state variables (in this case, either value functions or envelope theorems) and in essence are set-valued generalizations of standard dynamic programming methods. This approach (especially since the pioneering paper of Atkeson 1991) has found many important implementations to solve macroeconomic models with limited commitment or dynamically inconsistent preferences and is (in their structure) basically APS method extended to models with state variables.[2] These methods both verify the existence of subgame perfect equilibrium in a large class of dynamic/stochastic games, and they provide a systematic method for constructing all the sequential or subgame perfect equilibria in many dynamic macroeconomic models that can be formulated as a dynamic game.

In this chapter, we survey some of the important literature on macroeconomic models that use the methods of dynamic and stochastic games. We first discuss the literature and how dynamic and stochastic games naturally arise in dynamic

[2]Strategic dynamic programming methods were first described in the seminal papers of Abreu (1988) and Abreu et al. (1986, 1990), and they were used to construct the entire set of sequential equilibrium values for repeated games with discounting. These methods have been subsequently extended in the work of Atkeson (1991), Judd et al. (2003), and Sleet and Yeltekin (2016), among others.

general equilibrium models that are the workhorse of macroeconomic modeling. We then discuss strategic dynamic programming methods extending to setting with state variables that are very important for solving these models. We focus on strategic dynamic programming with states, as when these methods apply, they provide a systematic method for constructing *all* dynamic equilibria in the models. At the end of the chapter, we also discuss alternative optimization and Euler equation-based methods for solving these models, which have also been studied in the literature. These latter methods, although in some cases not explicitly game theoretic, provide powerful alternatives to the set-theoretic approaches that APS methods with state variables provide.

There are many prototype problems in macroeconomics that require the tools of dynamic game theory, and there are a number of alternative methods for studying these models. Take, for example, the paper of Phelan and Stacchetti (2001), where they consider optimal taxation in a model first described in Kydland and Prescott (1980), where the structure of optimal taxation in their model was studied as a sequential equilibrium of a dynamic game played between overlapping generations of government policymakers who are collectively tasked with choosing an optimal sequence of capital and/or labor taxes to finance a stream of government spending over an infinite horizon, where government policymakers maximize the representative agent's lifetime utility function in a sequential equilibrium. Further, as Kydland and Prescott (1980) showed, as labor and capital decisions in the private economy are made endogenously by households and firms, the resulting dynastic social objective function for the collective government is not dynamically consistent. This raised the interesting question of studying sustainable (or credible) optimal taxation policies, where constraints forcing the government to make time-consistent choices further restricted the set of optimal government policies (i.e., forced optimal government policies to satisfy a further restriction that all current plans about decisions by future generations of government policymakers are actually optimal for those successor generations of policymakers when their decisions have to be made). This situation was distinct from previous work in dynamic general equilibrium theory (as well as much of the subsequent work on optimal policy design over the decade after their paper) which assumed perfect commitment on the part of government policymakers.[3] In showing this (far from innocuous) assumption of perfect commitment in dynamic economies, Kydland and Prescott (1980) asked the question of how to resolve this fundamental credibility issue for optimal policy design. Their construction of dynamic equilibria incorporated explicitly the strategic considerations between current and future policy agents into the design of sequential equilibrium optimal plans.

The papers of Kydland and Prescott (1980) and Phelan and Stacchetti (2001) also provide a nice comparison and contrast of methods for studying macroeconomic

[3]The model they studied turned out to be closely related to the important work on optimal dynamic taxation in models with perfect commitment in the papers of Judd (1985) and Chamley (1986). For a recent discussion, see Straub and Werning (2014).

models with dynamic strategic interaction, dynamically inconsistent preferences, or limited commitment. Basically, the authors study very related dynamic economies (i.e., so-called „Ramsey optimal taxation models"), but their approaches to constructing time-consistent solutions are very different. Kydland and Prescott (1980) viewed the problem of constructing time-consistent optimal plans from the vantage point of optimization theory (with a side condition that is a fixed point problem that is used to guarantee time consistency). That is, they forced the decision-maker to respect the additional implicit constraint of time consistency by adding new endogenous state variables to further restrict the set of optimal plans from which government policymakers could choose, and the structure of that new endogenous state variable is determined by a (set-valued) fixed point problem. This "recursive optimization" approach has a long legacy in the theory of consistent plans and time-consistent optimization.[4]

Phelan and Stacchetti (2001) view the problem somewhat differently, as a dynamic game between successive generations of government policymakers. When viewing the problem this way, in the macroeconomics literature, the role for strategic dynamic programming provided the author a systematic methodology for both proving existence of, and potentially computing, sequential equilibria in macroeconomic models formulated as a dynamic/stochastic game.[5] As we shall discuss in the chapter, this difference in viewpoint has its roots in an old literature in economics on models with dynamically inconsistent preference beginning with Strotz (1955) and subsequent papers by Pollak (1968), Phelps and Pollak (1968), and Peleg and Yaari (1973).

One interesting feature of this particular application is that the methods differ in a sense from the standard strategic dynamic programming approach of APS for dynamic games with states. In particular, they differ by choice of expanded state variables, and this difference in choice is intimately related to the structure of dynamic macroeconomic models with strategically interacting agents. Phelan and Stacchetti (2001) note, as do Dominguez and Feng (2016a,b) and Feng (2015) subsequently, that an important technical feature of the optimal taxation problem is the presence of Euler equations for the private economy. This allows them to develop for optimal taxation problems a hybrid of the strategic dynamic programming methods of APS. That is, like APS, the recursive methods these authors develop employ enlarged states spaces, but unlike APS, in this particular case, these

[4]Indeed, in the original work of Strotz (1955), this was the approach taken. This approach was somehow criticized in the work of Pollak (1968), Phelps and Pollak (1968), and Peleg and Yaari (1973). See also Caplin and Leahy (2006) for a very nice discussion of this tradition.

[5]In some cases, researchers also seek further restrictions of the set of dynamic equilibria studied in these models, and they focus on Markov perfect equilibria. Hence, the question of memory in strategic dynamic programming methods has also been brought up. To answer this question, researchers have sought to generate the value correspondence in APS type methods using nonstationary Markov perfect equilibria. See Doraszelski and Escobar (2012) and Balbus and Woźny (2016) for discussion of these methods.

additional state variables are Karush-Kuhn-Tucker (KKT) multipliers or envelope theorems (e.g., as is also done by Feng et al. 2014).

These enlarged state space methods have also given rise to a new class of recursive optimization methods that incorporate strategic considerations and dynamic incentive constraints explicitly into dynamic optimization problems faced by social planners. For early work using this recursive optimization approach, see Rustichini (1998a) for a description of so-called primal optimization methods and also Rustichini (1998b) and Marcet and Marimon (1998) for related "dual" recursive optimization methods using recursive Lagrangian approaches.

Since Kydland and Prescott's work was published, over the last four decades, it had become clear that issues related to time inconsistency and limited commitment can play a key role in understanding many interesting issues in macroeconomics. For example, although the original papers of Kydland and Prescott focused on optimal fiscal policy primarily, the early papers by Fischer (1980a) and Barro and Gordon (1983) showed that similar problems arise in very simple monetary economies, when the question of optimal monetary policy design is studied. In such models, again, the sequential equilibrium of the private economy can create similar issues with dynamic consistency of objective functions used to study the optimal monetary policy rule question, and therefore the sequential optimization problem facing successive generations of central bankers generates optimal solutions that are not time consistent. In Barro and Gordon (1983), and subsequent important work by Chang (1998), Sleet (2001), Athey et al. (2005), and Sleet and Yeltekin (2007), one can then view the problem of designing optimal monetary policy as a dynamic game, with sequential equilibrium in the game implementing time-consistent optimal monetary policy.

But such strategic considerations have also appeared outside the realm of policy design and have become increasingly important in explaining many important phenomena observed in macroeconomic data. The recent work studying consumption-savings puzzles in the empirical data (e.g., why do people save so little?) has focused on hyperbolic discounting and dynamically inconsistent choice as a basis for an explanation. Following the pioneering paper by Strotz (1955), where he studied the question of time-consistent plans for decision-makers whose preferences are changing overtime, many researchers have attempted to study dynamic models where agents are endowed with preferences that are dynamically inconsistent (e.g., Harris and Laibson 2001, 2013; Krusell and Smith 2003, 2008; Laibson 1997). In such models, at any point in time, agents make decisions on current and future consumption-savings decisions, but their preferences exhibit the so-called present-bias. These models have also been used to explain sources of poverty (e.g., see Banerjee and Mullainathan 2010; Bernheim et al. 2015). More generally, the question of delay, procrastination, and the optimal timing of dynamic choices have been studied in O'Donoghue and Rabin (1999, 2001), which has started an important discussion of how to use models with dynamically inconsistent payoffs to explain observed behavior in a wide array of applications, including dynamic asset choice.

Additionally, when trying to explain the plethora of defaults that we observe in actual macroeconomies, and further address the question of how to sustain sovereign debt arrangements and debt repudiation, a new theory of asset markets with strategic default has emerged, where the role of limited commitment has generated a wide array of new models of dynamic insurance under incomplete markets with strategic default. These models have been applied to many important problems in international lending, where limited commitment plays a key role in understanding financial arrangements. Strategic default also plays a key role in the construction of dynamic models with endogenous borrowing constraints. These models have played a critical role in explaining various asset pricing puzzles in the macroeconomics literature. This literature began with the important early paper by Atkeson (1991) which studies international lending and debt repudiation; but the problem of sustainable debt under limited commitment has been studied in the early work of Kehoe and Levine (1993, 2001), as well as in Alvarez and Jermann (2000), and Hellwig and Lorenzoni (2009). Further, the issue of sovereign debt repudiation has been studied in a number of papers including Arellano (2008), Benjamin and Wright (2009), Yue (2010), and Broner et al. (2014, 2010).

One final prototype of a dynamic game in macroeconomics arises in models of economic growth with limited commitment. One common version of this sort of model arises in models of strategic altruism, where a dynastic household faces a collective choice problem between successive generations of families. Models in this spirit were first introduced in Phelps and Pollak (1968) and subsequently studied in Bernheim and Ray (1983), Leininger (1986), Amir (1996b), Nowak (2006c), Balbus et al. (2012, 2014, 2015a,b,c) and Woźny and Growiec (2012), among others. Another classic example of strategic growth models arises in the seminal work of Levhari and Mirman (1980), where the "great fishwar" was originally studied. In this model, a collection of agents face the problem of managing a common resource pool, where each period agents can consume from the existing stock of resources, with the remainder of that stock being used as input to a regeneration process (i.e., as investment into a social production function) that produces next period stock of resources. This problem has been extensively studied (e.g., Mirman (1979), Sundaram (1989a), Amir (1996b), Balbus and Nowak (2004), Nowak (2006a,b), Jaśkiewicz and Nowak (2015), and Fesselmeyer et al. (2016) among others).

As dynamic games have been introduced more extensively into macroeconomics, researchers have developed some very powerful methods for studying sequential or Markovian equilibrium in such models. For example, in the macroeconomic models where sequential optimization problems for agents have preferences that are *changing* over time, when searching for time-consistent optimal solutions, since the work of Strotz (1955) it has been known that additional constraints on the recursive optimization problem must be imposed. These constraints can be formulated as either backward- or forward-looking constraints. In Kydland and Prescott (1980), they proposed a very interesting resolution to the problem. In particular, they reformulate the optimal policy design problem recursively in the presence of additional *endogenous* state variables that are used to force optimal plans of the government decision-makers to be time consistent. That is, one can formulate an agent's incentive to deviate from a candidate dynamic equilibrium future choices by imposing a sequence of incentive constraints on current choices. Given these additional incentive constraints, one is led to a natural choice for a set of new endogenous state variables (e.g., value functions, Kuhn-Tucker multipliers, envelope theorems, etc.). Such added state variables also allow one to represent sequential equilibrium problems recursively. That is, they force optimal policies to condition on lagged values of Kuhn-Tucker multipliers. In these papers, one constructs the new state variable as the fixed point of a set-valued operator (similar, in spirit, to the methods discussed in Abreu et al. (1986, 1990) adapted to dynamic games. See Atkeson (1991) and Sleet and Yeltekin (2016)).[6]

The methods of Kydland and Prescott (1980) have been extended substantially using dynamic optimization techniques, where the presence of strategic interaction creates the need to further constrain these optimization problems with period-by-period dynamic incentive constraints. These problems have led to the development of "incentive-constrained" dynamic programming techniques (e.g., see Rustichini (1998a) for an early version of "primal" incentive-constrained dynamic programming methods and Rustichini (1998b) and Marcet and Marimon (1998) for early discussions of "dual" methods). Indeed, Kydland and Prescott's methodological approach was essentially the first "recursive dual" approach to a dynamic consistency problem. Unfortunately, in either formulation of the incentive-constrained dynamic programming approach, these optimization methods have some serious methodological issues associated with their implementation. For example, in some problems, these additional incentive constraints are often difficult to formulate (e.g., for models with quasi-hyperbolic discounting. See Pollak 1968). Further, when these constraints can be formulated, they often involve punishment schemes that are ad hoc (e.g., see Marcet and Marimon 1998).

Now, additionally, in "dual formulations" of these dynamic optimization approaches, problems with dual solutions not being primal feasible can arise even in convex formulations of these problems (e.g., see Messner and Pavoni 2016), dual variables can be very poorly behaved (e.g., see Rustichini 1998b), and the programs are not necessarily convex (hence, the existence of recursive saddle points is not known, and the existing duality theory is poorly developed. See Rustichini (1998a) for an early discussion and Messner et al. (2012, 2014) for a discussion of problems with recursive dual approaches). It bears mentioning, all these duality issues also arise in the methods proposed by Kydland and Prescott (1980). This dual approach has been extended in a number of recent papers to related problems, including Marcet and Marimon (2011), Cole and Kubler (2012), and Messner et al. (2012, 2014).

In addition to recursive dual approaches, incentive-constrained dynamic programming methods using "primal" formulations have been also proposed, and these

[6]The key difference between the standard APS methods and those using dual variables such as in Kydland and Prescott (1980), and Feng et al. (2014) is that in the former literature, value functions are used as the new state variables; hence, APS methods are closely related to "primal" methods, not dual methods.

methods do not exploit the dynamic structure of the set of Karush-Kuhn-Tucker multipliers associated with the recursive dual approach. As with dual dynamic optimization approaches, these primal methods also suffer from the problem that they are not concave programs. Further, characterizing optimal solutions can be very problematic.

Because of issues related to these "dynamic optimization" approaches, strategic dynamic programming has emerged as a systematic approach to this problem of constructing sequential equilibrium in dynamic macroeconomic models that can be formulated as a dynamic (or stochastic) game. For example, for the optimal taxation economy in Kydland and Prescott, where time-consistent optimal policies are viewed as subgame perfect equilibrium in a dynamic game played by successive generations of government policymakers, one can first construct a sequential equilibrium for the private economy for each sequential path for policy and then considers in the second stage a dynamic game played between successive generations of short-lived policymakers assuming no commitment (e.g., Dominguez and Feng 2016b; Phelan and Stacchetti 2001). This method, in some broad sense, can be thought of as a generalization of a "primal" incentive-constrained dynamic programming method, and this method has played a key role in the study of sustainable optimal government policy (e.g., see Sleet (2001) for an early discussion of using strategic dynamic programming methods for studying optimal monetary policy). In either case, one can additionally consider the role of reputation in the sustainability of optimal government plans (e.g., see Rogoff (1987) for an early discussion of this approach). In this latter approach, strategic dynamic programming methods that extend the seminal work of Abreu (1988) and Abreu et al. (1986, 1990) have been typically employed. We shall focus primarily on these strategic dynamic programming methods for studying strategic interaction in macroeconomic models that are formulated as a dynamic game in this chapter.

The rest of this chapter is laid out as follows: in the next section, we survey the application of dynamic and stochastic games in macroeconomics. In Sect. 3, we discuss the strategic dynamic programming approach to studying sequential equilibrium (and subgame perfect equilibrium) in these models more formally. We discuss both the extension of APS methods to models with states, as well as in Sect. 4 discuss some computational issues associated with strategic dynamic programming methods. In Sect. 5, we return to particular versions of the models discussed in Sect. 2 and discuss how to formulate sequential equilibrium in these models using strategic dynamic programming. In Sect. 6, we briefly discuss some alternative approaches to dynamic games in macroeconomics, and in the last section, we conclude.

2 Dynamic and Stochastic Games in Macroeconomics

The literature on dynamic and stochastic games in macroeconomics is extensive. These models often share a common structure and are dynamic general equilibrium models where some (or all) of the economic agents have dynamically inconsistent preferences or limited commitment generating a source of strategic interaction. In some models, the dynamic inconsistency problems stem from the primitive data of the model (e.g., models where agents have lifetime preferences that exhibit hyperbolic discounting). In other models, strategic interactions emerge because of the lack of commitment (i.e., as in dynastic models of economic growth where current generations care about future generations, but cannot control what future generations actually decide, or asset accumulations models with strategic default where one cannot assume borrowers will reply unless it is in their incentive to do so). Still in other models, the source of dynamic inconsistency comes from the structure of sequential equilibrium (e.g., preferences for government decision-makers designing optimal fiscal or monetary policy which are time inconsistent because of how the private economy responses in a sequential equilibrium to government policy). We now describe few prototypes of these models that we shall discuss in Sect. 4 of the chapter.

2.1 Hyperbolic Discounting

One prototype for dynamic games in macroeconomics is infinite horizon model of optimal economic growth or asset allocation where households have dynamically inconsistent preferences. The most studied version of this problem is economy where agents have preferences that exhibit hyperbolic discounting. This problem was first studied in Strotz (1955), subsequently by Pollak (1968) and Phelps and Pollak (1968), and has become the focus of an extensive literature in macroeconomics (e.g., see Barro 1999; Bernheim et al. 2015; Harris and Laibson 2001, 2013; Krusell et al. 2010; Krusell and Smith 2003; Laibson 1997).

The classical approach to studying the existence of time-consistent optimal plans for these problems has emphasized the language of recursive decision theory, as was discussed in the original paper by Strotz (1955). Unfortunately, as is well known, optimal dynamically consistent (including Markov) plans for such models need not exist, so the question of sufficient conditions for the existence of time-consistent optimal plans is a question of a great deal of study (e.g., see Pollak (1968), Peleg and Yaari (1973), and Caplin and Leahy (2006) for discussions of the nonexistence question).

One reason time-consistent plans may be nonexistent lies in the seemingly inherent presence of discontinuities in intertemporal preferences that arise very naturally in these problems when the recursive decision theory approach is applied. The reason for this lack of continuity is found in the inherent lack of commitment between the current "versions" of the dynamic decision-maker and all her continuation "selves." For example, from a decision theoretic perspective, when a "current" decision-maker is indifferent between some alternatives in the future, the earlier decision-maker ("planner") can still strictly prefer one of such alternatives in advance. As a result, he is willing to commit, yet lack access to a reasonable "commitment device" that would impose discipline on the choices of her future "selves" when tomorrow actually arrives. Due to this discontinuity, the optimal level of “commitment” may be nonexistent, and the dynamic maximization problem can turn out to be poorly defined (see, for example, Caplin and Leahy (2006) for an excellent discussion of this point).

An alternative way of obtaining a set of consistent plans for a dynamic choice problem with hyperbolic discounting is to view the dynamic choice problem as a dynamic game among different generations of “selves.” In this formulation of the decision problem, at any current period, the current “self” takes as given a set of continuation strategies of all her “future selves” and best responds to this continuation structure in the game. For example, in the context of optimal growth, one could search for Markov perfect equilibrium in this dynamic game played between successive “selves.” This is the approach advocated in the early work of Peleg and Yaari (1973) and in subsequent work by Laibson (1997), Barro (1999), Harris and Laibson (2001, 2013), Krusell and Smith (2003), Krusell et al. (2010), Balbus et al. (2015d), Balbus and Woźny (2016), and Bernheim et al. (2015). In this setting, one could take a candidate pure strategy continuation policy for savings/investment of one’s future “self” as given, generate a value from the program from tomorrow onward, and given this value function could determine an optimal savings/investment decision problem for the current self. A fixed point in this mapping between continuation savings/investment and current savings/investment would be a Markov perfect equilibrium.

The problem is finding a space with sufficient continuity to study this fixed point problem. For example, if you take the continuation decision on savings/investment as continuous, the value function it generates need not be concave in the income state; this then means the current decision problem is not concave (hence, the best reply correspondence does not generally admit a continuous selection). If the continuation policy is only semicontinuous, then the current generations best reply correspondence need not contain a semicontinuous selection. So finding sufficient continuity for the existence of even pure strategy Markov perfect plans is problematic. Similar issues arise when considering subgame perfect equilibrium.[7] Finally, when Markovian time-consistent plans do exist, they are difficult to characterize and compute, as these models often suffer from an indeterminacy of equilibria (e.g., Krusell and Smith 2003).

Perhaps the most well-studied version of hyperbolic discounting involves models where preferences exhibit quasi-hyperbolic discounting. In the quasi-hyperbolic model, agents have “$\beta - \delta$” preferences, where they have a “long-run” discount rate of $\delta \in (0, 1)$, and a “short-run” discount rate of $\beta \in (0, 1)$. In such models, agents have changing preferences, where at each period the preferences exhibit a bias toward current consumption. Such preferences often lead to an

[7] It bears mentioning that this continuity problem is related to difficulties that one finds in looking for continuity in best reply maps of the stage game given a continuation value function. It was explained nicely in the survey by Mirman (1979) for a related dynamic game in the context of equilibrium economic growth without commitment. See also the non-paternalistic altruism model first discussed in Ray (1987).

important role for public policy (e.g., Krusell et al. 2002, 2010). One class of models where the introduction of quasi-hyperbolic discounting has been shown to be important are models of asset accumulation (e.g., see the series of papers by Laibson (1994, 1997), Harris and Laibson (2001, 2013), as well as the recent paper by Bernheim et al. 2015). In these papers, the authors have shown using various methods that Markovian equilibrium savings behavior of models where agents have dynamically inconsistent preferences differ a great deal from models with standard time separable, dynamically consistent preferences. It is well known that in models with present-bias, savers consume more as a fraction of income than in models with dynamically consistent, time-separable preferences (also, see Diamond and Koszegi (2003) for examples of this in overlapping generations/life cycle models). In a very important paper, Laibson (1997) showed that in a standard asset accumulation model where agents possess preferences with quasi-hyperbolic preferences, and models enhanced with illiquid assets, the impact of present-bias preference can be mitigated by the presence of the illiquid asset. Indeed, illiquidity of assets can help constrain time-inconsistent behavior by working as a commitment device. His work suggests that financial innovation, therefore, can have a profound influence on equilibrium savings rates.

These models have also been used in the study of equilibrium economic growth. For example, Barro (1999) shows that in a version of the optimal growth model, under full commitment, and isoelastic period utility, agents save more and consume less; under imperfect commitment, saving rates and capital accumulation are lower. Krusell and Smith (2003) study a version of the optimal growth model and find additionally there exists a continuum of Markovian equilibria in their model without commitment. Krusell et al. (2002) produce a very interesting result for a particular parametric class of models. In particular, they show that for this particular parametric case, social planning solutions are strictly worse in welfare terms than a recursive equilibrium solution.

Extensions of this work of dynamic inconsistency in dynamic models have been numerous. The paper by O'Donoghue and Rabin (1999) extends class of Strotzian models to encompass models of procrastination. In their model, decision-makers are sophisticated or naive about their future structure of preferences (i.e., the nature of their future self-control problem), must undertake a single activity, and face intermediate costs and rewards associated with this activity. In the baseline model, "naive" decision-makers suffer procrastination ("acting too late") about undertaking a future activities with intermediate cost, while they act too soon relative to activities with intermediate future rewards. Sophistication about future self-control problems mitigates procrastination problems associated with dynamic inconsistency, while it makes the problem of preproperation ("acting too early") worse. In O'Donoghue and Rabin (2001), they extend this model to more general choice problems (with "menus" of choices).

In another line of related work, Fudenberg and Levine (2006) develop a "dual-selves" model of dynamically inconsistent choice and show that this model can explain both the choice in models with dynamically inconsistent preferences (e.g., Strotz/Laibson "$\beta - \delta$" models) and the O'Donoghue/Rabin models of procrastination. In their paper, they model the decision-maker as a "dual self," one being a long-run decision-maker, and a sequence of short-run myopic decision-makers, the dual self sharing preferences and playing a stage game.

There have been many different approaches in the literature to solve this problem. One approach is a recursive decision theory (Caplin and Leahy 2006; Kydland and Prescott 1980). In this approach, one attempts to introduce additional (implicit) constraints on dynamic decisions in a way that enforces time consistency. It is known that such decision theoretic resolutions in general can fail in some cases (e.g., time-consistent solutions do not necessarily exist).[8] Alternatively, one can view time-consistent plans as sequential (or subgame perfect) equilibrium in a dynamic game between successive generations of "selves." This was the approach first proposed in Pollak (1968) and Peleg and Yaari (1973). The set of subgame perfect equilibria in the resulting game using strategic dynamic programming methods is studied in the papers of Bernheim et al. (2015) and Balbus and Woźny (2016). The existence and characterization of Markov perfect stationary equilibria is studied in Harris and Laibson (2001), Balbus and Nowak (2008), and Balbus et al. (2015d, 2016). In the setting of risk-sensitive control, Jaśkiewicz and Nowak (2014) have studied the existence of Markov perfect stationary equilibria.

2.2 Economic Growth Without Commitment

Models of economic growth without commitment provide another important example of dynamic and stochastic games in macroeconomics. These models have arisen in many forms since the pioneering papers of Phelps and Pollak (1968), Peleg and Yaari (1973), Ray (1987), and Levhari and Mirman (1980).[9] For example, consider the model of altruistic growth without commitment as first described in Phelps and Pollak (1968) and Peleg and Yaari (1973) and extended in the work of Bernheim and Ray (1983), Leininger (1986), Amir (1996b), and Nowak (2006c). The model consists of a sequence of identical generations, each living one period, deriving utility from its own consumption, as well as the consumption of its successor generations. In any period of the economy, the current generation begins the period with a stock of output goods which it must either consume or invest in a technology that reproduces the output good tomorrow. The reproduction problem can be either deterministic or stochastic. Finally, because of the demographic structure of the model, there is no commitment assumed between generations. In this model, each generation of the dynastic household cares about the consumption

[8] For example, see Peleg and Yaari (1973), Bernheim and Ray (1983), and Caplin and Leahy (2006).

[9] For example, models of economic growth with strategic altruism under perfect commitment have also been studied extensively in the literature. For example, see Laitner (1979a,b, 1980, 2002), Loury (1981), and including more recent work of Alvarez (1999). Models of infinite-horizon growth with strategic interaction (e.g., "fishwars") are essentially versions of the seminal models of Cass (1965) and Brock and Mirman (1972), but without commitment.

of the continuation generation of the household, but it cannot control what the future generations choose. Further, as the current generation only lives a single period, it has an incentive to deviate from a given sequence of bequests to the next generation by consuming relatively more of the current wealth of the household (relative to, say, past generations) and leaving little (or nothing) of the dynastic wealth for subsequent generations. So the dynastic household faces a time-consistent planning problem.

Within this class of economies, conditions are known for the existence of semicontinuous Markov perfect stationary equilibria, and these conditions have been established under very general conditions via nonconstructive topological arguments (e.g., for deterministic versions of the game, in Bernheim and Ray 1987; Leininger 1986), and for stochastic versions of the game, by Amir (1996b), Nowak (2006c), and Balbus et al. (2015b,c). It bears mentioning that for stochastic games, existence results in spaces of continuous functions have been obtained in these latter papers. In recent work by Balbus et al. (2013), the authors give further conditions under which sharp characterizations of the set of pure strategy Markov stationary Nash equilibria (MSNE, henceforth) can be obtained. In particular, they show that the set of pure strategy MSNE forms an antichain, as well as develop sufficient conditions for the uniqueness of Markov perfect stationary equilibrium. This latter paper also provides sufficient conditions for globally stable approximate solutions relative to a unique nontrivial Markov equilibrium within a class of Lipschitz continuous functions. Finally, in Balbus et al. (2012), these models are extended to settings with elastic labor supply.

It turns out that relative to the set of subgame perfect equilibria, strategic dynamic programming methods can also be developed for these types of models (e.g., see Balbus and Woźny 2016).[10] This is interesting as APS type methods are typically only used in situations where players live an infinite number of periods. Although the promised utility approach has proven very useful in even this context, for models with altruistic growth without commitment, they suffer from some well-known limitations and complications. First, they need to impose discounting typically in this context. When studying the class of Markov perfect equilibria using more direct (fixed point) methods, one does not require this. Second, and more significantly, the presence of "continuous" noise in our class of dynamic games proves problematic for existing promised utility methods. In particular, this noise introduces significant complications associated with the measurability of value correspondences that represent continuation structures (as well as the possibility of constructing and characterizing measurable selections which are either equilibrium value function or pure strategies). We will discuss how this can be handled in versions of this model with discounting. Finally, characterizations of pure strategy equilibrium values (as well as implied pure strategies) is also difficult to obtain. So in this context, more direct methods studying the set of Markov perfect stationary equilibria can provide sharper characterizations of equilibria. Finally, it can be difficult to use promised utility continuation methods to obtain any characterization of the long-run

[10]Also see Balbus et al. (2012) section 5 for a discussion of these methods for this class of models.

stochastic properties of stochastic games (i.e., equilibrium invariant distributions or ergodic distributions).[11]

There are many other related models of economic growth without commitment that have also appeared in the literature. For example, in the paper of Levhari and Mirman (1980), the authors study a standard model of economic growth with many consumers but without commitment (the so-called great fishwar). In this model, in each period, there is a collective stock of output that a finite number of players can consume, with the remaining stock of output being used as an input into a productive process that regenerates output for the next period. This regeneration process can be either deterministic (e.g., as in Levhari and Mirman 1980 or Sundaram 1989a) or stochastic (as in Amir 1996a; Nowak 2006c).

As for results on these games in the literature, in Levhari and Mirman (1980), the authors study a parametric version of this dynamic game and prove existence of unique Cournot-Nash equilibrium. In this case, they obtain unique smooth Markov perfect stationary equilibria. In Sundaram (1989a), these results are extended to symmetric semicontinuous Markov perfect stationary equilibria in the game, but with more standard preferences and technologies.[12] Many of these results have been extended to more general versions of this game, including those in Fischer and Mirman (1992) and Fesselmeyer et al. (2016). In the papers of Dutta and Sundaram (1992) or Amir (1996a), the authors study stochastic versions of these games. In this setting, they are able to obtain the existence of continuous Markov perfect stationary Nash equilibrium under some additional conditions on the stochastic transitions of the game.

2.3 Optimal Policy Design Without Commitment

Another macroeconomic model where the tools of dynamic game theory play a critical role are models of optimal policy design where the government has limited commitment. In these models, again the issue of dynamic inconsistency appears. For example, there is a large literature studying optimal taxation problem in models under perfect commitment (e.g., Chamley 1986; Judd 1985. See also Straub and Werning 2014). In this problem, the government is faced with the problem of financing dynamic fiscal expenditures by choosing history-contingent paths for future taxation policies over capital and labor income under balanced budget constraints. When viewing the government as a dynastic family of policymakers, they collectively face a common agreed upon social objective (e.g., maximizing the representative agent's objective function along sequential equilibrium paths for the private economy). As mentioned in the introduction, this problem is often studied under limited commitment (e.g., in Kydland and Prescott 1980; Pearce and Stacchetti 1997; Phelan and Stacchetti 2001, and more recently Dominguez and

[11] For competitive economies, progress has been made. See Peralta-Alva and Santos (2010).

[12] See also the correction in Sundaram (1989b).

Feng 2016a,b, and Feng 2015). As the objective function in this class of models is generally not time consistent, the question of credible optimal government policy immediately arises.

The existence of time-consistent optimal plans for capital and labor income taxes was first studied in Kydland and Prescott (1980). In their formulation of the problem, the game was essentially a dynamic Stackelberg game, that is, in the "first stage," the agents in the private economy take sequences of tax instruments and government spending as given, and a sequential equilibrium for the private economy is determined. Then, in the second stage, this sequential equilibrium induces dynastic social preferences of government policymakers over these sequences of tax instruments and government spending (under a balanced budget rule). These preferences are essentially the discounted lifetime utility of a representative agent, and are maximized over the government's fiscal choice, which is not time consistent (therefore, as successive generations of government policymakers possess limited commitment across time, announced future plans will not necessarily be implemented by future government policymakers). To illustrate the basic problems of their model, we consider the following example:

Example 1. Consider a two-period economy with preferences given by

$$u(c_1) + \delta(u(c_2) + \gamma u(g)),$$

with linear production, full depreciation, and initial capital $k_0 > 0$. Then the economy's resource constraint is $c_1 + k = k_0$ and $c_2 + g = k$, where g is a public good level. Suppose that $\delta = 1$, $\gamma = 1$, and $u(c) = \log(\alpha + c)$ for some $\alpha > 0$.

The optimal, dictatorial solution (benevolent government choosing nonnegative k and g) to the welfare maximization problem is given by FOC:

$$u'(k_0 - k) = u'(k - g) = u'(g),$$

which gives $2g = k = \frac{2}{3}k_0$ with $c_1 = c_2 = g = \frac{1}{3}k_0$.

Now consider a competitive equilibrium economy, where the government finances public good g by levying a linear tax $\theta \in [0, 1]$ on capital income. The household budget is $c_1 + k = k_0$ and $c_2 = (1 - \theta)k$. Suppose that consumers have rational expectations and we look for a credible tax level θ under the balanced budget condition $g = \theta k$. For this reason suppose that θ is given and solve for competitive equilibrium investment k. The FOC gives:

$$u'(k_0 - k) = (1 - \theta)u'((1 - \theta)k),$$

which gives

$$k(\theta) = \frac{(1 - \theta)(\alpha + k_0) - \alpha}{2(1 - \theta)},$$

with $k'(\theta) < 0$. Now, knowing this reaction curve, the government chooses the optimal tax level solving the competitive equilibrium welfare maximization problem:

$$\max_{\theta \in [0,1]} u(k_0 - k(\theta)) + u((1-\theta)k(\theta)) + u(\theta k(\theta)).$$

Here the first-order condition requires

$$[-u'(k_0 - k(\theta)) + (1-\theta)u'((1-\theta)k(\theta))]k'(\theta) + [-u'((1-\theta)k(\theta)) \\ +u'(\theta k(\theta))]k(\theta) + \theta u'(\theta k(\theta))k'(\theta) = 0.$$

The last term (strictly negative) is the credibility adjustment which distorts the dynamically consistent solution from the optimal one. It indicates that in the dynamically consistent solution, when setting the tax level in the second period, the government must look backward for its impact on the first-period investment decision.

Comment: to achieve the optimal, dictatorial solution the government would need to promise $\theta = 0$ in the first period (so as not to distort investment) but then impose $\theta = \frac{1}{2}$ to finance the public good. Clearly it is a dynamically inconsistent solution.

This problem has led to a number of different approaches to solving it. One idea, found in the original paper of Kydland and Prescott (1980), was to construct optimal policy rules that respect a "backward"-looking endogenous constraint on future policy. This, in turn, implies optimal taxation policies must be defined on an enlarged set of (endogenous) state variables. That is, without access to a "commitment device" for the government policymakers, for future announcements about optimal policy to be credible, fiscal agents must constrain their policies to depend on additional endogenous state variables. This is the approach that is also related to the recursive optimization approaches of Rustichini (1998a), Marcet and Marimon (1998), and Messner et al. (2012), as well as the generalization of the original Kydland and Prescott method found in Feng et al. (2014) that is used to solve this problem in the recent work of Feng (2015).

Time-consistent polices can also be studied as a sequential equilibrium of a dynamic game between successive generations to determine the optimal mixture of policy instruments, where commitment to planned future policies is guaranteed in a sequential equilibrium in this dynamic game between generations of policymakers. These policies are credible optimal policies because these policies are subgame perfect equilibrium in this dynamic game. See also Chari et al. (1991, 1994) for a related discussion of this problem. This is the approach taken in Pearce and Stacchetti (1997), Phelan and Stacchetti (2001), Dominguez and Feng (2016a,b), among others.

Simple optimal policy problems also arise in the literature that studies optimal monetary policy rules; similar papers have been written in related macroeconomic models. This literature began with the important papers of Fischer (1980b) and Barro and Gordon (1983) (e.g., see also Rogoff (1987) for a nice survey of this work). More recent work studying the optimal design of monetary policy under limited commitment includes the papers of Chang (1998), Sleet (2001), Athey et al. (2005), and Sleet and Yeltekin (2007).

3 Strategic Dynamic Programming Methods

In this section, we lay out in detail the theoretical foundations of strategic dynamic programming methods for repeated and dynamic/stochastic games.

3.1 Repeated Models

An original strategic dynamic programming method was proposed by Abreu et al. (1986) and further developed in Abreu et al. (1990) for a class of repeated games with imperfect public information and perfect public equilibria. As the game is repeated, the original APS methods did not have "states" (e.g., in addition to promised utility). These methods have been used in macroeconomics (especially in dynamic contract theory, but also in policy games when considering the question of sustainable optimal monetary policy (e.g., see Chang 1998).[13]

Consider an infinitely repeated game between n-players with imperfect public information. Let $N = \{1, 2, \ldots, n\}$ be a set of players. In each period, each of n-players chooses simultaneously a strategy so that the strategy profile is $a = (a_1, a_2, \ldots, a_n) \in A$, where $A = \times_{i=1}^{n} A_i$, i.e., a Cartesian product of individual action sets. Each $a \in A$ induces a distribution over the realization of publicly observable signals $y \in Y$, where $Y \subset \mathbb{R}^k$ ($k \in \mathbb{N}$) given by $Q(dy|a)$. Each player $i \in N$ has a one-stage payoff given by $u_i(y, a_i)$, and its expectation is $g_i(a) := \int_Y u_i(y, a_i) Q(dy|a)$.

Remark 1. A repeated game with observable actions is a special case of this model, if $Y = A$ and $Q(\{a\}|a) = 1$ and zero otherwise.

For each $t > 1$, let H_t be a public history at the beginning of period t. Mathematically, it is a sequence of the signals before t, i.e., $H_t := Y^t$ with generic element $h^t := (y_0, y_1, \ldots, y_{t-1})$. A public strategy of player i is a sequence of functions $\sigma_i := (\sigma_{i,t})_{t=0}^{\infty}$, where each $\sigma_{i,t}$ maps histories H_t to probability

[13] Also, for repeated games with quasi-hyperbolic discounting, see Chade et al. (2008) and Obara and Park (2013).

distributions on A_i. A strategy $\sigma_{i,t}$ is pure if it maps histories H_t into A_i. A strategy profile is a product of strategies, i.e., $\sigma := (\sigma_t)_{t=0}^{\infty}$, where $\sigma_t := (\sigma_{1,t}, \ldots, \sigma_{n,t})$.

Let $H := Y^{\infty}$ be a set of all public histories with generic element $h := (y_0, y_1, \ldots)$. By Ionescu-Tulcea theorem, a transition probability Q and strategy σ induce the unique Borel probability measure on H. Let E^{σ} be an expectation associated with this measure.

Assume a common discount factor $\delta \in (0, 1)$; then the player i's expected payoff from the repeated game is given by:

$$U_i(\sigma) := (1-\delta)E^{\sigma}\left(\sum_{t=0}^{\infty} \delta^t g_i(\sigma_t(h^t))\right),$$

where $(1-\delta)$ normalization is used to make payoffs of the stage game and infinitely repeated game comparable.

We impose the following set of assumptions:

Assumption 1. (i) *A_i is finite for each $i \in N$,*

(ii) *for each $a \in A$, $Q(\cdot|a)$ is absolutely continuous probability measure with density $q(\cdot, a)$,*

(iii) *the support of $Q(\cdot|a)$ is independent of a, and without loss of generality assume that it is Y. That is*

$$Y := \{y \in Y : q(y|a) > 0, \textit{ for all } a \in A\},$$

(iv) *for each $i \in N$ and $a_i \in A_i$, $u_i(\cdot, a_i)$ is a continuous function,*

(v) *the one-shot strategic form game $(N, (A_i, g_i)_{i \in N})$ has a pure strategy Nash equilibrium.*

Let $\mathcal{V} := L^{\infty}(Y, \mathbb{R}^n)$ be a set of all equivalence classes of essentially bounded Lebesgue measurable functions from Y into $\mathbb{R}^n$. Endow $\mathcal{V}$ with its weak star topology. Similarly, denote the measurable functions from Y to any subsets of $\mathbb{R}^n$. Moreover, with a slight abuse of notation, we will denote the i-th component of $v \in \mathcal{V}$ by $v_i : Y \to \mathbb{R}$, and hence $v = (v_1, v_2, \ldots, v_n) \in \mathcal{V}$.

A standard tool to deal with discounted n-player repeated games is the class of one-shot auxiliary (strategic form) games $\Gamma(v) = (N, (A_i, \Pi_i(v_i))_{i \in N})$, where

$$\Pi_i(a)(v_i) = (1-\delta)g_i(a) + \delta \int_Y v_i(y) Q(dy|a)$$

is player i's payoff. Let $W \subset \mathbb{R}^n$. By $B(W)$ denote the set of all Nash equilibrium payoffs of the auxiliary game for some vector function $v \in \mathcal{V}$ having image in W. Formally:

$$B(W) := \{w \in \mathbb{R}^n : \text{ there is } a^* \in A, v \in \mathcal{V} \text{ such that}$$

$$\Pi_i(a^*)(v_i) \geq \Pi_i(a^*_{-i}, a_i)(v_i),$$

$$\text{for all } a_i \in A_i, i \in N, \text{ and } v(y) \in W, \text{ for almost all } y \in Y\}.$$

By the axiom of the choice, there is an operator $\eta : B(W) \to L^\infty(Y, W)$ and $\xi : W \to A$ such that for each $i \in N$ it holds

$$w_i = \Pi_i(\xi(w))(\eta(w, \cdot)) \geq \Pi_i(\xi_{-i}(w), a_i)(\eta(w, \cdot)).$$

Modifying on null sets if necessary, we may assume that $\eta(w, y) \in W$ for all y. We say that $W \subset \mathbb{R}^n$ is self-generating if $W \subset B(W)$. Denoting by $V^* \subset \mathbb{R}^n$ the set of all public perfect equilibrium vector payoffs and using self-generation argument one can show that $B(V^*) = V^*$. To see that we proceed in steps.

Lemma 1. *If W is self-generating, then $W \subset V^*$.*

Self-generation is an extension of the basic principle of optimality from dynamic programming. Let W be some self-generating set. Then, if $w \in W$, by self-generation, $w \in B(W)$. Consequently, we may find a sequence of functions $(v^t)_{t=1}^\infty$ such that $v^t : Y^t \to W$, for each $t > 1$ such that $v^1(y) := \eta(w, y)$ and for $t > 1$ $v^t(y_1, y_2, \ldots, y_t) := \eta(v^{t-1}(y_1, \ldots, y_{t-1}), y_t)$ and a sequence of functions $(\sigma_t)_{t=0}^\infty$ such that $\sigma_1 := \xi(w)$ and for $t > 0$ $\sigma_{t+1}(y_1, \ldots, y_t) := \xi(\eta(v^t(y_1, \ldots, y_t)))$. We claim that $\sigma := (\sigma_t)_{t=0}^\infty$ is a perfect Nash equilibrium in public strategies and $w_i = U_i(\sigma)$. Indeed, if player i deviates from σ until time T, choosing $\tilde{a}_i^t$ instead of a_i^t, then by definition of ξ and η:

$$w_i = U_i(\sigma) \geq (1 - \delta) \sum_{t=1}^{T} \delta^t g_i(\tilde{a}_i^t) + \delta^{T+1} U_i\left(J^T(\sigma)\right).$$

Here $J^T(\sigma) := (\sigma_{T+1}, \sigma_{T+2}, \ldots)$. Taking a limit with $T \to \infty$, we may conclude σ is a perfect Nash equilibrium in public strategies and $w \in V^*$. To formalize this thinking, we state the next theorem. Let $V \subset \mathbb{R}^n$ be some large set of possible payoffs, such that $V^* \subset V$. Then:

Theorem 1. *Suppose that Assumption 1 holds. Then,*

(i) $\bigcap_{t=1}^{\infty} B^t(V) \neq \emptyset$,

(ii) $\bigcap_{t=1}^{\infty} B^t(V)$ *is the greatest fixed point of B,*

(iii) $\bigcap_{t=1}^{\infty} B^t(V) = V^*$.

To see (i) of the aforementioned theorem, observe that V is a nonempty compact set. By Assumption 1 (ii), (iv) and (v), we may conclude that $B(V)$ is nonempty compact set and consequently that each $B^t(V)$ is nonempty and compact. Obviously, B is an increasing operator, mapping V into itself. $B^t(V)$ is a decreasing sequence; hence, its intersection is not empty. To see (ii) and (iii) of this theorem, observe that all sets $B^t(V)$ include any fixed point of B and, consequently, its intersection also. On the other hand, $\bigcap_{t=1}^{\infty} B^t(V)$ is self-generating, hence by Lemma 1

$$\bigcap_{t=1}^{\infty} B^t(V) \subset V^*. \tag{17.1}$$

By Assumption 1 (iii), we may conclude that V^* is self-generating; hence, $B(V^*) \subset V^*$. Consequently, $V^* = B(V^*)$; hence, $V^* \subset \bigcap_{t=1}^{\infty} B^t(V)$. Together with (17.1), we have points (ii) and (iii). Moreover, observe that V^* is compact. To see that, observe that V^* is bounded and its closure $cl(V^*)$ is compact. On the other hand, $V^* = B(V^*) \subset B(cl(V^*))$. By Assumption 1 (ii) and (iv), we have compactness of $B(cl(V^*))$ and consequently $cl(V^*) \subset B(cl(V^*))$; hence, by Lemma 1, $cl(V^*) \subset V^*$. As a result, V^* is closed and hence compact.

An interesting property of the method is that the equilibrium value set can be characterized using some extremal elements of the equilibrium value set only. Abreu et al. (1990) call it a bang-bang property. Cronshaw and Luenberger (1994) (and Abreu (1988) for some early examples) push this fact to the extreme and show that the equilibrium value set of a strongly symmetric subgame perfect equilibrium can be characterized using the worst punishment only. This observation has important implications on computation algorithms and applications.[14]

For each $W \subset \mathbb{R}^n$, let $co(W)$ be a convex hull of W. By $ext(W)$, we denote the set of extreme points of $co(W)$.

Definition 1. We say that the function $v \in L^\infty(Y, W)$ has bang-bang property if $v(y) \in ext(W)$ for almost all $y \in Y$.

Using Proposition 6.2 in Aumann (1965), we have:

Theorem 2. *Let $W \subset \mathbb{R}^n$ be a compact set. Let $a^* \in A$ and $v \in L^\infty(Y, co(W))$ be chosen such that a^* is Nash equilibrium in the game $\Gamma(v)$. Then, there exists a function $\tilde{v} \in L^\infty(Y, ext(W))$ such that a^* is Nash equilibrium of $\Gamma(\tilde{v})$.*

[14] See Dominguez (2005) for an application to models with public dept and time-consistency issues, for example.

Corollary 1. *If $W \subset \mathbb{R}^n$ is compact, then $B(W) = B(ext(W))$.*

Theorem 2 and its corollary show that we may choose $\eta(w, \cdot)$ to have bang-bang property. Moreover, if that continuation function has bang-bang properties, then we may easily calculate continuation function in any step. Especially, if Y is a subset of the real line, the set of extreme points is at most countable.

Finally, Abreu et al. (1990) present a monotone comparative statics result in the discount factor. The equilibrium value set V^* is increasing in the set inclusion order in δ. That is, the higher the discount factor, the larger is the set of attainable equilibrium values (as cooperation becomes easier).

3.2 Dynamic and Stochastic Models with States

We now consider an n-player, discounted, infinite horizon, stochastic game in discrete time. This is the basic APS tool used in numerous applications in macroeconomics (e.g., all the examples discussed in Sect. 2, but others too). Along these lines, consider the primitives of a class of stochastic games given by the tuple:

$$\left\{S, (A_i, \tilde{A}_i, \delta_i, u_i)_{i=1}^N, Q, s_0\right\},$$

where S is the state space, $A_i \subset \mathbb{R}^{k_i}$ is player i's action space with $A = \times_i A_i$, $\tilde{A}_i(s)$ the set of actions feasible for player i in state s, δ_i is the discount factor for player i, $u_i : S \times A \to \mathbb{R}$ is the one-period payoff function, Q denotes a transition function that specifies for any current state $s \in S$ and current action $a \in A$, a probability distribution over the realizations of the next period state $s' \in S$, and finally $s_0 \in S$ is the initial state of the game. We assume that $S = [0, \bar{S}] \subset \mathbb{R}$ and that $\tilde{A}_i(s)$ is a compact Euclidean subset of $\mathbb{R}^{k_i}$ for each s, i.

Remark 2. A dynamic game is a special case of this model, if Q is a deterministic transition.

Using this notation, a formal definition of a (Markov, stationary) strategy, payoff, and a Nash equilibrium can be stated as follows. A set of all possible histories of player i till period t is denoted by H_i^t. An element $h_i^t \in H_i^t$ is of the form $h_i^t = (s_0, a_0, s_1, a_1, \ldots, a_{t-1}, s_t)$. A *pure strategy* for a player i is denoted by $\sigma_i = (\sigma_{i,t})_{t=0}^{\infty}$ where $\sigma_{i,t} : H_i^t \to A_i$ is a measurable mapping specifying an action to be taken at stage t as a function of history, such that $\sigma_{i,t}(h_i^t) \in \tilde{A}_i(s_t)$. If, for some t and history $h_i^t \in H_i^t$, $\sigma_{i,t}(h_i^t)$ is a probability distribution on $\tilde{A}(s_t)$, then we say σ_i is a *behavior strategy*. If a strategy depends on a partition of histories limited to the current state s_t, then the resulting strategy is referred to as *Markov*. If for all stages t, we have a Markov strategy given as $\sigma_{i,t} = \gamma_i$, then a strategy identified with γ_i for player i is called a *Markov stationary strategy* and denoted simply by γ_i. For a strategy profile $\sigma = (\sigma_1, \sigma_2, \ldots, \sigma_n)$, and initial state $s \in S$, the expected payoff for player i can be denoted by:

$$U_i(\sigma, s_0) = (1 - \delta_i) \sum_{t=0}^{\infty} \delta_i^t \int u_i(s_t, \sigma_t(h^t)) dm_i^t(\sigma, s_0),$$

where m_i^t is the stage t marginal on A_i of the unique probability distribution (given by Ionescu-Tulcea theorem) induced on the space of all histories for σ. A strategy profile $\sigma^* = (\sigma_{-i}^*, \sigma_i^*)$ is a *Nash equilibrium* if and only if σ^* is feasible, and for any i, and all feasible σ_i, we have

$$U_i(\sigma_{-i}^*, \sigma_i^*, s_0) \geq U_i(\sigma_{-i}^*, \sigma_i, s_0).$$

Assumption 2. (i) *S is a standard Borel space,*

(ii) *A_i is a separable metric space and $\tilde{A}$ is a compact-valued measurable correspondence,*

(iii) *Each u_i is a uniformly bounded and jointly measurable function such that for each $s \in S$, $u_i(s, \cdot)$ is continuous on $\tilde{A}(s)$,*

(iv) *For each Borel measurable subset D of S, $(s, a) \mapsto Q(D|s, a)$ is jointly measurable and for each $s \in S$*

$$\lim_{n \to \infty} \sup_D |Q(D|s, a_n) - Q(D|s, a)| = 0$$

whenever $a_n \to a$.

When dealing with discounted n-player dynamic or stochastic games, the main tool is again the class of one-shot auxiliary (strategic form) games $\Gamma_s(v) = (N, (A_i, \Pi_i(s, \cdot)(v_i))_{i \in N})$, where $s \in S \subset \mathbb{R}^n$ is the current state, while $v = (v_1, v_2, \ldots, v_n)$, where each $v_i : S \to \mathbb{R}$ is the integrable continuation value and the payoffs are given by:

$$\Pi_i(s, a)(v_i) = (1 - \delta_i) u_i(s, a) + \delta_i \int_S v_i(s') Q(ds'|s, a).$$

Then, by $K \subset \mathbb{R}^n$ denote some initial compact set of attainable payoff vectors and consider the large compact valued correspondence $V : S \rightrightarrows K$. Let $W : S \rightrightarrows K$ be any correspondence. By $B(W)(s)$ denote the set of all payoff vectors of $\Gamma_s(v)$ in K, letting v varying through all integrable selections from W. Then showing that $B(W)$ is a measurable correspondence, and denoting by $a^*(s)(v)$ a Nash equilibrium of $\Gamma_s(v)$, one can define an operator B such that:

$$B(W)(s) := \{w \in K : \text{ there is integrable selection } v \text{ of } W$$

$$\text{and a measurable Nash equilibrium } a^*(s)(v) \text{ of } \Gamma_s(v) \text{ such that}$$

$$\text{for each } i \in N \text{ it holds } w_i = \Pi_i(s, a^*(s)(v))(v)\}.$$

It can be shown that B is an increasing operator; hence, starting from some large initial correspondence V_0, one can generate a decreasing sequence of sets $(V_t)_t$ (whose graphs are ordered under set inclusion) with $V_{t+1} = B(V_t)$. Then, one can show using self-generation arguments that there exists the greatest fixed point of B, say V^*. Obviously, as V is a measurable correspondence, $B(V)$ is a measurable correspondence. By induction, one can then show that all correspondences V_t are measurable (as well as nonempty and compact valued). Hence, by the Kuratowski and Ryll-Nardzewski selection theorem (Theorem 18.13 in Aliprantis and Border 2005), all of these sets admit measurable selections. By definition, $B(V^*) = V^*$; hence, for each state $s \in S$ and $w \in B(V^*)(s) \subset K$, there exists an integrable selection v'^* such that $w = \Pi(s, a^*(s)(v'))(v')$. Repeating this procedure in the obvious (measurable) way, one can construct an equilibrium strategy of the initial stochastic game. To summarize, we state the next theorem:

Theorem 3. *Suppose that Assumption 2 holds. Then,*

(i) $\bigcap_{t=1}^{\infty} B^t(V) \neq \emptyset$,

(ii) $\bigcap_{t=1}^{\infty} B^t(V)$ *is the greatest fixed point of* B,

(iii) $\bigcap_{t=1}^{\infty} B^t(V) = V^*$,

where V^* is the set of all values of subgame perfect behavior strategies.

The details of the argument are developed in Mertens and Parthasarathy (1987), restated in Mertens and Parthasarathy (2003), and nicely summarized by Mertens et al. (2015) (pages 397–398). See also Fudenberg and Yamamoto (2011) for similar concepts used in the study of irreducible stochastic games with imperfect monitoring, or Hörner et al. (2011) with a specific and intuitive characterization of equilibria payoffs of irreducible stochastic games, when discount factor tends to 1. See also Baldauf et al. (2015) for the case of a finite number of states.

3.3 Extensions and Discussion

The constructions presented in Sects 3.1 and 3.2 offer the tools needed to analyze appropriate equilibria of repeated, dynamic, or stochastic games. The intuition, assumptions, and possible extensions require some comments, however.

The method is useful to prove existence of a sequential or subgame perfect equilibrium in a dynamic or stochastic economy. Further, when applied to macroeconomic models, where Euler equations for the agents in the private economy are available, in other fields like in economics (e.g., industrial organization, political economy, etc.), the structure of their application can be modified (e.g., see Feng et al. (2014) for an extensive discussion of alternative choices of state variables). In all cases, when the method is available, it allows one to characterize the entire set of equilibrium values, as well as giving a constructive method to compute them.

Specifically, the existence of some fixed point of B is clear from Tarski fixed point theorem. That is, an increasing self-map on a nonempty complete lattice has a nonempty complete lattice of fixed points. In the case of strategic dynamic programming, B is monotone by construction under set inclusion, while the appropriate nonempty complete lattice is a set of all bounded correspondences ordered by set inclusion on their graphs (or simply value sets for a repeated game).[15] Further, under self-generation, it is only the largest fixed point of this operator that is of interest. So the real value added of the theorems, when it comes to applications, is characterization and computation of the greatest fixed point of B. Again, it exists by Tarski fixed point theorem.

However, to obtain convergence of iterations on B, one needs to have stronger continuity type conditions. This is easily obtained, if the number of states S (or Y for a repeated game) is countable, but typically requires some convexification by sunspots of the equilibrium values, when dealing with uncountably many states. This is not because of the fixed point argument (which does not rely on convexity); rather, it is because the weak star limit belongs pointwise only to the convex hull of the pointwise limits. Next, if the number of states S is uncountable, then one needs to work with correspondences having *measurable* selections. Moreover, one needs to show that B maps into the space of correspondences having some measurable selection. This can complicate matters a good bit for the case with uncountable states (e.g, see Balbus and Woźny (2016) for a discussion of this point). Finally, some Assumptions in 1 for a repeated game or Assumption 2 for a stochastic game are superfluous if one analyzes particular examples or equilibrium concepts.

As already mentioned, the convexification step is critical in many examples and applications of strategic dynamic programming. In particular, convexification is important not only to prove existence in models with uncountably many states but also to compute the equilibrium set (among other important issues). We refer the reader to Yamamoto (2010) for an extensive discussion of a role of public randomization in the strategic dynamic programming method. The paper is important not only for discussing the role of convexification in these methods but also provides an example with a *non-convex* set of equilibrium values (where the result on monotone comparative statics under set inclusion relative to increases in the discount rate found in the original APS papers does not hold as well).

Next, the characterization of the entire set of (particular) equilibrium values is important as it allows one to rule out behaviors that are not supported by any equilibrium strategy. However, in particular games and applications, one has to construct carefully the greatest fixed point of B that characterizes the set of all equilibrium values obtained in the public perfect, subgame perfect, or sequential equilibrium. This requires assumptions on the information structure (for example,

[15] See Baldauf et al. (2015) for a discussion of this fact.

assumptions on the structure of private signals, the observability of chosen actions, etc.). We shall return to this discussion shortly.

In particular, for the argument to work, and for the definition of operator B to make sense, one needs to guarantee that for every continuation value v, and every state s, there exists a Nash equilibrium of one-shot game $\Gamma_s(v)$. This can be done, for example, by using mixed strategies in each period (and hence mapping to behavior strategies of the extensive form game). Extensive form mixed strategies in general, when players do possess some private information in sequential or subgame perfect equilibria, cannot always be characterized in this way (as they do not possess recursive characterization). To see that, observe that in such equilibria, the strategic possibilities at every stage of the game is not necessarily common knowledge (as they can depend arbitrarily on private histories of particular players). This, for example, is not the case of public perfect equilibria (or sequential equilibrium with *full support* assumption as required by Assumption 1 (iii)) or for subgame perfect equilibria in stochastic games with no observability of players' past moves.

Another important extension of the methods applied to repeated games with public monitoring and public perfect equilibria was proposed by Ely et al. (2005). They analyze the class of repeated games with private information but study only the so called "belief-free" equilibria. Specifically, they consider a strong notion of sequential equilibrium, such that the strategy is constant with respect to the beliefs on others players' private information. Similarly, as Abreu et al. (1990), they provide a recursive formulation of all the belief-free equilibrium values of the repeated game under study and provide its characterizations. Important to mention, general payoff sets of repeated games with private information lack such recursive characterization (see Kandori 2002).

It is important to emphasize that the presented method is also very useful when dealing with nonstationary equilibrium in macroeconomic models, where an easy extension of the abovementioned procedure allows to obtain comparable existence results (see Bernheim and Ray (1983) for an early example of this fact for an economic growth model with altruism and limited commitment). But even in stationary economies, the equilibria obtained using APS method are only stationary as a function of the current state and future continuation value. Put differently, the equilibrium condition is satisfied for a set or correspondence of values, but not necessarily its particular selection, [16] say $\{v^*\} = B(\{v^*\})$. To map it on the histories of the game and obtain stronger stationarity results, one needs to either consider (extensive form) correlated equilibria or sunspot equilibria or semi-Markov equilibria (where the equilibrium strategy depends on both current and the previous period states). To obtain Markov stationary equilibrium, one needs to either assume that the number of states and actions is essentially finite or transitions are nonatomic or concentrate on specific classes of games.

[16] However, Berg and Kitti (2014) show that this characterization is satisfied for (elementary) paths of action profiles.

One way of restricting to a class of nonstationary Markov (or conditional Markov) strategies is possible by a careful redefinition of an operator B to work in function spaces. Such extensions were applied in the context of various macroeconomic models in the papers of Cole and Kocherlakota (2001), Doraszelski and Escobar (2012), or Kitti (2013) for countable number of states and Balbus and Woźny (2016) for uncountably many states. To see that, let us first redefine operator B to map the set of bounded measurable functions $\mathscr{V}$ (mapping $S \to \mathbb{R}^N$) the following way. If $W \subset \mathscr{V}$, then

$$
\begin{aligned}
B^f(W) := \{&(w_1, w_2, \ldots, w_n) \in \mathscr{V} \text{ and} \\
&\text{for all } s, i \text{ we have } w_i(s) = \Pi_i(s, a^*(s)(v))(v_i), \text{ where} \\
&v = (v_1, v_2, \ldots, v_n) \in W \text{ and each } v_i \text{ is an integrable function}\}.
\end{aligned}
$$

Again one can easily prove the existence of and approximate the greatest fixed point of B^f, say V_f^*. The difference between B and B^f is that B^f maps between spaces of functions not spaces of correspondences. The operator B^f is, hence, not defined pointwise as operator B. This difference implies that the constructed equilibrium strategy depends on the current state and future continuation value, but the future continuation value selection is constant among current states. This can be potentially very useful when concentrating on strategies that have more stationarity structure, i.e., in this case, they are Markov but not necessarily Markov stationary, so the construction of the APS value correspondence is generated by sequential or subgame perfect equilibria with *short memory*.

To see that formally, observe that from the definition of B and characterization of V^*, we have the following:

$$
\begin{aligned}
&(\forall s \in S)(\forall \text{ number } w \in V^*(s))(\exists \text{ measurable function} \\
&\quad v'^* \text{ s.t. } w = \Pi(s, a^*(s)(v'))(v')).
\end{aligned}
$$

Specifically, observe that continuation function v' can depend on w and s, and hence we shall denote it by $v'_{w,s}$. Now, consider operator B^f and its fixed point V_f^*. We have the following property:

$$
\begin{aligned}
&(\forall \text{ function } w \in V^*)(\exists \text{ measurable function } v_f'^* \text{ s.t. } (\forall s \in S)\, w(s) \\
&= \Pi(s, a^*(s)(v'))(v')).
\end{aligned}
$$

Hence, the continuation v' depends on w only, and we can denote it by v'_w.

Observe that in both methods, the profile of equilibrium decision rules: $a^*(s)(v')$ is *generalized Markov*, as it is enough to know state s and continuation function v' to make an optimal choice. In some cases in macroeconomic applications, this generalized Markov equilibrium can be defined using envelope theorems of continuation values (not the value function itself).[17] In construction of B^f, however, the dependence on the current state is direct: $s \to a^*(s)(v'_w)$. So we can easily verify properties of the generalized Markov policy, such as whether it is continuous or monotone in s. In the definition of operator B, however, one has the following: $s \to a^*(s)(v'_{w,s})$. So even if the Nash equilibrium is continuous in both variables, (generally) there is no way to control continuity of $s \to v'_{w,s}$. The best example of such discontinuous continuation selection in macroeconomics application of strategic dynamic programming is, perhaps, the time-consistency model (see Caplin and Leahy 2006) discussed later in the application section. These technical issues are also important when developing a computational technique that uses specific properties of (the profile) the equilibrium decision rules with respect to s (important especially when the state space is uncountable).

4 Numerical Implementations

4.1 Set Approximation Techniques

Judd et al. (2003) propose a set approximation techniques to compute the greatest fixed point of operator B of the APS paper. In order to accomplish this task, they introduce public randomization that technically convexifies each iteration on the operator B, which allows them to select and coordinate on one of the future values that should be played. This enhances the computational procedure substantially.

More specifically, they propose to compute the inner V^I and outer V^O approximation of V^*, where $V^I \subset V^* \subset V^O$. Both approximations use a particular approximation of values of operator B, i.e., an inner approximation B^I and an outer approximation B^O that are both monotone. Further, for any set W, the approximate operators preserve the order under set inclusion, i.e., $B^I(W) \subset B(W) \subset B^O(W)$. Having such lower and upper approximating sets, Judd et al. (2003) are able to compute the error bounds (and a stopping criterion) using the Hausdorff distance on bounded sets in $\mathbb{R}^n$, i.e.:

$$d(W^O, W^I) = \max_{w_O \in W^O} \min_{w_I \in W^I} ||w_I - w_O||.$$

Their method is particularly useful as they work with convex sets at every iteration and map them on $\mathbb{R}^m$ by using its m extremal points. That is, if one takes m points, say $Z \subset W \subset \mathbb{R}^n$, define $W^I = \text{co}Z$. Next, for the outer approximation, take m points for Z on the boundary of a convex set W, and let $W^O = \bigcap_{l=1}^m \{z \in \mathbb{R}^n : g_l \cdot z \leq g_l \cdot z_l\}$ for a vector of m subgradients oriented such that $(z_l - w) \cdot g_l > 0$. To start iterating toward the inner approximation, one needs to find some equilibrium

[17] See Kydland and Prescott (1980), Phelan and Stacchetti (2001), Feng et al. (2014), and Feng (2015).

values from V^*, while to start iterating toward the outer, one needs to start from the largest possible set of values, say given by minimal and maximal bounds of the payoff vector.

Importantly, recent work by Abreu and Sannikov (2014) provides an interesting technique of limiting the number of extreme points of V^* for the finite number of action repeated games with perfect observability. In principle, this procedure could easily be incorporated in the methods of Judd et al. (2003). Further, an alternative procedure to approximate V^* was proposed by Chang (1998), who uses discretization instead of extremal points of the convex set. One final alternative is given by Cronshaw (1997), who proposes a Newton method for equilibrium value set approximation, where the mapping of sets W and $B(W)$ on $\mathbb{R}^m$ is done by computing the maximal weighted values of the players' payoffs (for given weights).[18]

Finally, and more recently, Berg and Kitti (2014) developed a method for computing the subgame perfect equilibrium value of a game with perfect monitoring using fractal geometry. Specifically, their method is interesting as it allows computation of the equilibrium value set with no public randomization, sunspot, or convexification. To obtain their result, they characterize the set V^* using (elementary) subpaths, i.e., (finite or infinite) paths of repeated action profiles, and compute them using the Hausdorff distance.[19]

4.2 Correspondence Approximation Techniques

The method proposed by Judd et al. (2003) was generalized to dynamic games (with endogenous and exogenous states) by Judd et al. (2015). As already mentioned, an appropriate version of the strategic dynamic programming method uses correspondences V^* defined on the state space to handle equilibrium values. Then authors propose methods to compute inner and outer (pointwise) approximations of V^*, where for given state s, $V^*(s)$ is approximated using original Judd et al. (2003) method. In order to convexify the values of V^*, the authors introduce sunspots.

Further, Sleet and Yeltekin (2016) consider a class of games with a finite number of exogenous states S and a compact set of endogenous states K. In their case, correspondence V^* maps on $S \times K$. Again the authors introduce sunspots to convexify the values of V^*; this, however, does not guarantee that $V^*(s, \cdot)$ is convex. Still the authors approximate correspondences by using step (convex-valued) correspondences applying constructions of Beer (1980).

Similar methods are used by Feng et al. (2014) to study sequential equilibria of dynamic economies. Here, the focus is often also on equilibrium policy functions (as

[18] Both of these early proposals suffer from some well-known issues, including curse of dimensionality or lack of convergence.

[19] See Rockafellar and Wets (2009), chaps. 4 and 5, for theory of approximating sets and correspondences.

opposed to value functions). In either case (of approximating values or policies), the authors concentrate on outer approximation only and discretize both the arguments and the spaces of values. Interestingly, Feng et al. (2014), based on Santos and Miao (2005), propose a numerical technique to simulate the moments of invariant distributions resulting from the set of sequential equilibria. In particular, after approximating the greatest fixed point of B, they convexify the image of $B(V)$ and approximate some invariant measure on $A \times S$ by selecting some policy functions from the approximated equilibrium value set V^*.

Finally, Balbus and Woźny (2016) propose a step correspondence approximation method to approximate function sets without the use of convexification for a class of short-memory equilibria. See also Kitti (2016) for a fractal geometry argument for computing (pointwise) equilibria in stochastic games without convexification.

5 Macroeconomic Applications of Strategic Dynamic Programming

In this section, we apply strategic dynamic programming methods to the canonical examples discussed in Sect. 2.

5.1 Hyperbolic Discounting

As already mentioned in Sect. 2.1, one important application of strategic dynamic programming methods that is particularly useful in macroeconomics is finding the time-consistent solutions to the quasi-hyperbolic discounting optimization problem.[20] We now present this application in more detail and provide sufficient conditions to construct all the consistent plans for this class of models.

Our environment is a version of a $\beta - \delta$ quasi-hyperbolic discounting model that has been studied extensively in the literature. We envision an agent to be a sequence of selves indexed in discrete time $t \in \mathbb{N} \cup \{0\}$. A "current self" or "self t" enters the period in given state $s_t \in S$, where for some $\bar{S} \in \mathbb{R}_+$, $S := [0, \bar{S}]$, and chooses an action denoted by $c_t \in [0, s_t]$. This choice determines a transition to the next period state s_{t+1} given by $s_{t+1} = f(s_t - c_t)$. The period utility function for the consumer is given by (bounded) utility function u that satisfies standard conditions. The discount factor from today (t) to tomorrow ($t + 1$) is $\beta\delta$; thereafter, it equals δ between any two future dates $t + \tau$ and $t + \tau + 1$ for $\tau > 0$. Thus, preferences (discount factor) depend on τ.

Let $h^t = (s_0, s_1, \ldots, s_{t-1}, s_t) \in H_t$ be the history of states realized up to period t, with $h^0 = \emptyset$. We can now define preferences and a subgame perfect equilibrium for the quasi-hyperbolic consumer.

[20] See, e.g., Harris and Laibson (2001) or Balbus et al. (2015d).

Definition 2. The sequence of functions $\sigma := (\sigma_t)_{t\in\mathbb{N}}$ is subgame perfect, if there is a sequence $(v_t)_{t\in\mathbb{N}}$, such that for each $t \in \mathbb{N}$ and $s \in S$

$$\sigma_t(h^t) \in \arg\max_{c\in[0,s_t]} \left\{(1-\delta)u(c) + \beta\delta v_{t+1}((h^t, f(s_t - c)))\right\},$$

and

$$v_t(h^t) = (1-\delta)u(\sigma_t(h^t)) + \delta v_{t+1}((h^t, f(s_t - \sigma_t(h^t)))).$$

Here, for uniformly bounded v_t, we have the following payoffs:

$$v_t(h^t) = \sum_{\tau=1}^{\infty} \delta^{\tau-1} u(\sigma_{t+\tau}(h^{t+\tau})). \tag{17.2}$$

Intuitively, current self best responds to the value v_{t+1} discounted by $\beta\delta$ and that continuation value v_{t+1} summarizes payoffs from future "selfs" strategies $(\sigma_\tau)_{\tau=t+1}^{\infty}$. Such a best response is then used to update v_{t+1} discounted by δ to v_t.

In order to construct a subset of SPNE, we proceed with the following construction. Put:

$$\Pi^\kappa(s,c)(v) := (1-\delta)u(c) + \kappa v(f(s-c))$$

for $\kappa \in [0,1]$. The operator B defined for a correspondence $W : S \rightrightarrows \mathbb{R}$ is given by:

$$B(W)(s) := \Big\{v \in \mathbb{R} :\ v = \Pi^\delta(s, a(s))(w),\ \text{for some } a, w$$

$$\text{s.t. } a \in \arg\max_{c\in[0,s]} \Pi^{\beta\delta}(s,c)(w),\ \text{and } w \in W(s)\Big\}.$$

Based on the operator B, one can prove the existence of a subgame perfect equilibrium in this intrapersonal game and also compute the equilibrium correspondence. We should note that this basic approach can be generalized to include nonstationary transitions $\{f_t\}$ or credit constraints (see, e.g., Bernheim et al. (2015), who compute the greatest fixed point of operator B for a specific example of CIES utility function). Also, Chade et al. (2008) pursue a similar approach for a version of this particular game, where at each period n, consumers play a strategic form game. In this case, the operator B must be adopted to require that a is not the optimal choice, but rather a Nash equilibrium of the stage game.

Finally, Balbus and Woźny (2016) show how to generalize this method to include a stochastic transition and concentrate on short-memory (Markov) equilibria (or Markov perfect Nash equilibria, MPNE henceforth). To illustrate the approach to short memory discussed in Sect. 3.3, we present an application of the strategic dynamic programming method for a class of strategies where each σ_t depends on s_t only, but in the context of a version of the game with stochastic transitions. Let $s_{t+1} \sim Q(\cdot|s_t - c_t)$, and CM be a set of nondecreasing, Lipschitz continuous (with modulus 1) functions $h : S \to S$, such that $\forall s \in S\ h(s) \in [0, s]$. Clearly, as CM is equicontinuous and closed, it is a nonempty, convex, and compact set when endowed with the topology of uniform convergence. Then the discounted sum in (17.2) is evaluated under E_s^σ that is an expectation relative to the unique probability measure (existence and uniqueness of such a measure follows from standard Ionescu-Tulcea theorem) on histories h^t determined by initial state $s_0 \in S$ and a strategy profile σ.

For given $\bar{S} \in \mathbb{R}_+$, $S = [0, \bar{S}]$, define a function space:

$$V := \{v : S \to \mathbb{R}_+ : v \text{ is nondecreasing and u.s.c. bounded by } u(0) \text{ and } u(\bar{S})\}.$$

And let:

$$V^* = \{v \in V : \exists \text{ MPNE } (\sigma_t)_{t \in \mathbb{N}},$$

where each $\sigma_t \in CM$, s.t $v(s) = U((\sigma_t)_{t\in\mathbb{N}})(s)\ \forall s \in S\}$.
In such a case, the stage payoff is

$$\Pi^\kappa(s, c)(v) := (1 - \delta)u(c) + \kappa \int_S v(s')Q(ds'|s - c),$$

and operator B^f defined on 2^V is given by:

$$B^f(W) := \bigcup_{w \in W} \Big\{ v \in V : (\forall s \in S)\ v(s) = \Pi^\delta(s, a(s))(w), \text{ for some } a : S \to S,$$

$$\text{s.t. } a(s) \in \arg\max_{c \in [0,s]} \Pi^{\beta\delta}(s, c)(w) \text{ for all } s \in S \Big\}.$$

Balbus and Woźny (2016) prove that the greatest fixed point of B^f characterizes the set of all MPNE values in V generated with short memory, and they also discuss how to compute the set of all such equilibrium values.

5.2 Optimal Growth Without Commitment

Similar methods for (nonstationary) Markov perfect equilibrium can be developed for a class of optimal growth models without commitment between consecutive generations. As discussed in Sect. 2.2, this is often formalized using a class of paternalistic bequest games. We now present a detailed application of this class of models.

For the sake of illustration, consider a simple model of stochastic growth without commitment where there is an infinite sequence of generations labeled by $t \in \mathbb{N}$. In the economy, there is one commodity which may be either consumed or invested. Every generation lives one period and derives utility u from its own consumption and utility v from consumption of its immediate descendant. Generation t receives the endowment $s_t \in S$ and chooses consumption level $c_t \in A(s_t) := [0, s_t]$. The investment of $y_t := s_t - c_t$ determines the endowment of its successor according to some stochastic transition probability Q_t from S to S which depends on y_t.

Let P be the set of (bounded by a common bound) Borel measurable functions $p : S \mapsto \mathbb{R}_+$. A strategy for generation t is a function $\sigma_t \in \Sigma$, where Σ is a set of Borel measurable functions such that $\sigma(s) \in A(s)$ for each $s \in S$. The expected utility of generation t is defined as follows:

$$u(c) + \int_S v(\sigma_{t+1}(s'))Q(ds'|s - c, s), \tag{17.3}$$

where $u : S \mapsto \mathbb{R}_+$ is a bounded function, whereas $v : S \mapsto \mathbb{R}_+$ is bounded and Borel measurable. We endow P with its weak star topology and order 2^P (the set of all subsets of P) by set inclusion order.

Then, in sect. 5 of their paper, Balbus et al. (2012) define an operator B^f on 2^P:

$$B^f(W) = \bigcup_{p \in W} \Big\{ p' \in \mathcal{P} : p'(s) = v(a_p^*(s)), \text{ where}$$

$$a_p^*(s) \in \arg\max_{c \in A(s)} \{u(c) + \int_S p(s')Q(ds'|s - c)\} \Big\}.$$

Clearly, each selection of values $\{v_t^*\}$ from the greatest fixed point $V^* = B^f(V^*)$ generates a MPNE strategy $\{\sigma_t^*\}$, where $\sigma_t^*(s) \in \arg\max\{u(c) + \int_S v_t^*(s')Q(ds'|s - c)\}$. Hence, using operator B^f, not only is the existence of MPNE established, but also a direct computational procedure can be used to compute the entire set of sustainable MPNE values. Here we note that a similar technique was used by Bernheim and Ray (1983) to study MPNE of a nonstationary bequest game.

Another direct application of the strategic dynamic programming presented above was proposed by Atkeson (1991) to study the problem of international lending with moral hazard and risk of repudiation. Specifically, using the currently available income (net of repayment) as a state variable and correspondences of possible continuation utilities, he characterizes the set of Pareto optimal allocations constrained to satisfy individual rationality and incentive compatibility (including no repudiation constraint), using the techniques advocated in Sect. 3.1.

5.3 Optimal Fiscal Policy Without Commitment

As already introduced in Sect. 2.3, in their seminal paper, Kydland and Prescott (1980) have proposed a recursive method to solve for the optimal tax policy of the dynamic economy. Their approach resembles APS method for dynamic games, but is different as it incorporates dual variables as states of the dynamic program. Such a state variable can be constructed because of the dynamic Stackelberg structure of the game. That is, equilibrium in the private economy is constructed first, and these agents are "small" players in the game, and take as given sequences of government tax policies, and simply optimize. As these problems are convex, standard Euler equations govern the dynamic equilibrium in this economy. Then, in the approach of Kydland-Prescott, in the second stage of the game, successive generations of governments design time-consistent policies by forcing successive generations of governments to condition optimal choices on the lagged values of Lagrange/KKT multipliers.

To illustrate how this approach works, consider an infinite horizon economy with a representative consumer solving:

$$\max_{\{a_t\}} \sum_{t=0}^{\infty} \delta^t u(c_t, n_t, g_t),$$

where $a_t = (c_t, n_t, k_{t+1})$ is choice of consumption, labor, and next period capital, subject to the budget constraints $k_{t+1} + c_t \leq k_t + (1 - \theta_t)r_t k_t + (1 - \tau_t)w_t n_t$ and feasibility constraint $a_t \geq 0$, $n_t \leq 1$. Here, θ_t and τ_t are the government tax rates and g_t their spendings. Formulating Lagrangian and writing the first-order conditions, together with standard firm's profit maximization conditions, one obtains $u_c(c_t, n_t, g_t) = \lambda_t$, $u_n(c_t, n_t, g_t) = -\lambda_t(1 - \tau_t)w_t$ and $\delta[1 + (1 - \theta_{t+1}) f_k(k_{t+1}, n_{t+1})]\lambda_{t+1} = \lambda_t$.[21]

Next the government solves:

$$\max_{\{\pi_t\}} \sum_{t=0}^{\infty} \delta^t u(c_t, n_t, g_t),$$

where $\pi_t = (g_t, \tau_t, \theta_t)$ under the above-given first-order conditions of the consumer and budget balance constraint: $g_t \leq \theta_t f_k(k_t, n_t)k_t + \tau_t f_n(k_t, n_t)n_t$. It is well known that the solution to this problem (on the natural state space k_t) is time inconsistent. That is, the solution of the problem, e.g., π_{t+s} chosen at time t, is different from the solution of the same problem at time $t + s$. Hence, standard dynamic programming techniques cannot be applied.

[21] Phelan and Stacchetti (2001) prove that a sequential equilibrium exists in this economy for each feasible sequence of tax rates and expenditures.

Kydland and Prescott (1980) then propose, however, a new method to make the problem recursive by adding a pseudo-state variable λ_{t-1}. Relative to this new state space, one can then develop a recursive optimization approach to the time-consistency problem that resembles the strategic dynamic programming methods of APS. To see this, omitting the time subscripts, Kydland and Prescott (1980) rewrite the problem of the government recursively by

$$v(k, \lambda_{-1}) = \max_{a,\pi,\lambda}\{u(c, n, g) + \delta v(k', \lambda)\}$$

under the budget balance, the first-order conditions of the consumer, and requiring that $(k'^*, \lambda^*) \in V^*$. Here V^* is the set of such $\{k_t, \lambda_t\}$, for which there exists an equilibrium policy $\{a_s, \pi_s, \lambda_s\}_{s=t}^{\infty}$ consistent with or supportable by these choices. This formalizes the constraint needed to impose time-consistent solutions on government choices. To characterize set V^*, Kydland and Prescott (1980) use the following APS type operator:

$$\begin{aligned} B(W) = \{&(k, \lambda_{-1}) \in [0, k_{\max}] \times [\lambda_{\min}, \lambda_{\max}] : \text{ there exists} \\ &(a, \pi, \lambda)\text{satisfying budget balance constraints and consumer FOCs, with} \\ &(k', \lambda) \in W\}. \end{aligned}$$

They show that V^* is the largest fixed point of B and this way characterize the set of all optimal equilibrium policies. Such are time consistent on the expanded state space (k, λ_{-1}), but not on the natural state space k.

This approach was later extended and formalized by Phelan and Stacchetti (2001). They study the Ramsey optimal taxation problem in the symmetric sequential equilibrium of the underlying economy. They consider a dynamic game and also use Lagrange multipliers to characterize the continuation values. Moreover, instead of focusing on optimal Ramsey policies, they study symmetric sequential equilibrium of the economy and hence incorporate some private state variables. Specifically, in the direct extension of the strategic dynamic programming technique with private states, one should consider a distribution of private states (say capital) and (for each state) a possible continuation value function. But as all the households are ex ante identical, and sharing the same belief about the future continuations, they have the same functions characterizing the first-order conditions, although evaluated at different points, in fact only at the values of the Lagrange multipliers that keep track of the sequential equilibrium dynamics. In order to characterize the equilibrium conditions by FOCs, Phelan and Stacchetti (2001) add a public sunspot $s \in [0, 1]$ that allows to convexify the equilibrium set under study. The operator B in their paper is then defined as follows:

$$\begin{aligned} B(W)(k) = \text{co}\{&(\lambda, v) : \text{ there exists } (a, \lambda', v') \text{ satisfying consumer equilibrium} \\ &\text{FOCs, with } (\lambda', v') \in W(k) \text{ and the government deviation} \\ &\text{is punished by the minimax (worst) equilibrium value}\}. \end{aligned}$$

Here, λ is the after tax marginal utility of consumption and v is household equilibrium value. Notice that, as opposed to the methods in Kydland and Prescott (1980), these authors integrate the household and government problem into one operator equation. Phelan and Stacchetti (2001) finish with the characterization of the best steady states of the symmetric sequential equilibrium.

Finally, this approach was more recently extended by Feng et al. (2014) (and Santos and Miao (2005) earlier), as a generalization of the strategic dynamic programming method to characterize all sequential equilibria of the more general dynamic stochastic general equilibrium economy. They follow the Phelan and Stacchetti (2001) approach and map the sequential equilibrium values to the space of continuation Lagrange multipliers values. Specifically, they consider a general economy with many agents in discrete time, with endogenous choices $a \in A$ and countably many exogenous shocks $s \in S$, drawn each period from distribution $Q(\cdot|s)$. Denoting the vector of endogenous variables by y, they assume the model dynamics is given by a condition $\phi(a', a, y, s) = 0$ specifying the budget and technological constraints. Next, denoting by $\lambda \in \Lambda$ the marginal values of all the investments of all the agents, they consider a function $\lambda = h(a, y, s)$. Finally, the necessary and sufficient first-order conditions for the household problems are given by

$$\Phi(a, y, s, \sum_{s' \in S} \lambda'(s') Q(ds'|s)) = 0,$$

where λ' is the next period continuation marginal value as a function on S. Next, they characterize the correspondence V^* mapping $A \times S$ to Λ, as the greatest fixed point of the correspondence-based operator:

$$\begin{aligned} B(W)(a, s) := \{\lambda : \lambda = h(a, y, s) \text{ for some } y, a', \lambda' \text{ with} \\ \Phi(a, y, s, \sum_{s' \in S} \lambda'(s') Q(ds'|s)) = 0,\ \phi(a', a, y, s) = 0 \\ \text{and } \lambda'(s') \in W(a', s')\}. \end{aligned}$$

To characterize V^*, they operate on the set of all upper hemi-continuous correspondences and under standard continuity conditions show that B maps W with compact graph into correspondence $B(W)$ with compact graph. Using the intersection theorem, along with a standard measurable selection theorem, they select a policy function a' as function of (a, s, λ) (hence, Markovian on the expanded state space including Lagrange multipliers λ).[22]

[22]It bears mentioning that in Phelan and Stacchetti (2001) and Feng et al. (2014) the authors actually used envelope theorems essentially as the new state variables. But, of course, assuming a dual representation of the sequential primal problem, this will then be summarized essentially by the KKT/Lagrange multipliers.

In order to compute the equilibrium correspondence, they use Beer (1980) algorithm of approximating correspondences by step correspondences on the discretized domain and co-domain grids as already discussed. Feng et al. (2014) conclude their paper with applications of the above framework to nonoptimal growth models with taxes, monetary economies, or asset prices with incomplete markets. See also Dominguez and Feng (2016b) and Feng (2015) for a recent application of the Feng et al. (2014) strategic dynamic programing method to a large class of optimal Ramsey taxation problems with and without constitutional constraints. In these papers, the authors are able to quantify the value of commitment technologies in optimal taxation problems (e.g., constitutional constraints) as opposed to imposing simply time-consistent solutions.

Finally, it is worth mentioning that the Feng et al. (2014) method is also useful to a class of OLG economies, hence with short-lived agents. See also Sleet (1998) (chap. 3) for such a model.

5.4 Optimal Monetary Policy Without Commitment

We should briefly mention that an extension of the Kydland and Prescott (1980) approach in the study of policy games was proposed by Sleet (2001). He analyzes a game between the private economy and the government or central bank possessing some private information. Instead of analyzing optimal tax policies like Kydland and Prescott (1980), he concentrates on optimal, credible, and incentive-compatible monetary policies. Technically, similar to Kydland and Prescott (1980), he introduces Lagrange multipliers that, apart from payoffs as state variables, allow to characterize the equilibrium set. He then applies the computational techniques of Judd et al. (2003) to compute dynamic equilibrium and then recovers the equilibrium allocation and prices. This extension of the methods makes it possible to incorporate private signals, as was later developed by Sleet and Yeltekin (2007).

We should also mention applications of strategic dynamic programming without states to optimal sustainable monetary policy due to Chang (1998) or Athey et al. (2005) in their study of optimal discretion of the monetary policy in a more specific model of monetary policy.

6 Alternative Techniques

We conclude with a few remarks concerning alternative methods to strategic dynamic programming for constructing dynamic equilibria in the macroeconomic models with strategically interacting agents. In particular, we focus on two widely used approaches that have been proposed in the literature, each providing significant advantages relative to strategic dynamic programming per characterizing *some* dynamic equilibrium (when applicable).

6.1 Incentive-Constrained Dynamic Programming

The first alternative approach to strategic dynamic programming methods are incentive-constrained dynamic programming methods (and their associated dual methods, often referred to in the literature as "recursive saddle point" or "recursive dual" methods). These methods develop recursive representations of sequential incentive-constrained optimization problems that are used to represent dynamic equilibria in macroeconomic models that are also dynamic/stochastic games. The methods are in the spirit of the recursive optimization approaches we discussed in Sect. 2 (e.g., the recursive optimization approaches to models with dynamically inconsistent payoffs or limited commitment such as models that are studied as interpersonal games between successive generations as in models with quasi-hyperbolic agents or growth models with limited commitment). The seminal early work on these methods is found in Rustichini (1998a,b) and Marcet and Marimon (1998), but a wealth of recent work has extended many of their ideas. These methods are used in models where agents face sequential optimization problems, but have incentives to change future optimal continuation plans when future states actually arise. Therefore, incentive-constrained programming methods add further constraints on sequential optimal decisions that agents face in the form of period-by-period dynamic incentive and participation constraints. These constraints are imposed to guarantee optimal decisions are time consistent (or, in some cases, subgame perfect) along equilibrium paths and therefore further restrict sequential optimal choices of economic agents. Then, incentive-constrained dynamic programming methods seek to find recursive primal or dual representations of these sequential incentive-constrained optimization problems. Such recursive representations help sharpen the characterization of dynamic equilibria strategies/policies.

Many applications of these incentive-constrained programming methods have arisen in the macroeconomic literature. For example, in dynamic asset pricing models with limited commitment and strategic default, where incentive constraints are used to model endogenous borrowing constraints that restrict current asset-consumption choices to be consistent with households not defaulting on outstanding debt obligations in any state the continuation periods (e.g., see Alvarez and Jermann 2000; Hellwig and Lorenzoni 2009; Kehoe and Levine 1993, 2001). Such solvency constraints force households to make current decisions that are consistent with them being able to credibly commit to future repayment schemes, and not to default on their debt obligations, making repayment schemes self-enforcing and sustainable. Similar recursive optimization approaches to imposing dynamic incentives for credible commitment to future actions arise in models of sustainable plans for the government in models in dynamic optimal taxation. Such problems have been studied extensively in the literature, including models of complete information and incomplete information (e.g., see the work of Chari et al. 1991; Chari and Kehoe 1990; Farhi et al. 2012; Sleet and Yeltekin 2006a,b).

Unfortunately, the technical limitations of these methods are substantial. In particular, the presence of dynamic incentive constraints greatly complicates the analysis of the resulting incentive-constrained sequential optimization problem (and hence recursive representation of this sequential incentive-constrained problem). For example, even in models where the primitive data under perfect commitment imply the sequential optimization problem generates value functions that are concave over initial states in models with limited commitment and state variables (e.g., capital stocks, asset holdings, etc.), the constraint set in such problems is no longer convex-valued in states. Therefore, as value function in the sequential problem ends up generally not being concave, it is not in general differentiable, and so developing useful recursive primal or recursive dual representations of the optimal incentive-constrained solutions (e.g., Euler inequalities) is challenging (e.g., see Rustichini (1998a) and Messner et al. (2014) for a discussion). That is, given this fact, an immediate complication for characterizing incentive-constrained solutions is that value functions associated with recursive reformulations of these problems are generally not differentiable (e.g., see Rincón-Zapatero and Santos (2009) and Morand et al. (2015) for discussion). This implies that standard (smooth) Euler inequalities, which are always useful for characterizing optimal incentive-constrained solutions, fail to exist. Further, as the recursive primal/dual is not concave, even if necessary first-order conditions can be constructed, they are not sufficient. These facts, together, greatly complicate the development of rigorous recursive primal methods for construction and characterization of optimal incentive-constrained sequential solutions (even if conditions for the existence of a value function in the recursive primal/dual exist). This also implies that even when such sequential problems can be recursively formulated, they cannot be conjugated with *saddle points* using any known recursive dual approach. See Messner et al. (2012, 2014) for a discussion.[23]

Now, when trying to construct recursive representations of the sequential incentive-constrained primal problem, new problems emerge. For example, these problems cannot be solved generally by standard dynamic programming type arguments (e.g., standard methods for solving Bellman equations in dynamic programming). In particular, the resulting operator equation that must be solved is not, in general, a contraction (e.g., see Rustichini 1998a). Rustichini (1998a) shows that although the standard dynamic programming tools do not apply to sequential optimization problems with dynamic incentive constraints, one can develop a monotone iterative method based on a nonlinear operator (that has the spirit of a Bellman operator) that computes recursive solutions to the sequential incentive-constrained optimization problems. When sufficient conditions for a fixed point for the resulting functional equation in the recursive primal problem can be given, the recursive primal approach provides an alternative to APS/strategic

[23]It is worth mentioning that Messner et al. (2012, 2014) often do not have sufficient conditions on primitives to guarantee that dynamic games studied using their recursive dual approaches have recursive saddle point solutions for models with state variables. Most interesting applications of game theory in macroeconomics involve states variable (i.e., they are dynamic or stochastic games).

dynamic programming methods as some dynamic equilibrium value in the model can be computed *if* the Markovian policies can be computed and characterized. Unfortunately, without a dual method for its implementation, computing the set of incentive-constrained optimal solutions that achieve this value (i.e., in these models, dynamic equilibria) is in general very difficult.

One other technical limitation of this method relative to strategic dynamic programming is often the punishment scheme used to sustain dynamic equilibria in general is ad hoc. That is, in strategic dynamic programming/APS methods, the punishment schemes used to construct sequential/subgame values are *endogenous*; in the standard version of an incentive-constrained dynamic programming problem, the punishment schemes are exogenous.

The primal formulation of incentive-constrained dynamic programming has been applied to many important macroeconomic models. In his original paper, Rustichini (1998a) shows how by adding period-by-period incentive constraints to the relevant decision-makers' problems in some important macroeconomic models with limited commitment incentive-constrained dynamic programming can be used to prove the existence of time-consistent or sustainable optimal policies. If the question is the existence of dynamic equilibria in such models, the primal versions of these recursive primal methods are very powerful. The problem with these methods is that it is challenging to compute the incentive-constrained optimal solutions themselves. Two interesting applications he makes in his paper are to optimal Ramsey taxation problems under limited commitment and models of economic growth without commitment. Since the publication of his paper, other applications of these methods have arisen. For example, they have been applied to studying optimal solutions to household's problem in dynamic asset accumulation models with limited commitment, a government optimal taxation problem with time inconsistent preferences, sustaining sovereign debt in models of international finance, and contract enforcement problems in models with human capital. See, for example, Koeppl (2007), Durdu et al. (2013), and Krebs et al. (2015), among many others, for a discussion.

To address the question of the computation of incentive-constrained optimal solutions, an extensive new literature has arisen. This recent work was motivated by the original paper of Kydland and Prescott (1980), as well as Marcet and Marimon (1998), where "dual variables" were used as "pseudo-state" variables to construct time-consistent optimal solutions. Indeed, in these two papers, the authors show how to apply recursive dual approaches to a plethora of dynamic macroeconomic and dynamic contracting problems. In the original paper by Kydland and Prescott (1980), a recursive method for constructing generalized Markov equilibria was proposed where by adding the lagged values of Karush-Kuhn-Tucker multipliers to the set of state variables to optimal taxation rules, which forced the resulting government policymaker's taxation policies to respect a "backward-looking" constraint, would in turn force the resulting optimal solution to the time-inconsistent Euler equations (under the additional implied constraints) to be time consistent. See also Feng et al. (2014) for a significant extension of this method.

In the important paper by Marcet and Marimon (1998), the authors extend the ideas of Kydland and Prescott (1980) to the setting of a recursive dual optimization method, where the restrictions implied in the Kydland-Prescott method were explored more systematically. In the approach of Marcet and Marimon (1998), this restriction was embedded more formally into an extended dual recursive optimization approach, where KKT multipliers are added as state variables, and where in principle sufficient conditions can be developed such that this dual method will deliver incentive-constrained solutions to the primal recursive optimization methods ala Rustichini (1998a). The success of this dual approach critically relies on the existence of a recursive representation of saddle points, and in dynamic models where their dual recursive saddle point methods remain strictly concave, it can be proven that the methods of Marcet and Marimon (1998) compute primal incentive-constrained optimal solutions. The problem with this method is that in very simple concave problems, serious issues with duality can arise (e.g., see Cole and Kubler (2012) and Messner and Pavoni (2016) for discussion). The first problem is that in simple dynamic contracting problems, dual solutions can fail to be primal feasible (e.g., see the example in Messner and Pavoni 2016). In some cases, this issue can be resolved by extending the recursive saddle point method to weakly concave settings by introducing lotteries into the framework. In particular, see Cole and Kubler (2012). So even when recursive saddle points exist, some technical issues with the method can arise. Very importantly, Rustichini (1998b) shows even in concave settings, the dual variables/KKT multipliers can be poorly behaved from a duality perspective (see also Le Van and Saglam (2004) and Rincón-Zapatero and Santos (2009) for details).

In a series of recent papers by Messner et al. (2012, 2014), the authors further develop this recursive dual method. In these papers, they develop sufficient conditions for the equivalence of sequential primal and recursive dual formulations. For example, similar technical issues arise for recursive dual methods per existence of value functions that satisfy the functional equation that must be solved to represent the dual sequential incentive-constrained programming problem with a recursive dual (e.g., see Messner et al. 2014). Relative to the question of the existence of a recursive dual version of the recursive primal problem, Messner et al. (2014) provide the most general conditions under which a recursive dual formulation exists for a large class of dynamic models with incentive constraints.[24] In this paper, also many important questions concerning the equivalence of recursive primal and recursive dual solutions are addressed, as well as the question of sequential and recursive dual equivalence. In Messner et al. (2012), for example, the authors provide equivalence in models without backward-looking constraints (e.g., constraints generated by state variables such as capital in time-consistent optimal taxation problems á la Kydland and Prescott 1980) and many models with linear forward-looking

[24]For example, it is not a contraction in a the "sup" or "weighted sup" metric. It is a contraction (or a local contraction) under some reasonable conditions in the Thompson metric. See Messner et al. (2014) for details.

incentive constraints (e.g., models with incentive constraints, models with limited commitment, etc.). In Messner et al. (2014), they give new sufficient conditions to extend these results to settings with backward-looking states/constraints, as well as models with general forward-looking constraints (including models with nonseparabilities across states). In this second paper, they also give new conditions for the existence of recursive dual value functions using contraction mapping arguments (in the Thompson metric). This series of papers represents a significant advancement of the recursive dual approach; yet, many of the results in these papers still critically hinge upon the existence of recursive saddle point solutions, and conditions on primitives of the model are not provided for these critical hypotheses. But critically, in this recursive dual reformulations, the properties of Lagrangians can be problematic (e.g., see Rustichini 1998b).

6.2 Generalized Euler Equation Methods

A second class of methods that have found use to construct Markov equilibrium in macroeconomic models that are dynamic games are generalized Euler equation methods. These methods were pioneered in the important papers by Harris and Laibson (2001) and Krusell et al. (2002), but have subsequently been used in a number of other recent papers. In these methods, one develops a so-called generalized Euler equation that is derived from the local first- and second-order properties relative to the theory of derivatives of local functions of bounded variation of an equilibrium value function (or value functions) that govern a recursive representation of agents' sequential optimization problem. Then, from these local representations of the value function, one can construct a generalized first-order representation of any Markovian equilibrium (i.e., a generalized Euler equation, which is a natural extension of a standard Euler) using this more general language of nonsmooth analysis. From this recursive representation of the agents' sequential optimization problem, plus this related generalized Euler equation, one can then construct an approximate solution to the actual pair of functional equations that are used to characterize a Markov perfect equilibrium, and Markov perfect equilibrium values and pure strategies can then be computed. The original method based on the theory of local functions of bounded variation was proposed in Harris and Laibson (2001) , and this method remains the most general, but some authors have assumed the Markovian equilibrium being computed is continuously differentiable, which greatly sharpens the generalized Euler equation method.[25]

[25]By "most general", we mean has the weakest assumptions on the *assumed* structure of Markov perfect stationary equilibria. That is, in other implementations of the generalized Euler equation method, authors often assume *smooth* Markov perfect stationary equilibria exist. In none of these cases do the authors actually appear to prove the existence of Markov perfect stationary equilibria within the class postulated.

These methods have applied in a large number of papers in the literature. Relative to the models discussed in Sect. 2, Harris and Laibson (2001), Krusell et al. (2002), and Maliar and Maliar (2005) (among many others) have used generalized Euler equation methods to solve dynamic general equilibrium models with a representative quasi-hyperbolic consumers. In Maliar and Maliar (2006b), a version of the method is applied to dynamic economies with heterogeneous agents, each of which has quasi-hyperbolic preferences. In Maliar and Maliar (2006a, 2016), some important issues with the implementation of generalized Euler equations are discussed (in particular, they show that there is a continuum of smooth solutions that arise using these methods for models with quasi-hyperbolic consumers. In Maliar and Maliar (2016), the authors propose an interesting resolution to this problem by using the turnpike properties of the dynamic models to pin down the set of dynamic equilibria being computed.

They have also been applied in the optimal taxation literature.[26] For example, in Klein et al. (2008), the authors study a similar problem to the optimal time-consistent taxation problem of Kydland and Prescott (1980) and Phelan and Stacchetti (2001). In their paper, they assume that a *differentiable* Markov perfect equilibrium exists, and then proceed to characterize and compute Markov perfect stationary equilibria using a generalized Euler equation method in the spirit of Harris and Laibson (2001). Assuming that such smooth Markov perfect equilibria exist, their characterization of dynamic equilibria is much sharper than those obtained using the calculus of functions of bounded variation. The methods also provide a much sharper characterization of dynamic equilibrium than obtained using strategic dynamic programming. In particular, Markov equilibrium strategies can be computed and characterized directly. They find that only taxation method available to the Markovian government is capital income taxation. This appears in contrast to the findings about optimal time-consistent policies using strategic dynamic programming methods in Phelan and Stacchetti (2001), as well as the findings in Klein and Ríos-Rull (2003). In Klein et al. (2005), the results are extended to two country models with endogenous labor supply and capital mobility.

There are numerous problems with this approach as it has been applied in the current literature. First and foremost, relative to the work assuming that the Markov perfect equilibrium is smooth, this assumption seems exceedingly strong as in very few models of dynamic games in the literature, when Markov equilibria are known to exist, they are smooth. That is, conditions on the primitives of these games that guarantee such smooth Markov perfect equilibria exist are never verified. Further, even relative to applications of these methods using local results for functions of bounded variation, the problem is, although the method can solve the resulting generalized Euler equation, that solution cannot be tied to any particular value function in the actual game that generates this solution as satisfying the *sufficient* condition for a best reply map in the actual game. So it is not clear how to relate the solutions using these methods to the actual solutions in the dynamic or stochastic game.

---

[26] See for example Klein and Ríos-Rull (2003) and Klein et al. (2008).

In Balbus et al. (2015d), the authors develop sufficient conditions on the underlying stochastic game that a generalized Bellman approach can be applied to construct Markov perfect stationary equilibria. In their approach, the Markov perfect equilibria computing in the stochastic game can be directly related to the recursive optimization approach first advocated in, for example, Strotz (1955) (and later, Caplin and Leahy 2006). In Balbus et al. (2016), sufficient conditions for the uniqueness of Markov perfect equilibria are given. In principle, one could study if these equilibria are smooth (and hence, rigorously apply the generalized Euler equation method). Further, of course, in some versions of the quasi-hyperbolic discounting problem, closed-form solutions are available. But even in such cases, as Maliar and Maliar (2016) note, numerical solutions using some type of generalized Euler equation method need not converge to the actual closed-form solution.

7 Conclusion

Strategic interactions play a critical role in many dynamic models in macroeconomics. The introduction of such strategic elements into dynamic general equilibrium models has expanded greatly since the seminal work of Kydland and Prescott (1977), as well as early papers by Phelps and Pollak (1968), Peleg and Yaari (1973), Bernheim and Ray (1983), and Levhari and Mirman (1980). It is now a common feature of many models in macro, including models of economic fluctuations, public policy, asset pricing, models of the behavioral aspects of consumption-savings problems, models of economic growth with limited commitment or strategic altruism, among others. In this chapter, we have presented a number of canonical situations, where strategic considerations arise in the study of dynamic equilibria in macroeconomics. Then, we have discussed how the tools of dynamic and stochastic game theory can be used to study equilibria in such problems.

The introduction of such strategic dimensions into macroeconomics greatly complicates the analysis of equilibria. Still, rigorous and general methods are available for constructing, characterizing, and computing them. We have argued that strategic dynamic programming methods, first pioneered in Abreu et al. (1986, 1990) for repeated games, when extended to settings with state variables, provide a powerful systematic set of tools to construct and compute equilibria in such macroeconomic models. Also, we have mentioned that in some cases, for particular subclasses of sequential or subgame perfect equilibria (e.g., Markov perfect equilibria), these methods can be improved upon using recursive primal/dual methods or generalized Euler equation methods. Unfortunately, relative to strategic dynamic programming methods, these methods are known to suffer from serious technical limitations in some dynamic models with state variables. As the majority of the models studied in macroeconomics are dynamic, and include states, strategic dynamic programming offers the most systematic approach to such models; hence, in this chapter, we have discussed what these methods are and how they can be applied to a number of interesting models in dynamic macroeconomics.

References

Abreu D (1988) On the theory of infinitely repeated games with discounting. Econometrica 56(2):383–396

Abreu D, Pearce D, Stacchetti E (1986) Optimal cartel equilibria with imperfect monitoring. J Econ Theory 39(1):251–269

Abreu D, Pearce D, Stacchetti E (1990) Toward a theory of discounted repeated games with imperfect monitoring. Econometrica 58(5):1041–1063

Abreu D, Sannikov Y (2014) An algorithm for two-player repeated games with perfect monitoring. Theor Econ 9(2):313–338

Aliprantis CD, Border KC (2005) Infinite dimensional analysis. A Hitchhiker's guide. Springer, Heidelberg

Alvarez F (1999) Social mobility: the Barro-Becker children meet the Laitner-Loury dynasties. Rev Econ Stud 2(1):65–103

Alvarez F, Jermann UJ (2000) Efficiency, equilibrium, and asset pricing with risk of default. Econometrica 68(4):775–798

Amir R (1996a) Continuous stochastic games of capital accumulation with convex transitions. Games Econ Behav 15(2):111–131

Amir R (1996b) Strategic intergenerational bequests with stochastic convex production. Econ Theory 8(2):367–376

Arellano C (2008) Default risk and income fluctuations in emerging economies. Am Econ Rev 98(3):690–712

Athey S, Atkeson A, Kehoe PJ (2005) The optimal degree of discretion in monetary policy. Econometrica 73(5):1431–1475

Atkeson A (1991) International lending with moral hazard and risk of repudiation. Econometrica 59(4):1069–1089

Aumann R (1965) Integrals of set-valued functions. J Math Anal Appl 12(1):1–12

Balbus L, Jaśkiewicz A, Nowak AS (2015a) Bequest games with unbounded utility functions. J Math Anal Appl 427(1):515–524

Balbus Ł, Jaśkiewicz A, Nowak AS (2015b) Existence of stationary Markov perfect equilibria in stochastic altruistic growth economies. J Optim Theory Appl 165(1):295–315

Balbus Ł, Jaśkiewicz A, Nowak AS (2015c) Stochastic bequest games. Games Econ Behav 90(C):247–256

Balbus Ł, Nowak AS (2004) Construction of Nash equilibria in symmetric stochastic games of capital accumulation. Math Meth Oper Res 60:267–277

Balbus Ł, Nowak AS (2008) Existence of perfect equilibria in a class of multigenerational stochastic games of capital accumulation. Automatica 44(6):1471–1479

Balbus Ł, Reffett K, Woźny Ł (2012) Stationary Markovian equilibrium in altruistic stochastic OLG models with limited commitment. J Math Econ 48(2):115–132

Balbus Ł, Reffett K, Woźny Ł (2013) A constructive geometrical approach to the uniqueness of Markov perfect equilibrium in stochastic games of intergenerational altruism. J Econ Dyn Control 37(5):1019–1039

Balbus Ł, Reffett K, Woźny Ł (2014) A constructive study of Markov equilibria in stochastic games with strategic complementarities. J Econ Theory 150:815–840

Balbus Ł, Reffett K, Woźny Ł (2015d) Time consistent Markov policies in dynamic economies with quasi-hyperbolic consumers. Int J Game Theory 44(1):83–112

Balbus Ł, Reffett K, Woźny Ł (2016) On uniqueness of time-consistent Markov policies for quasi-hyperbolic consumers under uncertainty. Working paper no. WP KAE 2016_020, Collegium of Economic Analysis Working Paper Series, Warsaw School of Economics

Balbus Ł, Woźny Ł (2016) A strategic dynamic programming method for studying short-memory equilibria of stochastic games with uncountable number of states. Dyn Games Appl 6(2): 187–208

Baldauf M, Judd K, Mollner J, Yeltekin S (2015) Supergames with states. Discussion paper, MS

Banerjee A, Mullainathan S (2010) The shape of temptation: implications for the economics lives of the poor. Working paper No. 15973, NBER
Barro RJ (1999) Ramsey meets Laibson in the neoclassical growth model. Q J Econ 114(4): 1125–1152
Barro RJ, Gordon DB (1983) Rules, discretion and reputation in a model of monetary policy. J Monet Econ 12(1):101–121
Beer G (1980) The approximation of upper semicontinuous multifunction by step multifunction. Pac J Math 87(1):11–19
Benjamin D, Wright MLJ (2009) Recovery before redemption: a theory of delays in sovereign debt renegotiations. Discussion paper 2009-2015, CAMA working papers
Berg K, Kitti M (2014) Fractal geometry of equilibrium payoffs in discounted supergames. Fractals 22(4)
Bernheim D, Ray D (1983) Altruistic growth economies. I. Existence of bequest equilibria. Technical report no. 419, Institute for Mathematical Studies in the Social Sciences
Bernheim D, Ray D (1987) Economic growth with intergenerational altruism. Rev Econ Stud 54:227–242
Bernheim BD, Ray D, Yeltekin S (2015) Poverty and self-control. Econometrica 85(5):1877–1911
Brock W, Mirman L (1972) Optimal economic growth and uncertainty: the discounted case. J Econ Theory 4:479–513
Broner F, Erce A, Martin A, Ventura J (2014) Sovereign debt markets in turbulent times: creditor discrimination and crowding-out effects. J Monet Econ 61(C):114–142
Broner F, Martin A, Ventura J (2010) Sovereign risk and secondary markets. Am Econ Rev 100(4):1523–1555
Caplin A, Leahy J (2006) The recursive approach to time inconsistency. J Econ Theory 131(1): 134–156
Cass D (1965) Optimum growth in an aggregative model of capital accumulation. Rev Econ Stud 32:233–240
Chade H, Prokopovych P, Smith L (2008) Repeated games with present-biased preferences. J Econ Theory 139(1):157–175
Chamley C (1986) Optimal taxation of capital income in general equilibrium with infinite lives. Econometrica 54(3):607–622
Chang R (1998) Credible monetary policy in an infinite horizon model: recursive approaches. J Econ Theory 81(2):431–461
Chari V, Christiano L, Kehoe P (1991) Optimal fiscal and monetary policy: some recent results. J Money Credit Bank 23:519–539
Chari V, Christiano L, Kehoe P (1994) Optimal fiscal policy in a business cycle model. J Polit Econ 102(4):617–652
Chari V, Kehoe P (1990) Sustainable plans. J Polit Econ 98(4):783–802
Cole HL, Kocherlakota N (2001) Dynamic games with hidden actions and hidden states. J Econ Theory 98(1):114–126
Cole H, Kubler F (2012) Recursive contracts, lotteries and weakly concave Pareto sets. Rev Econ Dyn 15(4):479–500
Cronshaw MB (1997) Algorithms for finding repeated game equilibria. Comput Econ 10(2): 139–168
Cronshaw MB, Luenberger DG (1994) Strongly symmetric subgame perfect equilibria in infinitely repeated games with perfect monitoring and discounting. Games Econ Behav 6(2):220–237
Diamond P, Koszegi B (2003) Quasi-hyperbolic discounting and retirement. J Public Econ 87:1839–1872
Dominguez B (2005) Reputation in a model with a limited debt structure. Rev Econ Dyn 8(3): 600–622
Dominguez B, Feng Z (2016a, forthcoming) An evaluation of constitutional constraints on capital taxation. Macroecon Dyn. https://doi.org/10.1017/S1365100515000978
Dominguez B, Feng Z (2016b) The time-inconsistency problem of labor taxes and constitutional constraints. Dyn Games Appl 6(2):225–242

Doraszelski U, Escobar JF (2012) Restricted feedback in long term relationships. J Econ Theory 147(1):142–161
Durdu C, Nunes R, Sapriza H (2013) News and sovereign default risk in small open economies. J Int Econ 91:1–17
Dutta PK, Sundaram R (1992) Markovian equilibrium in a class of stochastic games: existence theorems for discounted and undiscounted models. Econ. Theory 2(2):197–214
Ely JC, Hörner J, Olszewski W (2005) Belief-free equilibria in repeated games. Econometrica 73(2):377–415
Farhi E, Sleet C, Werning I, Yeltekin S (2012) Nonlinear capital taxation without commitment. Rev Econ Stud 79:1469–1493
Feng Z (2015) Time consistent optimal fiscal policy over the business cycle. Quant Econ 6(1): 189–221
Feng Z, Miao J, Peralta-Alva A, Santos MS (2014) Numerical simulation of nonoptimal dynamic equilibrium models. Int Econ Rev 55(1):83–110
Fesselmeyer E, Mirman LJ, Santugini M (2016) Strategic interactions in a one-sector growth model. Dyn Games Appl 6(2):209–224
Fischer S (1980a) Dynamic inconsistency, cooperation and the benevolent dissembling government. J Econ Dyn Control 2(1):93–107
Fischer S (1980b) On activist monetary policy and rational expectations. In: Fischer S (ed) Rational expectations and economic policy. University of Chicago Press, Chicago, pp 211–247
Fischer RD, Mirman LJ (1992) Strategic dynamic interaction: fish wars. J Econ Dyn Control 16(2):267–287
Fudenberg D, Levine DK (2006) A dual-self model of impulse control. Am Econ Rev 96(5): 1449–1476
Fudenberg D, Yamamoto Y (2011) The folk theorem for irreducible stochastic games with imperfect public monitoring. J Econ Theory 146(4):1664–1683
Harris C, Laibson D (2001) Dynamic choices of hyperbolic consumers. Econometrica 69(4): 935–957
Harris C, Laibson D (2013) Instantaneous gratification. Q J Econ 128(1):205–248
Hellwig C, Lorenzoni G (2009) Bubbles and self-enforcing debt. Econometrica 77(4):1137–1164
Hörner J, Sugaya T, Takahashi S, Vieille N (2011) Recursive methods in discounted stochastic games: an algorithm for $\delta \to 1$ and a folk theorem. Econometrica 79(4):1277–1318
Jaśkiewicz A, Nowak AS (2014) Stationary Markov perfect equilibria in risk sensitive stochastic overlapping generations models. J Econ Theory 151:411–447
Jaśkiewicz A, Nowak AS (2015) Stochastic games of resource extraction. Automatica 54:310–316
Judd K (1985) Redistributive taxation in a simple perfect foresight model. J Public Econ 28(1): 59–83
Judd K, Cai Y, Yeltekin S (2015) Computing equilibria of dynamic games. Discussion paper, MS
Judd K, Yeltekin S, Conklin J (2003) Computing supergame equilibria. Econometrica 71(4): 1239–1254
Kandori M (2002) Introduction to repeated games with private monitoring. J Econ Theory 102(1):1–15
Kehoe TJ, Levine DK (1993) Debt-constrained asset markets. Rev Econ Stud 60(4):865–888
Kehoe TJ, Levine DK (2001) Liquidity constrained markets versus debt constrained markets. Econometrica 69(3):575–98
Kitti M (2013) Conditional Markov equilibria in discounted dynamic games. Math Meth Oper Res 78(1):77–100
Kitti M (2016) Subgame perfect equilibria in discounted stochastic games. J Math Anal Appl 435(1):253–266
Klein P, Krusell P, Ríos-Rull J-V (2008) Time-consistent public policies. Rev Econ Stud 75: 789–808
Klein P, Quadrini V, Ríos-Rull J-V (2005) Optimal time-consistent taxation with international mobility of capital. BE J Macroecon 5(1):1–36

Klein P, Ríos-Rull J-V (2003) Time-consistent optimal fiscal policy. Int Econ Rev 44(4): 1217–1245
Koeppl T (2007) Optimal dynamic risk sharing when enforcement is a decision variable. J Econ Theory 134:34–69
Krebs T, Kuhn M, Wright M (2015) Human capital risk, contract enforcement, and the macroeconomy. Am Econ Rev 105:3223–3272
Krusell P, Kurusçu B, Smith A (2002) Equilibrium welfare and government policy with quasi-geometric discounting. J Econ Theory 105(1):42–72
Krusell P, Kurusçu B, Smith A (2010) Temptation and taxation. Econometrica 78(6):2063–2084
Krusell P, Smith A (2003) Consumption–savings decisions with quasi–geometric discounting. Econometrica 71(1):365–375
Krusell P, Smith A (2008) Consumption-savings decisions with quasi-geometric discounting: the case with a discrete domain. MS. Yale University
Kydland F, Prescott E (1977) Rules rather than discretion: the inconsistency of optimal plans. J Polit Econ 85(3):473–491
Kydland F, Prescott E (1980) Dynamic optimal taxation, rational expectations and optimal control. J Econ Dyn Control 2(1):79–91
Laibson D (1994) Self-control and savings. PhD thesis, MIT
Laibson D (1997) Golden eggs and hyperbolic discounting. Q J Econ 112(2):443–477
Laitner J (1979a) Household bequest behaviour and the national distribution of wealth. Rev Econ Stud 46(3):467–83
Laitner J (1979b) Household bequests, perfect expectations, and the national distribution of wealth. Econometrica 47(5):1175–1193
Laitner J (1980) Intergenerational preference differences and optimal national saving. J Econ Theory 22(1):56–66
Laitner J (2002) Wealth inequality and altruistic bequests. Am Econ Rev 92(2):270–273
Leininger W (1986) The existence of perfect equilibria in model of growth with altruism between generations. Rev Econ Stud 53(3):349-368
Le Van C, Saglam H (2004) Optimal growth models and the Lagrange multiplier. J Math Econ 40:393–410
Levhari D, Mirman L (1980) The great fish war: an example using a dynamic Cournot-Nash solution. Bell J Econ 11(1):322–334
Loury G (1981) Intergenerational transfers and the distributions of earnings. Econometrica 49: 843–867
Maliar L, Maliar S (2005) Solving the neoclassical growth model with quasi-geometric discounting: a grid-based Euler-equation method. Comput Econ 26:163–172
Maliar L, Maliar S (2006a) Indeterminacy in a log-linearized neoclassical growth model with quasi-geometric discounting. Econ Model 23/3:492–505
Maliar L, Maliar S (2006b) The neoclassical growth model with heterogeneous quasi-geometric consumers. J Money Credit Bank 38:635–654
Maliar L, Maliar S (2016) Ruling out multiplicity of smooth equilibria in dynamic games: a hyperbolic discounting example. Dyn Games Appl 6(2):243–261
Marcet A, Marimon R (1998) Recursive contracts. Economics working papers, European University Institute
Marcet A, Marimon R (2011) Recursive contracts. Discussion paper. Barcelona GSE working paper series working paper no. 552
Mertens J-F, Parthasarathy T (1987) Equilibria for discounted stochastic games. C.O.R.E. discussion paper 8750
Mertens J-F, Parthasarathy T (2003) Equilibria for discounted stochastic games. In: Neyman A, Sorin S (eds) Stochastic games and applications. NATO advanced science institutes series D: behavioural and social sciences. Kluwer Academic, Boston
Mertens J-F, Sorin S, Zamir S (2015) Repeated games. Cambridge University Press, New York
Messner M, Pavoni N (2016) On the recursive saddle point method. Dyn Games Appl 6(2): 161–173

Messner M, Pavoni N, Sleet C (2012) Recursive methods for incentive problems. Rev Econ Dyn 15(4):501–525
Messner M, Pavoni N, Sleet C (2014) The dual approach to recursive optimization: theory and examples. Discussion paper 1267, Society for Economic Dynamics
Mirman L (1979) Dynamic models of fishing: a heuristic approach. In: Liu PT, Sutinen JG (ed) Control theory in mathematical economics. Decker, New York, pp 39–73
Morand O, Reffett K, Tarafdar S (2015) A nonsmooth approach to envelope theorems. J Math Econ 61(C):157–165
Nowak AS (2006a) A multigenerational dynamic game of resource extraction. Math Soc Sci 51(3):327–336
Nowak AS (2006b) A note on an equilibrium in the great fish war game. Econ Bull 17(2): 1–10
Nowak AS (2006c) On perfect equilibria in stochastic models of growth with intergenerational altruism. Econ Theory 28(1):73–83
Obara I, Park J (2013) Repeated games with general time preference. Paper available at: https://editorialexpress.com/cgi-bin/conference/download.cgi?db_name=ESWC2015&paper_id=548
O'Donoghue T, Rabin M (1999) Doing it now or later. Am Econ Rev 89(1):103–124
O'Donoghue T, Rabin M (2001) Choice and procrastination. Q J Econ 116(1):121–160
Pearce D, Stacchetti E (1997) Time consistent taxation by a government with redistributive goals. J Econ Theory 72(2):282–305
Peleg B, Yaari ME (1973) On the existence of a consistent course of action when tastes are changing. Rev Econ Stud 40(3):391–401
Peralta-Alva A, Santos M (2010) Problems in the numerical simulation of models with heterogeneous agents and economic distortions. J Eur Econ Assoc 8:617–625
Phelan C, Stacchetti E (2001) Sequential equilibria in a Ramsey tax model. Econometrica 69(6):1491–1518
Phelps E, Pollak R (1968) On second best national savings and game equilibrium growth. Rev Econ Stud 35:195–199
Pollak RA (1968) Consistent planning. Rev Econ Stud 35(2):201–208
Ray D (1987) Nonpaternalistic intergenerational altruism. J Econ Theory 41:112–132
Rincón-Zapatero JP, Santos MS (2009) Differentiability of the value function without interiority assumptions. J Econ Theory 144(5):1948–1964
Rockafellar TR, Wets RJ-B (2009) Variational analysis. Springer, Berlin
Rogoff K (1987) Reputational constraints on monetary policy. Carn-Roch Ser Public Policy 26:141–182
Rustichini A (1998a) Dynamic programming solution of incentive constrained problems. J Econ Theory 78(2):329–354
Rustichini A (1998b) Lagrange multipliers in incentive-constrained problems. J Math Econ 29(4):365–380
Santos M, Miao J (2005) Numerical solution of dynamic non-optimal economies. Discussion paper, Society for Economic Dynamics, 2005 meeting papers
Sleet C (1998) Recursive models for government policy. PhD thesis, Stanford University
Sleet C (2001) On credible monetary policy and private government information. J Econ Theory 99(1–2):338–376
Sleet C, Yeltekin S (2006a) Credibility and endogenous social discounting. Rev Econ Dyn 9: 410–437
Sleet C, Yeltekin S (2006b) Optimal taxation with endogenously incomplete debt markets. J Econ Theory 127:36–73
Sleet C, Yeltekin S (2007) Recursive monetary policy games with incomplete information. J Econ Dyn Control 31(5):1557–1583
Sleet C, Yeltekin S (2016) On the computation of value correspondences. Dyn Games Appl 6(2):174–186
Straub L, Werning I (2014) Positive long run capital taxation: Chamley-Judd revisited. Discussion paper 20441, National Bureau of Economic Research

Strotz RH (1955) Myopia and inconsistency in dynamic utility maximization. Rev Econ Stud 23(3):165–180

Sundaram RK (1989a) Perfect equilibrium in non-randomized strategies in a class of symmetric dynamic games. J Econ Theory 47(1):153–177

Sundaram RK (1989b) Perfect equilibrium in non-randomized strategies in a class of symmetric dynamic games: corrigendu. J Econ Theory 49:358–387

Woźny Ł, Growiec J (2012) Intergenerational interactions in human capital accumulation. BE J Theor Econ 12

Yamamoto Y (2010) The use of public randomization in discounted repeated games. Int J Game Theory 39(3):431–443

Yue VZ (2010) Sovereign default and debt renegotiation. J Int Econ 80(2):176–187

Differential Games in Industrial Organization

18

Luca Colombo and Paola Labrecciosa

Contents

Abstract

In this chapter, we provide an overview of continuous-time games in industrial organization, covering classical papers on adjustment costs, sticky prices, and R&D races, as well as more recent contributions on oligopolistic exploitation of renewable productive assets and strategic investments under uncertainty.

Keywords

Differential oligopoly games · Adjustment costs · Sticky prices · Productive assets · Innovation

L. Colombo (✉)
Department of Economics, Deakin University, Burwood, VIC, Australia
e-mail: luca.colombo@deakin.edu.au

P. Labrecciosa
Department of Economics, Monash University, Clayton, VIC, Australia
e-mail: paola.labrecciosa@monash.edu

T. Başar, G. Zaccour (eds.), *Handbook of Dynamic Game Theory*,
https://doi.org/10.1007/978-3-319-44374-4_17

1 Introduction

In the last three decades, dynamic noncooperative game theory has become one of the main tools of analysis in the field of industrial organization, broadly defined. The reason is that dynamic games provide a better understanding of the dynamic strategic interactions among firms than static games, in which the time dimension is not explicitly taken into account. Many decisions that firms make in a competitive environment are dynamic in nature: not only they have long-lasting effects but they also shape the future environment in which firms will operate. In a dynamic competitive environment, a firm has incentives to behave strategically (e.g., trying to condition rivals' responses) so as to make the future environment more profitable. The importance of a dynamic perspective on oligopoly markets has been stressed in Cabral (2012), who argues that, "... dynamic oligopoly models provide considerable value added with respect to static models."

In this chapter, we provide a selective survey of differential game models applied to industrial organization. Differential games constitute a class of decision problems wherein the evolution of the state is described by a differential equation and the players act throughout a time interval (see Başar and Olsder 1995; Dockner et al. 2000; Haurie et al. 2012, Ch. 7). As pointed out in Vives (1999), "Using the tools provided by differential games, the analysis can be extended to competition in continuous time.[1] The result is a rich theory which explains a variety of dynamic patterns ...." We focus on differential oligopoly games in which firms use Markov strategies. A Markov strategy is a decision rule that specifies a firm's action as a function of the current state (assumed to be observable by all firms), which encapsulates all the relevant history of the game. This type of strategy permits a firm's decision to be responsive to the actual evolution of the state of the system (unlike open-loop strategies, according to which firms condition their actions only on calendar time and are able to commit themselves to a pre-announced path over the remaining planning period). As argued in Maskin and Tirole (2001), the interest in Markov strategies is motivated by at least three reasons. First, they produce subgame perfect equilibria. Secondly, they are widely used in applied dynamic game theory, thus justifying further theoretical attention. Thirdly, their simplicity makes econometric estimation and inference easier, and they can readily be simulated. Throughout this chapter, firms are assumed to maximize the discounted sum of their own profit streams subject to the relevant dynamic constraints, whereas consumers are static maximizers, and their behavior can be summarized by either a static demand function (the same for all periods) or a demand function augmented by a state variable.[2]

[1] Continuous time can be interpreted as "discrete time, but with a grid that is infinitely fine" (see Simon and Stinchcombe 1989, p.1171).

[2] An interesting area where dynamic games have been fruitfully applied in industrial organization is where firms face sophisticated customers that can foresee their future needs. For a survey of this literature, see Long (2015).

In Sect. 2, we review differential games with adjustment costs. The idea behind the papers considered in this section is that some payoff relevant variables are costly to adjust. For instance, capacity can be costly to expand because firms need to buy new machineries and equipment (and hire extra workers) and re-optimize the production plan. Or prices can be costly to adjust due to menu costs and costs associated with providing consumers with information about price reductions and promotions. Since the open-loop equilibria of these dynamic games coincide with the equilibria of the corresponding static games (in which state variables become control variables), and open-loop equilibria are not always subgame perfect, the lesson that can be drawn is that static models do not capture the long-run stable relationships among firms in industries where adjustment costs are important. The degree of competitiveness is either under- or overestimated by static games, depending on whether Markov control-state substitutability or Markov control-state complementarity prevails.

In Sect. 3, we present the classical sticky price model of oligopolistic competition and some important extensions and applications. Price stickiness is relevant in many real-world markets where firms have control over their output levels but the market price takes time to adjust to the level indicated by the demand function. The demand function derives from a utility function that depends on both current and past consumption levels. The take-away message is that, even in the limiting case in which the price adjusts instantaneously (therefore dynamics is removed), the static Cournot equilibrium price does not correspond to the limit of the steady-state feedback equilibrium price, implying that the static model cannot be considered as a reduced form of a full-fledged dynamic model with price stickiness. Interestingly, the most efficient outcome can be supported as a subgame perfect equilibrium of the dynamic game for some initial conditions, provided that firms use nonlinear feedback strategies and that the discount rate tends to zero.

In Sect. 4, we deal with productive asset games of oligopolistic competition, focussing on renewable assets. In this class of games, many of the traditional results from static oligopoly theory are challenged, giving rise to new policy implications. For instance, in a dynamic Cournot game where production requires exploitation of a common-pool renewable productive asset, an increase in market competition, captured by an increase in the number of firms, could be detrimental to welfare, with important consequences for merger analysis. Moreover, allowing for differentiated products, Cournot competition could become more efficient than Bertrand competition. The main reason for these novel results to arise is the presence of an intertemporal negative externality (due to the lack of property rights on the asset), which is not present in static models.

In Sect. 5, we consider games of innovation, either stochastic or deterministic. In the former, firms compete in R&D *for* potential markets, in the sense that they engage in R&D races to make technological breakthrough which enable them to get monopolistic rents. In the latter, firms compete in R&D *in* existing markets, meaning that all firms innovate over time while competing in the market place. From the collection of papers considered in this section, we learn how incentives for firms to innovate, either noncooperatively or cooperatively, depend on a number of factors, including the length of patent protection, the degree to which own knowledge spills over to other firms, firms location, as well as the nature and the degree of market competition.

In Sect. 6, we examine recent contributions that emphasize timing decisions, following a real options approach. This is a stream of literature which is growing in size and recognition. It extends traditional real options theory, which studies investments under uncertainty (without competition), to a competitive environment where the optimal exercise decision of a firm depends not only on the value of the underlying economic variables (e.g., the average growth rate and the volatility of a market) but also on rivals' actions. Real options models provide new insights into the evolution of an industry structure, in particular in relation to the entry process (timing, strategic deterrence, and investment patterns). Concluding remarks are provided in Sect. 7.

2 Adjustment Costs

In this section, we present a number of differential oligopoly games whose common denominator is given by the presence of adjustment costs.[3] We first consider costly output adjustments, with and without sluggish demand, and costly price adjustments. Next, we turn our attention to capacity investments with capital adjustment costs.

We start from the general dynamic Cournot duopoly model with output adjustment costs analyzed in Dockner (1992), who generalizes Driskill and McCafferty (1989) to account for asymmetric technologies across firms and explores the relationship between dynamic duopolistic competition and static conjectural variations equilibria.[4] The product price is related to industry output by means of an inverse demand function $p(Q(t))$, decreasing in Q, where $p(\cdot)$ is the price at time t and $Q(t) = q_1(t) + q_2(t)$ is industry output at time t, i.e., the sum of outputs produced by each firm. The cost of production is given by $C_i(q_i(t))$, with $C_i'' > p'$. In addition, it is assumed that each firm faces adjustment costs when scaling up (or down) output. Let $u_i = \dot{q}_i$ denote the rate of change of output of Firm i at time t. Adjustment costs are described by the cost function $A_i(u_i(t))$, assumed to be convex. Firms choose their production plans over an infinite time horizon with the aim to maximize the discounted sum of profits. The common discount factor is denoted by r.

Under the open-loop information structure, the dynamic game under consideration admits a steady-state equilibrium that coincides with the Cournot equilibrium

[3] Adjustment costs are important in quite a few industries as evidenced by several empirical studies (e.g., Hamermesh and Pfann 1996; Karp and Perloff 1989, 1993a,b).

[4] Driskill and McCafferty (1989) derive a closed-loop equilibrium in the case in which quantity-setting firms face a linear demand for a homogeneous product and bear symmetric quadratic costs for changing their output levels.

of the corresponding static game. This implies that the static Cournot game is able to capture the long-run stable dynamic interactions between firms, i.e., the static Cournot outcome is a good prediction of the equilibrium of the infinite horizon dynamic game. However, as is well-known in the differential game literature, under the open-loop information structure, firms base their strategies only on the information available at the beginning of the game, and they do not revise their plans as new information becomes available. For this reason, the open-loop information is typically not subgame perfect.

Assuming firms use feedback strategies, Firm i's problem can be written as $(i, j = 1, 2, j \neq i)$

$$\begin{cases} \max_{u_i} \int_0^{\infty} e^{-rt} \left[p(Q(t))q_i(t) - C_i(q_i(t)) - A_i(u_i(t)) \right] dt \\ \text{s.t. } \dot{q}_i = u_i(t), \dot{q}_j = u_j(q_1(t), q_2(t)), q_i(0) = q_{i0} \geq 0, \end{cases} \tag{18.1}$$

where $u_j(q_1, q_2)$ denotes Firm j's feedback strategy. In this context, a feedback strategy is a decision rule that specifies a firm's current rate of change of its own output as a function of the current state (q_1, q_2). This type of strategy permits a firm's decisions to be responsive to the actual evolution of the state of the system (in contrast to open-loop strategies). A Nash equilibrium in feedback strategies has the perfectness property. That is, feedback Nash equilibrium (or FNE) strategies are equilibrium strategies for any subgame that begins from a time $t \geq 0$ and associated state $(q_1(t), q_2(t))$, even if $(q_1(t), q_2(t))$ is not on the equilibrium state trajectory that begins at $(q_1(0), q_2(0))$.

Recall that, in the static Cournot model, when choosing its optimal production, Firm i takes the production of Firm j as a given. This is referred to as Cournot-Nash conjecture. Alternatively, we can assume that firms have nonzero conjectural variations, i.e., $dq_j/dq_i = \xi \neq 0$. In this case, from the first-order condition for profit maximization, we obtain

$$p(Q) \left[1 - \frac{s_i}{\eta} (1 + \xi) \right] = C_i'(q_i), \tag{18.2}$$

where $s_i = q_i/Q$ is the market share of Firm i and $\eta = -p/[p'Q]$ is the price elasticity of industry demand. Equation (18.2) characterizes a conjectural variations equilibrium. Dockner (1992) shows that the steady-state equilibrium of the dynamic game with firms playing Markov strategies can be viewed as a conjectural variations equilibrium of a static game and that the conjectural variation that Firm i has about the reaction of its rival can be written as

$$\xi = \frac{\partial u_j^*}{\partial q_i} \left[r - \frac{\partial u_j^*}{\partial q_j} \right]^{-1},$$

where u_j^* denotes Firm j's equilibrium strategies. Assuming a linear demand $p = \max\{a - Q, 0\}$, $a > 0$, quadratic costs $A_i = ku_i^2/2$, $k > 0$, and

$C_i = c_i q_i + bq_i^2/2$, $c_i < a$, with a, b, c_i, k constants, equilibrium strategies can be written as

$$u_i^* = \frac{1}{k}\left(\alpha_i + \beta q_i + \gamma q_j\right),$$

where α_i, β, and γ are equilibrium values of the endogenous parameters that depend on the exogenous parameters of the model, with β and γ satisfying $\beta < \gamma < 0$ (for global asymptotic stability). It follows that

$$\xi = \frac{\gamma}{rk - \beta}, \tag{18.3}$$

which is constant and negative (since $\beta, \gamma < 0$). Hence, the steady-state equilibrium of the dynamic game coincides with a conjectural variations equilibrium of the corresponding static game with constant and negative conjectures, given in (18.3). Each firm conjectures that own output expansion will elicit an output contraction by the rival. This will result in a more competitive behavior than in the static Cournot game, where firms make their output decision only on the basis of the residual demand curve, ignoring the reactions of the rivals.

Jun and Vives (2004) extend the symmetric duopoly model with undifferentiated products and costly output adjustments to a differentiated duopoly. A representative consumer has a quadratic and symmetric utility function for the differentiated goods (and utility is linear in money) of the form $U(q_1, q_2) = A(q_1 + q_2) - [B(q_1^2 + q_2^2) + 2Cq_1q_2]/2$ (see Singh and Vives 1984). In the region where prices and quantities are positive, maximization of $U(q_1, q_2)$ w.r.t. q_1 and q_2 yields the following inverse demands $P_i(q_1, q_2) = A - Bq_i - Cq_j$, which can be inverted to obtain $D_i(p_1, p_2) = a - bp_i + cp_j$, with $a = A/(B+C)$, $b = B/(B^2 - C^2)$, and $c = C/(B^2 - C^2)$, $B > |C| \geq 0$. When $B = C$, the two products are homogeneous (from the consumer's view point); when $C = 0$, the products are independent; and when $C < 0$, they are complements. Firms compete either in quantities (à la Cournot) or in prices (à la Bertrand) over an infinite time horizon. The variable that is costly to adjust for each firm is either its production or the price level. Adjustment costs are assumed to be quadratic. The following four cases are considered: (i) output is costly to adjust and firms compete in quantities, (ii) price is costly to adjust and firms compete in prices, (iii) output is costly to adjust and firms compete in prices, and (iv) price is costly to adjust and firms compete in quantities. For all cases, a (linear) feedback equilibrium is derived.

In (i), Firm i's problem can be written as ($i, j = 1, 2$, $j \neq i$)

$$\begin{cases} \max\limits_{u_i} \int_0^\infty e^{-rt}\left[\left(A - Bq_i - Cq_j\right)q_i - \frac{\beta}{2}u_i^2\right]dt \\ \text{s.t. } \dot{q}_i = u_i,\ \dot{q}_j = u_j(q_1, q_2),\ q_i(0) = q_{i0} \geq 0, \end{cases} \tag{18.4}$$

where $\beta u_i^2/2$ is the cost of output adjustments, with $\beta > 0$, u_i is Firm i's rate of output adjustments, $r > 0$ is the discount rate (common to both firms), and $u_j(q_1, q_2)$ is

Firm j's feedback strategy, which Firm i takes as a given. Let $V_i(q_1, q_2)$ denote Firm i's value function, i.e., the discounted value of profits for Firm i for a equilibrium of the game that begins at (q_1, q_2). The pair of strategies $(u_1^*(q_1, q_2),\ u_2^*(q_1, q_2))$ constitutes a feedback equilibrium if $u_1^*(q_1, q_2)$ is a solution of

$$rV_1(q_1, q_2) = \max_{u_1} \left\{ (A - Bq_1 - Cq_2)\, q_1 - \frac{\beta}{2} u_1^2 + \frac{\partial V_1}{\partial q_1} u_1 + \frac{\partial V_1}{\partial q_2} u_2^*(q_1, q_2) \right\},$$

and $u_2^*(q_1, q_2)$ is a solution of

$$rV_2(q_1, q_2) = \max_{u_2} \left\{ (A - Bq_2 - Cq_1)\, q_2 - \frac{\beta}{2} u_2^2 + \frac{\partial V_2}{\partial q_2} u_2 + \frac{\partial V_2}{\partial q_1} u_1^*(q_1, q_2) \right\},$$

and $\lim_{t\to\infty} e^{-rt} V_1(q_1(t), q_2(t)) = \lim_{t\to\infty} e^{-rt} V_2(q_1(t), q_2(t)) = 0$, where $q_i(t)$ is the time path of q_i induced by $(u_1^*(q_1, q_2), u_2^*(q_1, q_2))$. The first-order condition on the control variable yields

$$u_i = \frac{1}{\beta} \frac{\partial V_i}{\partial q_i}. \tag{18.5}$$

The term $\partial V_i / \partial q_i$ captures the shadow price of Firm i's output, i.e., the impact of a marginal increase in Firm i's output on Firm i's discounted sum of profits. Note that the optimality condition (18.5) holds in the region where $P_i(q_1, q_2) \geq 0$ and $q_i \geq 0$, $i = 1, 2$ and that u_i can take either positive or negative values, i.e., it is possible for Firm i to either increase or decrease its production level.[5]

Using (18.5), we obtain $(i, j = 1, 2, i \neq j)$

$$rV_i(q_1, q_2) = \left(A - Bq_i - Cq_j\right) q_i + \frac{1}{2\beta} \left(\frac{\partial V_i}{\partial q_i} \right)^2 + \frac{1}{\beta} \frac{\partial V_i}{\partial q_j} \frac{\partial V_j}{\partial q_j}.$$

Given the linear-quadratic structure of the game, we guess value functions of the form

$$V_i(q_1, q_2) = \kappa_1 q_i^2/2 + \kappa_2 q_j^2/2 + \kappa_3 q_i + \kappa_4 q_j + \kappa_5 q_i q_j + \kappa_6,$$

which imply that

$$u_i = \frac{1}{\beta} \left(\kappa_1 q_i + \kappa_3 + \kappa_5 q_j \right).$$

[5] For a discussion about the nature of capacity investments (reversible vs. irreversible) and its strategic implications in games with adjustment costs, see Dockner et al. (2000, Ch. 9).

There exist six candidates for a feedback equilibrium. The equilibrium values of the endogenous parameters κ_1, κ_3, and κ_5 as functions of the exogenous parameters of the model can be found by identification. Let $K = \kappa_5^2$, which turns out to be the solution of the cubic equation

$$81K^3 + \alpha_1 K^2 + \alpha_2 K + \alpha_3 = 0, \tag{18.6}$$

where $\alpha_1 = -18\beta(8B+\beta r^2)$, $\alpha_2 = \beta^2[8C^2+(8B+\beta r^2)^2]$, and $\alpha_3 = C^2\beta^3(8B+\beta r^2)$; the equilibrium values of κ_1 and κ_3 are given by

$$\kappa_1 = \frac{1}{2}\left(r\beta \pm \sqrt{r^2\beta^2 + 8B\beta - 8K}\right), \tag{18.7}$$

and

$$\kappa_3 = \frac{A\beta}{r\beta - \kappa_1 - \kappa_5}\left[1 + \frac{\kappa_5^2(\kappa_5 - 2\kappa_1 + r\beta)}{(2\kappa_1 - r\beta)(\kappa_1 - r\beta)(\kappa_1 + \kappa_5 - r\beta)}\right]^{-1}. \tag{18.8}$$

Out of the six candidates for a feedback equilibrium, there exists only one stabilizing the state (q_1, q_2). The coefficients of the equilibrium strategy u_i^* satisfy $\kappa_1 < \kappa_5 < 0$ and $\kappa_3 > 0$. The couple (u_1^*, u_2^*) induces a trajectory of q_i that converges to $q_\infty = -\kappa_3/(\kappa_1+\kappa_5)$ for every possible initial conditions. The fact that $\partial u_j^*/\partial q_i < 0$ (since $\kappa_5 < 0$) implies that there exists intertemporal strategic substitutability: an output expansion by Firm i today leads to an output contraction by the rival in the future.[6] Since the game is symmetric, each firm has a strategic incentive to expand its own output so as to make the rival smaller. This incentive is not present in the open-loop equilibrium, where firms do not take into account the impact of a change in the state variables on their optimal strategies. At the steady-state, Markovian behavior turns out to be more aggressive than open-loop (static) behavior.

In (ii), Firm i's problem can be written as $(i, j = 1, 2, j \neq i)$

$$\begin{cases} \max\limits_{u_i} \int_0^\infty e^{-rt}\left[\left(a - bp_i + cp_j\right)p_i - \frac{\beta}{2}u_i^2\right]dt \\ \text{s.t. } \dot{p}_i = u_i,\ \dot{p}_j = u_j(p_1, p_2),\ p_i(0) = p_{i0} \geq 0, \end{cases} \tag{18.9}$$

which is formally identical to (18.4) once one replaces q_i with p_i, A with a, B with b, and C with $-c$. Indeed, as pointed out in Jun and Vives (2004), Bertrand competition with costly price adjustment is the dual of Cournot competition with costly output adjustment. The equilibrium strategy can be found by setting $A = a$, $B = b$, and $C = -c$ in (18.6), (18.7), and (18.8). By exploiting duality, we can see that $\partial u_j^*/\partial p_i > 0$, implying that there exists intertemporal strategic complementarity. A firm has a strategic incentive to price more softly today so as to

[6]This is also referred to as Markov control-state substitutability (see Long 2010).

induce a less aggressive price behavior by the rival in the future. The steady-state price turns out to be higher than the static Bertrand equilibrium price.

In (iii), Firm i's problem can be written as $(i, j = 1, 2, j \neq i)$

$$\begin{cases} \max_{u_i} \int_0^\infty e^{-rt} \left[\left(a - bp_i + cp_j\right) p_i - \frac{\beta \left(\dot{q}_i\right)^2}{2} \right] dt \\ \text{s.t. } \dot{p}_i = u_i, \dot{p}_j = u_j(p_1, p_2), p_i(0) = p_{i0} \geq 0, \end{cases} \tag{18.10}$$

where $\dot{q}_i = -b\dot{p}_i + c\dot{p}_j$, implying that the cost for Firm i to adjust its output is directly affected by Firm j. This complicates the analysis of the game compared with (i) and (ii).

Firm i's HJB equation is $(i, j = 1, 2, j \neq i)$

$$rV_i(p_1, p_2) = \max_{u_i} \left\{ \left(a - bp_i + cp_j\right) p_i - \frac{\beta}{2} \left[bu_i - cu_j(p_1, p_2)\right]^2 + \frac{\partial V_i}{\partial p_i} u_i + \frac{\partial V_i}{\partial p_j} u_j(p_1, p_2) \right\}. \tag{18.11}$$

Performing the maximization indicated in (18.11) gives Firm i's instantaneous best response,

$$u_i = \frac{1}{b} \left[\frac{1}{b\beta} \frac{\partial V_i}{\partial p_i} + cu_j \right],$$

which can be used to derive the equilibrium of the instantaneous game given p_1 and p_2,

$$u_i^* = \frac{1}{b\beta\left(b^2 - c^2\right)} \left(b \frac{\partial V_i}{\partial p_i} + c \frac{\partial V_j}{\partial p_j} \right). \tag{18.12}$$

Using (18.12), (18.11) becomes

$$rV_i\left(p_1, p_2\right) = \left(a - bp_i + cp_j\right) p_i + \frac{1}{2b^2\beta\left(b^2 - c^2\right)} \times \left\{ \left(\frac{\partial V_i}{\partial p_i}\right)^2 \left(b^2 + c^2\right) + 2b \left[b \frac{\partial V_i}{\partial p_j} \frac{\partial V_j}{\partial p_j} + c \frac{\partial V_i}{\partial p_i} \left(\frac{\partial V_i}{\partial p_j} + \frac{\partial V_j}{\partial p_j} \right) \right] \right\}.$$

Guessing quadratic value functions of the form

$$V_i(p_1, p_2) = \kappa_1 p_i^2/2 + \kappa_2 p_j^2/2 + \kappa_3 p_i + \kappa_4 p_j + \kappa_5 p_i p_j + \kappa_6,$$

the equilibrium values of the coefficients in u_i^*, κ_1, κ_3, and κ_5 can be obtained by identification. The couple of equilibrium strategies (u_1^*, u_2^*) induces a trajectory of p_i that converges to $p_\infty = -\kappa_3/(\kappa_1 + \kappa_5)$. Contrary to the previous case, $\partial u_j^*/\partial p_i < 0$, meaning that static strategic complementary is turned into intertemporal strategic substitutability. The intuition is that price cutting by a firm today leads to an increase in its rival's marginal cost, thus making the rival less efficient. This induces the rival to price more softly in the future. Each firm has an incentive to cut prices today to soften future competition, and this leads to a steady-state price which is below the static Bertrand equilibrium price. For industries characterized by price competition and output adjustment costs, the static game tends to overestimate the long-run price.

In (iv), Firm i's problem can be written as $(i, j = 1, 2, j \neq i)$

$$\begin{cases} \max_{u_i} \int_0^\infty e^{-rt} \left[\left(A - Bq_i - Cq_j\right) q_i - \frac{\beta\left(\dot{p}_i\right)^2}{2} \right] dt \\ \text{s.t. } \dot{q}_i = u_i, \dot{q}_j = u_j(q_1, q_2), q_i(0) = q_{i0} \geq 0, \end{cases} \tag{18.13}$$

where $\dot{p}_i = -B\dot{q}_i - C\dot{q}_j$, implying that the cost for Firm i to adjust its price is directly affected by Firm j. Note that (18.13) is formally identical to (18.10) once one replaces p_i with q_i, a with A, b with B, and c with $-C$. We obtain the results for Cournot competition with costly price adjustment by duality.

From the above analysis, we can conclude that whether or not Markovian behavior turns out to be more aggressive than open-loop/static behavior depends on the variable that is costly to adjust, not the character of competition (Cournot or Bertrand). Using the framework presented in Long (2010), when the variable that is costly to adjust is output, there exists Markov control-state substitutability. Consequently, Markovian behavior turns out to be more aggressive than open-loop (static) behavior. The opposite holds true in the case in which the variable that is costly to adjust is price. In this case, there exists Markov control-state complementarity, and Markovian behavior turns out to be less aggressive than open-loop (static) behavior.[7]

Wirl (2010) considers an interesting variation of the Cournot duopoly game with costly output adjustment presented above. The following dynamic reduced form model of demand is assumed,

$$\dot{x}(t) = s\left[D(p(t)) - x(t)\right].$$

The idea is that, for any finite $s > 0$, demand at time t, $x(t)$, does not adjust instantaneously to the level specified by the long-run demand, $D(p(t))$; only

[7]Figuières (2009) compares closed-loop and open-loop equilibria of a widely used class of differential games, showing how the payoff structure of the game leads to Markov substitutability or Markov complementarity. Focusing on the steady-states equilibria, he shows that competition intensifies (softens) in games with Markov substitutability (complementarity). Markov substitutability (complementarity) can be considered as the dynamic counterparts of strategic substitutability (complementarity) in Bulow et al. (1985).

when $s \to \infty$, $x(t)$ adjusts instantaneously to $D(p(t))$. Sluggishness of demand characterizes many important real-world nondurable goods markets (e.g., fuels and electricity) in which current demand depends on (long-lasting) appliances and equipment. To obtain analytical solutions, Wirl (2010) assumes that $D(p) = \max\{1-p, 0\}$. Note that there is no product differentiation in this model. Apart from the demand side, the rest of the model is as in case (i).

Equating demand and supply at t gives

$$x(t) = \sum_{i=1}^{n} q_i(t), \quad \dot{x}(t) = \sum_{i=1}^{n} u_i(t).$$

The market clearing price is

$$p(t) = 1 - \sum_{i=1}^{n} q_i(t) - \frac{1}{s} \sum_{i=1}^{n} u_i(t).$$

Wirl (2010) focusses on a symmetric ($q_i = q$ and $u_i = u$ $\forall i$) feedback equilibrium. Firm i's HJB equation can be written as ($i, j = 1, 2, \ldots, n$, $j \neq i$)

$$rV_i(q_i, q) = \max_{u_i} \left\{ \left[1 - \sum_{i=1}^{n} q_i - \frac{1}{s} \left(u_i + \sum_{j \neq i}^{n} u_j(q_i, q) \right) \right] q_i - \frac{\beta}{2} u_i^2 + \frac{\partial V_i}{\partial q_i} u_i \right\}. \tag{18.14}$$

Maximization of the RHS of (18.14) yields

$$u_i = \frac{1}{\beta} \left[\frac{\partial V_i}{\partial q_i} - \frac{1}{s} q_i \right],$$

which can be used to derive the following equilibrium strategy,

$$u^* = \zeta (q_\infty - q),$$

with

$$q_\infty = \frac{2\beta}{\beta(2 + r\tau)(1 + n) + (n - 1)\tau \left[2(n - 1)\tau - \sqrt{\kappa} \right]},$$

where ζ is a positive constant that depends on the parameter of the model, $\tau = 1/s$, and $\kappa = (\beta r + 2n\tau)^2 + 8\beta - 4(2n - 1)\tau^2$. Since u^* is decreasing in q then there exists intertemporal strategic substitutability: higher supply of competitors reduces own expansion. Markov strategies can lead to long-run supply exceeding the static Cournot equilibrium, such that preemption due to strategic interactions (if i increases supply, this deters the expansion of j) outweighs the impact of demand sluggishness. If demand adjusts instantaneously ($s \to \infty$), then feedback (and open-loop) steady-state supply coincides with the static one. Indeed, $\lim_{\tau \to 0} q_\infty = 1/(1 + n)$. Interestingly, the long-run supply is non-monotone in s, first decreasing then increasing, and can be either higher or lower than the static benchmark. In conclusion, what drives the competitiveness of a market in relation to the static benchmark is the speed of demand adjustment: when s is small (demand adjusts slowly), the long-run supply is lower than static supply; the opposite holds true when s is large.

We conclude this section with the case in which capacity rather than output is costly to adjust. The difference between output and capacity becomes relevant when capacity is subject to depreciation. Reynolds (1991) assumes that, at any point in time, Firm i's production is equal to its stock of capital, $i = 1, 2, \ldots, n$, with capital depreciating over time at a constant rate $\delta \geq 0$.[8] Firms accumulate capacity according to the following state equations ($i = 1, 2, \ldots, n$),

$$\dot{k}_i(t) = u_i(t) - \delta k_i(t), \quad k_i(0) = k_{i0} \geq 0,$$

where u_i denotes Firm i's rate of capacity investment.[9] Note that capacity is reversible even if $\delta = 0$ since u_i may be negative. A linear inverse demand for homogeneous output is assumed, $p(t) = \max\{0, a - \sum_{i=1}^{n} k_i(t)\}$, where $a > 0$ captures market size. The cost for adjusting the capital stock is given by $C_i(u_i) = \beta u_i^2/2$, $\beta > 0$.[10]

Let $k = (k_1, k_2, \ldots, k_n)$. Firm i's HJB equation is

$$rV_i(k) = \max_{u_i} \left\{ \left(a - \sum_{i=1}^{n} k_i \right) k_i - \frac{\beta}{2} u_i^2 + \frac{\partial V_i}{\partial k_i} u_i + \sum_{j \neq i}^{n} \frac{\partial V_i}{\partial k_j} u_j(k) \right\}. \tag{18.15}$$

Performing the maximization indicated in (18.15) yields

$$u_i = \frac{1}{\beta} \frac{\partial V_i}{\partial k_i}. \tag{18.16}$$

After substituting u_i from (18.16) into (18.15), we obtain

$$rV_i(k) = \left(a - \sum_{i=1}^{n} k_i \right) k_i + \frac{1}{\beta} \left[\frac{1}{2} \left(\frac{\partial V_i}{\partial k_i} \right)^2 + \sum_{j \neq i}^{n} \frac{\partial V_i}{\partial k_j} \frac{\partial V_j}{\partial k_j} \right]. \tag{18.17}$$

[8] The same assumption is made in the duopoly model analyzed in Reynolds (1987).

[9] A different capital accumulation dynamics is considered in Cellini and Lambertini (1998, 2007a), who assume that firms' unsold output is transformed into capital, i.e., firms' capacity is accumulated à la Ramsey (1928).

[10] In most of the literature on capacity investments with adjustment costs, it is assumed that adjustment costs are stock independent. A notable exception is Dockner and Mosburger (2007), who assume that marginal adjustment costs are either increasing or decreasing in the stock of capital.

The value function V_i is specified as a quadratic form in k_i and k_{-i}

$$V_i(k_i, k_{-i}) = \kappa_1 k_i^2/2 + \kappa_2 (k_{-i})^2/2 + \kappa_3 k_i + \kappa_4 k_{-i} + \kappa_5 k_i k_{-i} + \kappa_6,$$

where $k_{-i} = \sum_{j\neq i}^{n} k_i$. If $n = 2$, $0 < \beta \leq 1$, and $0 < r + 2\delta \leq 1$, then there exists only one couple of (linear) strategies stabilizing the state (k_1, k_2) and inducing a trajectory of k_i that converges to

$$k_\infty = \frac{a}{3 + \beta\delta (r + \delta) - y^*/[x^* - \beta (r + \delta)]},$$

for every k_{i0}, where x^* and y^* are constants that depend on the parameters of the model satisfying $x^* < y^* < 0$. For $n \geq 2$ and $\delta = 0$, there exists an $\hat{r}$ such that for $r < \hat{r}$ the vector of (linear) strategies stabilizing the state k induces a trajectory that converges to

$$k_\infty = \frac{a}{1 + n - (n-1) y^*/[x^* + (n-2) y^* - \beta r]},$$

for every k_{i0}. Note that when $\delta = 0$, the game can be reinterpreted as an n-firm Cournot game with output adjustment costs (an extension of Driskill and McCafferty 1989, to an oligopolistic setting). Equilibrium strategies have the property that each firm's investment rate is decreasing in capacity held by any rival (since $\partial u_j^*/\partial k_i = y^*/\beta < 0$). By expanding its own capacity, a firm preempts investments by the rivals in the future. As in the duopoly case, the strategic incentives for investment that are present in the feedback equilibrium lead to a steady-state outcome that is above the open-loop steady state.

As to the comparison between k_∞ and the static Cournot equilibrium output $q^C = a/(1+n)$, since $x^* < y^* < 0$, then $k_\infty > q^C$. The incentives to preempt subsequent investment of rivals in the dynamic model lead to a steady state with capacities and outputs greater than the static Cournot equilibrium output levels. This holds true also in the limiting case in which adjustment costs tend to zero ($\beta \to 0$). However, as $\beta \to \infty$, k_∞ approaches q^C. Finally, for large n, Reynolds (1991) shows that total output is approximately equal to the socially optimal output, not only at the steady state but also in the transition phase.

3 Sticky Prices

In this section, we review classical papers in industrial organization that belong to the class of Cournot games with sticky prices, whose roots can be traced back to Roos (1925, 1927). The main reference is the seminal paper by Fershtman and Kamien (1987).[11] The key feature of sticky price games is that the price for a homogenous product supplied by oligopolistic firms does not adjust instantaneously to the level specified by its demand function (for a given level of output). The demand function can be derived from the maximization of a utility function that depends on current and past consumption, $U(\cdot) = u(q(t), z(t)) + y(t)$, where $q(t)$ represents current consumption, $z(t)$ represents exponentially weighted accumulated past consumption, and $y(t) = m - p(t)q(t)$ represents the money left over after purchasing q units of the good, with $m \geq 0$ denoting consumers' income (exogenously given). Maximization of $U(\cdot)$ w.r.t. q yields $p(\cdot) = \partial u(q(t), z(t))/\partial q(t)$. If the marginal utility of current consumption is concentrated entirely on present consumption, then the price at time t will depend only on current consumption; otherwise it will depend on both current and past consumption. Fershtman and Kamien (1987) consider the following differential equation governing the change in price,

$$\dot{p}(t) = s\left[a - q(t) - p(t)\right], \quad p(0) = p_0 \tag{18.18}$$

where $a \geq 0$ is the demand intercept, $q = \sum_{i=1}^{n} q_i$ is industry output (the sum of outputs produced by each firm), and $s \geq 0$ denotes the speed at which the price converges to its level on the demand function, with $a - q(t)$ being the price on the demand function for the given level of output.[12] (18.18) can be rewritten as a dynamic demand function of the form

$$p(t) = a - q(t) - \frac{\dot{p}(t)}{s}.$$

In the remainder, we set $a = 1$. The production cost of Firm i is given by $C(q_i) = cq_i + q_i^2/2$, with $c \geq 0$. Production costs are assumed to be quadratic so as to make instantaneous profits concave in the control variable.

Firm i's problem can be written as $(i, j = 1, 2, \ldots, n,\ j \neq i)$

$$\begin{cases} \max\limits_{q_i} \displaystyle\int_0^\infty e^{-rt} \pi_i \, dt \\ \text{s.t. } \dot{p} = s\left[1 - q_i - \sum_{j=1, j\neq i}^{n} q_j(p) - p\right], \ p(0) = p_0, \end{cases}$$

[11] Applications of sticky price models of oligopolistic competition with international trade can be found in Dockner and Haug (1990, 1991), Driskill and McCafferty (1996), and Fujiwara (2009).

[12] Equation (18.18) can be derived from the inverse demand function $p(t) = a - s\int_0^t e^{-s(t-\tau)} q(\tau)\, d\tau$. Simaan and Takayama (1978) assume that the speed of price adjustment, s, is equal to one. Fershtman and Kamien (1987) analyze the duopoly game. For the n-firm oligopoly game, see Dockner (1988) and Cellini and Lambertini (2004). Differentiated products are considered in Cellini and Lambertini (2007b).

where $\pi_i = pq_i - C(q_i)$ is Firm i's instantaneous profit, $r > 0$ is the discount rate, common to all firms, and $q_j(p)$ is Firm j's feedback strategy.[13] Firm i's HJB equation is

$$rV_i(p) = \max_{q_i} \left\{ (p-c)\, q_i - \frac{q_i^2}{2} + \frac{\partial V_i}{\partial p} s \left[1 - q_i - \sum_{j=1, j \neq i}^{n} q_j(p) - p \right] \right\}. \tag{18.19}$$

Maximization of the RHS of (18.19) yields (assuming inner solutions exist)

$$q_i = p - c - s\frac{\partial V_i}{\partial p}. \tag{18.20}$$

Note that when $s = 0$, the price is constant. Firm i becomes price taker and thus behaves as a competitive firm, i.e., it chooses a production level such that the price is equal to its marginal cost $(c + q_i)$. When $s > 0$, Firm i realizes that an increase in own production will lead to a decrease in price. The extent of price reduction is governed by s: a larger s is associated with a larger price reduction. Using (18.20), (18.19) becomes

$$\begin{aligned} rV_i =& (p-c)\left(p - c - s\frac{\partial V_i}{\partial p}\right) - \frac{1}{2}\left(p - c - s\frac{\partial V_i}{\partial p}\right)^2 \\ &+ \frac{\partial V_i}{\partial p} s \left[1 - \left(p - c - s\frac{\partial V_i}{\partial p}\right) - \sum_{j=1, j\neq i}^{n} \left(p - c - s\frac{\partial V_j}{\partial p}\right) - p \right]. \end{aligned} \tag{18.21}$$

Given the linear-quadratic structure of the game, the "guess" for the value functions is as follows,

$$V_i(p) = \frac{\kappa_1}{2} p^2 + \kappa_2 p + \kappa_3,$$

yielding the following equilibrium strategies,

$$q_i^* = p(1 - s\kappa_1) - c - s\kappa_2,$$

where

$$\kappa_1 = \frac{r + 2s(1+n) - \sqrt{[r + 2s(1+n)]^2 - 4s^2(2n-1)}}{2s^2(2n-1)},$$

[13]In a finite horizon model, Fershtman and Kamien (1990) analyze the case in which firms use nonstationary feedback strategies.

and

$$\kappa_2 = \frac{s\kappa_1 (1 + cn) - c}{r + s(1 + n) + s^2\kappa_1 (1 - 2n)}.$$

These equilibrium strategies are admissible as long as $p > \widehat{p} = (c + s\kappa_2)/(1 - s\kappa_1)$. When $p \leq \widehat{p}$, corner solutions prevail. Indeed, there exists a threshold of p below which firms refrain from producing. Equilibrium strategies are increasing in p (since $1 - s\kappa_1 > 0$), implying that firms increase their production as the price increases.[14] Each firm knows that an increase in own production, by causing the price to decrease, will lead to a decrease in rivals' production. This is exactly what happens in a conjectural variations equilibrium with negative conjectures (see Dockner 1992). Firm i knows that each of its rivals will react by making an opposite move, and this reaction will be taken into account in the formulation of its optimal strategy. Any attempt to move along the reaction function will cause the reaction function to shift. As in the Cournot game with output or capital adjustment costs, for each firm, there exists a dynamic strategic incentive to expand own production so as to make the rivals smaller (intertemporal strategic substitutability). Intertemporal strategic substitutability turns out to be responsible for a lower steady-state equilibrium price compared with the open-loop case.[15]

The following holds,

$$p_\infty^{CL} = \frac{1 + n(c + s\kappa_2)}{1 + n(1 - s\kappa_1)} < p_\infty^{OL} = \frac{r + 2s + cn(r + s)}{r + 2s + n(r + s)} < p^C = \frac{2 + cn}{2 + n},$$

where p_∞^{CL} and p_∞^{OL} stand for steady-state closed-loop (feedback) and open-loop equilibrium price, respectively, and p^C stands for static Cournot equilibrium price. The above inequality implies that Markovian behavior turns out to be more aggressive than open-loop behavior, which turns out to be more aggressive than static behavior. In the limiting case in which the price jumps instantaneously to the level indicated by the demand function, we have

$$\lim_{s\to\infty} p_\infty^{CL} < \lim_{s\to\infty} p_\infty^{OL} = p^C.$$

Hence, the result that Markovian behavior turns out to be more aggressive than open-loop behavior persists also in the limiting case $s \to \infty$.[16] This is in contrast with Wirl (2010), in which Markovian behavior turns out to be more aggressive than open-loop behavior only for a finite s. As s tends to infinity, feedback and

[14]Firms' capacity constraints are considered in Tsutsui (1996).

[15]For an off-steady-state analysis, see Wiszniewska-Matyszkiel et al. (2015).

[16]Note that when $s \to \infty$, the differential game becomes a continuous time repeated game.

open-loop steady-state prices converge to the same price, corresponding to the static Cournot equilibrium price. Intuitively, in Wirl (2010), as $s \to \infty$, the price does not depend on the control variables anymore; therefore, the preemption effect vanishes. This is not the case in Fershtman and Kamien (1987), in which the price depends on the control variables also in the limiting case of instantaneous price adjustment. In line with Wirl (2010), once price stickiness is removed, open-loop and static behavior coincide. Finally, as $s \to 0$ or $n \to \infty$, feedback and open-loop equilibria converge to perfect competition. The main conclusion of the above analysis is that, compared with the static benchmark, the presence of price stickiness makes the oligopoly equilibrium closer to the competitive equilibrium, even when price stickiness is almost nil. It follows that the static Cournot model tends to overestimate the long-run price prevailing in industries characterized by price rigidities.

A question of interest is how an increase in market competition, captured by an increase in the number of firms, impacts on the equilibrium strategies of the dynamic Cournot game with sticky prices analyzed in Fershtman and Kamien (1987). Dockner (1988) shows that, as the number of firms goes to infinity, the stationary equilibrium price p_{∞}^{CL} approaches the competitive level. This property is referred to as quasi-competitiveness of the Cournot equilibrium.[17] As demonstrated in Dockner and Gaunersdorfer (2002), such a property holds irrespective of the time horizon and irrespective of whether firms play open-loop or closed-loop (feedback) strategies.

Dockner and Gaunersdorfer (2001) analyze the profitability and welfare consequences of a merger, modeled as an exogenous change in the number of firms in the industry from n to $n-m+1$, where m is the number of merging firms. The merged entity seeks to maximize the sum of the discounted profits of the participating firms with the remaining $n-m$ firms playing noncooperatively à la Cournot. The post-merger equilibrium corresponds to a noncooperative equilibrium of the game played by $n-m+1$ firms. The post-merger equilibrium is the solution of the following system of HJB equations,

$$rV_M(p) = \max_{q_1,\ldots,q_m} \left\{ (p-c) \sum_{i=1}^{m} q_i - \frac{1}{2} \sum_{i=1}^{m} q_i^2 + \frac{\partial V_M}{\partial p} s\dot{p} \right\},$$

and

$$rV_j(p) = \max_{q_j} \left\{ (p-c)\, q_j - \frac{1}{2} q_j^2 + \frac{\partial V_j}{\partial p} s\dot{p} \right\},$$

[17] Classical references on the quasi-competitiveness property for a Cournot oligopoly are Ruffin (1971) and Okuguchi (1973).

with $j = 1, 2, \ldots, n - m$, where V_M is the value function of the merged entity and V_j the value function of a generic firm outside the merger. For any subgame that starts at p, a merger is profitable if

$$\frac{V_M(p)}{m} > V_{NC}(p),$$

where $V_{NC}(p)$ corresponds to the value function of one of the n firms in the non cooperative equilibrium (in which no merger takes place). Focussing on the limiting case in which the speed of adjustment of the price tends to infinity ($s \to \infty$), Dockner and Gaunersdorfer (2001) show that, in contrast with static oligopoly theory, according to which the number of merging firms must be sufficiently high for a merger to be profitable, any merger is profitable, independent of the number of merging firms.[18] Since the equilibrium price after a merger always increases, consumers' surplus decreases, and overall welfare decreases.

So far we have focussed on the case in which firms use linear feedback strategies. However, as shown in Tsutsui and Mino (1990), in addition to the linear strategy characterized in Fershtman and Kamien (1987), there exists also a continuum of nonlinear strategies (although the unique strategy that can be defined over the entire state space is the linear one). Instead of working with the ordinary differential equations in the value functions, Tsutsui and Mino (1990) derive a system of differential equations in the shadow prices.[19] Assuming twice differentiability of the value functions, and considering the duopoly case, differentiation of (21) yields ($i, j = 1, 2,\ j \neq i$)

$$r\frac{\partial V_i}{\partial p} = -c - s^2\frac{\partial V_i}{\partial p}\frac{\partial^2 V_i}{\partial p^2} + p + s\left\{\frac{\partial^2 V_i}{\partial p^2}\left[1 - 3p + 2c + s\left(\frac{\partial V_i}{\partial p} + \frac{\partial V_j}{\partial p}\right)\right]\right.$$
$$\left. + \frac{\partial V_i}{\partial p}\left[s\left(\frac{\partial^2 V_i}{\partial p^2} + \frac{\partial^2 V_j}{\partial p^2}\right) - 3\right]\right\}.$$

Solving for the highest derivative and using symmetry gives

$$\frac{\partial y}{\partial p} = \frac{c - p + y(r + 3s)}{s[2(c - p) + 3sy - p + 1]}, \tag{18.22}$$

where $y = \partial V/\partial p$. Following Tsutsui and Mino (1990), it can be checked that the lowest and the highest price that can be supported as steady-state prices by nonlinear

[18]This result can also be found in Benchekroun (2003b)

[19]On the shadow price system approach, see also Wirl and Dockner (1995), Rincón-Zapatero et al. (1998), and Dockner and Wagener (2014), among others. Rincón-Zapatero et al. (1998) and Dockner and Wagener (2014), in particular, develop alternative solution methods that can be applied to derive symmetric Markov perfect Nash equilibria for games with a single-state variable and functional forms that can go beyond linear quadratic structures.

feedback strategies are given by $\underline{p}$ and $\overline{p}$, respectively, with

$$\underline{p} = \frac{1+2c}{3}, \overline{p} = \frac{2\left[c\left(2r+s\right)+r+s\right]+s}{2\left(3r+2s\right)+s}.$$

For s sufficiently high, both the stationary open-loop and the static Cournot equilibrium price lie in the interval of steady-state prices that can be supported by nonlinear feedback strategies, $[\underline{p}, \overline{p}]$, for some initial conditions. This implies that, once the assumption of linear strategies is relaxed, Markovian behavior can be either more or less competitive than open-loop and static behavior.[20] For all the prices between the steady-state price supported by the linear strategy, p_{∞}^{CL}, and the upper bound $\overline{p}$, steady-state profits are higher than those in Fershtman and Kamien (1987).

Joint profit maximization yields the following steady-state price,

$$p_{\infty}^{J} = \frac{\left(1+2c\right)\left(r+s\right)+2s}{2\left(r+2s\right)+r+s},$$

which is above $\overline{p}$. The fact that $p_{\infty}^{J} > \overline{p}$ implies that, for any $r > 0$, it is not possible to construct a stationary feedback equilibrium which supports the collusive stationary price, p_{∞}^{J}. However, as firms become infinitely patient, the upper bound $\overline{p}$ asymptotically approaches the collusive stationary price p_{∞}^{J}. Indeed,

$$\lim_{r\to 0} p_{\infty}^{J} = \lim_{r\to 0} \overline{p} = \frac{1+2\left(1+c\right)}{5}.$$

This is in line with the folk theorem, a well-known result in repeated games (see Friedman 1971 and Fudenberg and Maskin 1986): any individually rational payoff vector can be supported as a subgame perfect equilibrium of the dynamic game provided that the discount rate is sufficiently low. As argued in Dockner et al. (1993), who characterize the continuum of nonlinear feedback equilibria in a differential game between polluting countries, if the discount rate is sufficiently low, the use of nonlinear feedback strategies can be considered as a substitute for fully coordinating behavior. Firms may be able to reach a self-enforcing agreement which performs almost as good as fully coordinated policies.

4 Productive Assets

The dynamic game literature on productive assets can be partitioned into two subgroups, according to how players are considered: (i) as direct consumers of the asset; (ii) as firms using the asset to produce an output to be sold in the

[20] A similar result is obtained in Wirl (1996) considering nonlinear feedback strategies in a differential game between agents who voluntarily contribute to the provision of a public good.

marketplace.[21] In this section, we review recent contributions on (ii), focussing on renewable assets.[22]

Consider an n-firm oligopolistic industry where firms share access to a common-pool renewable asset over time $t \in [0, \infty)$. The rate of transformation of the asset into the output is one unit of output per unit of asset employed. Let $x(t)$ denote the asset stock available for harvesting at t. The dynamics of the asset stock is given by

$$\dot{x}(t) = h(x(t)) - \sum_{i=1}^{n} q_i(t), \ \ x(0) = x_0, \tag{18.23}$$

where

$$h(x) = \begin{cases} \delta x & \text{for } x \leq \overline{x}/2 \\ \delta(\overline{x} - x) & \text{for } x > \overline{x}/2. \end{cases}$$

The growth function $h(x)$, introduced in Benchekroun (2003a), and generalized in Benchekroun (2008), can be thought of as a "linearization" of the classical logistic growth curve.[23] $\delta > 0$ represents the implicit growth rate of the asset, $\overline{x}$ the maximum habitat carrying capacity, and $\delta\overline{x}/2$ the maximum sustainable yield.[24] Firms are assumed to sell their harvest in the same output market at a price $p = \max\{1 - Q, 0\}$, where $Q = \sum_{i=1}^{n} q_i$. Firm i's problem can be written as $(i, j = 1, 2, \ldots, n, j \neq i)$

$$\begin{cases} \max\limits_{q_i} \displaystyle\int_0^{\infty} e^{-rt} \pi_i \, dt \\ \text{s.t. } \dot{x} = h(x) - q_i - \sum_{j=1, j \neq i}^{n} q_j(x), \, x(0) = x_0 \end{cases}$$

---

[21] Papers belonging to (i) include Levhari and Mirman (1980), Clemhout and Wan (1985), Benhabib and Radner (1992), Dutta and Sundaram (1993a,b), Fisher and Mirman (1992, 1996), and Dockner and Sorger (1996). In these papers, agents' instantaneous payoffs do not depend on rivals' exploitation rates. The asset is solely used as a consumption good.

[22] Classical papers on oligopoly exploitation of nonrenewable resources are Lewis and Schmalensee (1980), Loury (1986), Reinganum and Stokey (1985), Karp (1992a,b), and Gaudet and Long (1994). For more recent contributions, see Benchekroun and Long (2006) and Benchekroun et al. (2009, 2010).

[23] The linearized logistic growth function has been used in several other oligopoly games, including Benchekroun et al. (2014), Benchekroun and Gaudet (2015), Benchekroun and Long (2016), and Colombo and Labrecciosa (2013a, 2015). Others have considered only the increasing part of the "tent", e.g., Benchekroun and Long (2002), Sorger (2005), Fujiwara (2008, 2011), Colombo and Labrecciosa (2013b), and Lambertini and Mantovani (2014). A nonlinear dynamics is considered in Jørgensen and Yeung (1996).

[24] Classical examples of $h(x)$ are fishery and forest stand dynamics. As to the former, with a small population and abundant food supply, the fish population is not limited by any habitat constraint. As the fish stock increases, limits on food supply and living space slow the rate of population growth, and beyond a certain threshold the growth of the population starts declining. As to the latter, the volume of a stand of trees increases at an increasing rate for very young trees. Then it slows and increases at a decreasing rate. Finally, when the trees are very old, they begin to have negative growth as they rot, decay, and become subject to disease and pests.

where $\pi_i = pq_i$ is Firm i's instantaneous profit, $r > 0$ is the common discount rate, and $q_j(x)$ is Firm j's feedback strategy. In order for equilibrium strategies to be well defined, and induce a trajectory of the asset stock that converges to admissible steady states, it is assumed that $\delta > \widehat{\delta}$, with $\widehat{\delta}$ depending on the parameters of the model. Firm i's HJB equation is

$$rV_i(x) = \max_{q_i} \left\{ \left[1 - q_i - \sum_{j=1, j\neq i}^{n} q_j(x) \right] q_i + \frac{\partial V_i}{\partial x} \left[h(x) - q_i - \sum_{j=1, j\neq i}^{n} q_j(x) \right] \right\}, \tag{18.24}$$

where $\partial V_i/\partial x$ can be interpreted as the shadow price of the asset for Firm i, which depends on x, r, δ, and n. A change in n, for instance, will have an impact on output strategies not only through the usual static channel but also through $\partial V_i/\partial x$. Maximization of the RHS of (18.24) yields Firm i's instantaneous best response (assuming inner solutions exist),

$$q_i = \frac{1}{2}\left(1 - \sum_{j=1, j\neq i}^{n} q_j - \frac{\partial V_i}{\partial x}\right),$$

which, exploiting symmetry, can be used to derive the equilibrium of the instantaneous game given x,

$$q^* = \frac{1}{1+n}\left(1 - \frac{\partial V}{\partial x}\right).$$

Note that when $\partial V/\partial x = 0$, q^* corresponds to the (per firm) static Cournot equilibrium output. For $\partial V/\partial x > 0$, any attempt to move along the reaction function triggers a shift in the reaction function itself, since $\partial V/\partial x$ is a function of x, and x changes as output changes. In Benchekroun (2003a, 2008), it is shown that equilibrium strategies are given by

$$q^* = \begin{cases} 0 & \text{for } 0 \leq x \leq x_1 \\ \alpha x + \beta & \text{for } x_1 < x \leq x_2 \\ q^C & \text{for } x_2 < x, \end{cases} \tag{18.25}$$

where α, β, x_1, and x_2 are constants that depend on the parameters of the model and q^C corresponds to the static Cournot equilibrium output. The following holds: $\alpha > 0$, $\beta < 0$, $\overline{x}/2 > x_2 > x_1 > 0$. For asset stocks below x_1, firms abstain from producing, the reason being that the asset is too valuable to be harvested, and firms are better off waiting for it to grow until reaching the maturity threshold, x_1. For asset stocks above x_2, the asset becomes too abundant to have any value, and firms behave as in the static Cournot game. For $x > x_2$, the static Cournot equilibrium can be sustained as a subgame perfect equilibrium, either temporarily or permanently, depending on the implicit growth rate of the asset and the initial asset stock. The fact that $\alpha > 0$ implies that there exists intertemporal strategic substitutability. Each firm has an incentive to increase output (leading to a decrease in the asset stock), so as to make the rival smaller in the future (given that equilibrium strategies are increasing in the asset stock). Benchekroun (2008) shows that there exists a range of asset stocks such that an increase in the number of firms, n, leads to an increase in both individual and industry output (in the short-run). This result is in contrast with traditional static Cournot analysis. In response to an increase in n, although equilibrium strategies become flatter, exploitation of the asset starts sooner. This implies that, when the asset stock is relatively scarce, new comparative statics results are obtained.

From (18.25), it follows that

$$\dot{x} = \begin{cases} \delta x \quad \text{for } 0 \leq x \leq x_1 \\ (\delta - n\alpha)\, x - n\beta \quad \text{for } x_1 < x \leq x_2 \\ \delta x - nq_C \quad \text{for } x_2 < x \leq \overline{x}/2 \\ \delta\,(\overline{x} - x) - nq^C \quad \text{for } \overline{x}/2 < x. \end{cases}$$

For δ sufficiently large, there exist three steady-state asset stocks, given by

$$x_\infty = \frac{n\beta}{\delta - n\alpha} \in (x_1, x_2)\,,\ \hat{x}_\infty = \frac{nq^C}{\delta} \in \left(x_2, \frac{\overline{x}}{2}\right),\ \tilde{x}_\infty = \overline{x} - \frac{nq^C}{\delta} \in \left(\frac{\overline{x}}{2}, \infty\right).$$

We can see immediately that $\widehat{x}_\infty$ is unstable and that $\tilde{x}_\infty$ is stable. Since $\delta - n\alpha < 0$, x_∞ is also stable. For $x_0 < \widehat{x}_\infty$, the system converges to x_∞; for $x_0 > \widehat{x}_\infty$, the system converges to $\tilde{x}_\infty$. For $x_0 < \overline{x}/2$, firms always underproduce compared with static Cournot; for $x_0 \in (x_2, \widehat{x}_\infty)$ firms start by playing the static equilibrium, and then when x reaches x_2, they switch to the nondegenerate feedback strategy; for $x_0 \in (\widehat{x}_\infty, \infty)$, firms always behave as in the static Cournot game. The static Cournot equilibrium can be sustained ad infinitum as a subgame perfect equilibrium of the dynamic game. For δ sufficiently small, there exists only one (globally asymptotically stable) steady state, x_∞. In this case, even starting with a very large stock, the static Cournot equilibrium can be sustained only in the short run. Irrespective of initial conditions, equilibrium strategies induce a trajectory of the asset stock that converges asymptotically to x_∞.

A question of interest is how an increase in the number of firms impacts on $p_\infty = 1 - \delta x_\infty$.[25] In contrast with static oligopoly theory, it turns out that p_∞ is increasing in n, implying that increased competition is detrimental to long-run welfare. This has important implications for mergers, modelled as an exogenous

[25]The impact of an increase in the number of firms on the steady-state equilibrium price is also analyzed in Colombo and Labrecciosa (2013a), who departs from Benchekroun (2008) by assuming that, instead of being common property, the asset is parcelled out (before exploitation begins). The qualitative results of the comparative statics results in Colombo and Labrecciosa (2013a) are in line with those in Benchekroun (2008).

change in the number of firms in the industry from n to $n - m + 1$, where m is the number of merging firms. Let $\Pi(m, n, x) = V(n - m + 1, x)$ be the equilibrium value of a merger, with $V(n - m + 1, x)$ denoting the discounted value of profits of the merged entity with $n - m + 1$ firms in the industry. Since each member of the merger receives $\Pi(m, n, x)/m$, for any subgame that starts at x, a merger is profitable if $\Pi(m, n, x)/m > \Pi(1, n, x)$. Benchekroun and Gaudet (2015) show that there exists an interval of initial asset stocks such that any merger is profitable. This holds true also for mergers otherwise unprofitable in the static model. Recall from static oligopoly theory that a merger involving $m < n$ firms in a linear Cournot game with constant marginal cost is profitable if $n < m + \sqrt{m} - 1$. Let $m = \alpha n$, where α is the share of firms that merge. We can see that a merger is profitable if $\alpha > 1 - (\sqrt{4n + 5} - 3)/(2n) \geq 0.8$. Hence, a merger involving less than 80% of the existing firms is never profitable (see Salant et al. 1983). In particular, a merger of two firms is never profitable, unless it results in a monopoly. In what follows, we consider an illustrative example of a merger involving two firms that is profitable in the dynamic game with a productive asset but unprofitable in the static game. We set $n = 3$ and $m = 2$ and assume that $x_0 \in (x_1, x_2)$ (so that firms play nondegenerate feedback strategies for all t). It follows that

$$\Pi(1, 3, x) = A\frac{x^2}{2} + Bx + C,$$

where $A = 8(r - 2\delta)/9$, $B = 5(2\delta - r)/(9\delta)$, and $C = [25r(r - 2\delta) + 9\delta^2]/(144r\delta^2)$, and

$$\Pi(2, 3, x) = \widehat{A}\frac{x^2}{2} + \widehat{B}x + \widehat{C},$$

where $\widehat{A} = 9(r - 2\delta)/8$, $\widehat{B} = 5(2\delta - r)/(8\delta)$, and $\widehat{C} = [25r(r - 2\delta) + 16\delta^2]/(144r\delta^2)$. A merger is then profitable if

$$\widetilde{A}\frac{x^2}{2} + \widetilde{B}x + \widetilde{C} > 0,$$

where $\widetilde{A} = 47(2\delta - r)/72$, $\widetilde{B} = 35(r - 2\delta)/(72\delta)$, and $\widetilde{C} = [25r(2\delta - r) - 2\delta^2]/(144r\delta^2)$. Setting $r = 0.1$ and $\delta = 11/9 > \widehat{\delta} = 10/9$, we can see that a merger involving two firms is profitable for $x \in (x_1, x_3)$, with $x_1 = 0.0754 < x_3 = 0.165\,8 < x_2 = 0.4545$.

Colombo and Labrecciosa (2015) extend the duopoly model in Benchekroun (2003a) to a differentiated duopoly and derive linear and nonlinear feedback equilibria under the assumption that firms compete either in quantities (à la Cournot) or in prices (à la Bertrand). The main objective of their analysis is to show that the traditional static result that the Bertrand equilibrium is more efficient than the Cournot equilibrium does not necessarily carry over to a Markovian environment. The inverse demand function for Firm i's product at t is given by $P_i(q_i(t), q_j(t)) =$

$\max\{0, 1 - q_i(t) - \gamma q_j(t)\}$, $i, j = 1, 2$, $j \neq i$, where $\gamma \in [0, 1)$ denotes the degree of product substitutability as perceived by the representative consumer: when $\gamma = 0$ demands for the two goods are independent, whereas when $\gamma \to 1$ the two goods are perceived as identical. For any $\gamma \in (0, 1)$ there exists imperfect product substitutability. In the context of a common-pool resource, imperfect product substitutability can be attributed to differences in firms' harvesting practices impacting on consumer preferences and/or differences in the way the resource is processed once harvested. The direct demand for Firm i's product at t is given by $D_i(p_i(t), p_j(t)) = \max\{0, 1/(1+\gamma) - p_i(t)/(1-\gamma^2) + \gamma p_j(t)/(1-\gamma^2)\}$. Colombo and Labrecciosa (2015) show that the Cournot equilibrium can be more efficient than the Bertrand equilibrium and can lead to a Pareto-superior outcome. In particular, there exists an interval of initial asset stocks such that the Cournot equilibrium dominates the Bertrand equilibrium in terms of short-run, stationary, and discounted consumer surplus (or welfare) and profits.

Focussing on the Bertrand competition case, Firm i's problem is

$$\begin{cases} \max\limits_{p_i} \displaystyle\int_0^{\infty} e^{-rt} \pi_i \, dt \\ \text{s.t. } \dot{x} = h(x) - D_i(p_i, p_j(x)) - D_j(p_i, p_j(x)), \, x(0) = x_0 \end{cases}$$

where $\pi_i = D_i(p_i, p_j(x))p_i$ is Firm i's instantaneous profit, $r > 0$ is the common discount rate, and $p_j(x)$ is Firm j's feedback strategy. Firm i's HJB equation is

$$rV_i(x) = \max_{p_i} \left\{ \left[1 - \frac{p_i}{1-\gamma} + \frac{\gamma p_j(x)}{1-\gamma} \right] \frac{p_i}{1+\gamma} + \frac{\partial V_i}{\partial x} \left[h(x) - \frac{2 - p_i - p_j(x)}{1+\gamma} \right] \right\}. \tag{18.26}$$

Maximization of the RHS of (18.26) yields Firm i's instantaneous best response (assuming inner solutions exist),

$$p_i = \frac{1-\gamma}{2} \left(1 + \frac{\partial V_i}{\partial x} + \frac{\gamma p_j}{1-\gamma} \right),$$

which, exploiting symmetry, can be used to derive the equilibrium of the instantaneous game given x,

$$p^* = \frac{1-\gamma}{2-\gamma} \left(1 + \frac{\partial V}{\partial x} \right).$$

Since $\partial V/\partial x \geq 0$ then the feedback Bertrand equilibrium is never more competitive than the static Bertrand equilibrium, given by $p^B = (1-\gamma)/(2-\gamma)$. This is in contrast with the result that Markovian behaviors are systematically more aggressive than static behaviors, which has been established in various classes of games, including capital accumulation games and games with production adjustment costs (e.g., Dockner 1992; Driskill 2001; Driskill and McCafferty 1989; Reynolds 1987, 1991).[26] In the case in which firms use linear feedback strategies, $\partial V/\partial x$ is nonincreasing in x. This has important strategic implications. Specifically, static strategic complementarity is turned into intertemporal strategic substitutability. A lower price by Firm i causes the asset stock to decrease thus inducing the rival to price less aggressively. When the implicit growth rate of the asset is sufficiently low, linear feedback strategies induce a trajectory of the asset stock that converges to a steady-state equilibrium price which, in the admissible parameter range, is above the steady-state equilibrium price of the corresponding Cournot game. This implies that the Cournot equilibrium turns out to be more efficient than the Bertrand equilibrium. The new efficiency result is also found when comparing the discounted sum of welfare in the two games. The intuitive explanation is that price-setting firms value the resource more than their quantity-setting counterparts and therefore tend to be more conservative in the exploitation of the asset. Put it differently, it is more harmful for firms to be resource constrained in Bertrand than in Cournot.

Proceeding as in Tsutsui and Mino (1990) and Dockner and Long (1993), Colombo and Labrecciosa (2015) show that there exists a continuum of asset stocks that can be supported as steady-state asset stocks by nonlinear feedback strategies. For $r > 0$, the stationary asset stock associated with the collusive price cannot be supported by any nonlinear feedback strategy. However, as $r \to 0$, in line with the folk theorem, firms are able to sustain the most efficient outcome. Interestingly, as argued in Dockner and Long (1993), the use of nonlinear strategies can be a substitute for fully coordinated behaviors. The static Bertrand (Cournot) equilibrium can never be sustained as a subgame perfect equilibrium of the dynamic Bertrand (Cournot) game. This is in contrast with the case in which firms use linear strategies. In this case, the equilibrium of the static game can be sustained ad infinitum, provided that the initial asset stock is sufficiently large and the asset stock grows sufficiently fast.

Colombo and Labrecciosa (2013b), analyzing a homogeneous product Cournot duopoly model in which the asset stocks are privately owned, show that the static Cournot equilibrium is the limit of the equilibrium output trajectory as time goes to infinity. This is true irrespective of initial conditions, the implication being that the static Cournot model can be considered as a reduced form model of a more complex dynamic model. The asset stocks evolve according to the following differential equation

$$\dot{x}_i(t) = \delta x_i(t) - q_i(t),\ x_i(0) = x_{i0}, \tag{18.27}$$

[26] A notable exception is represented by Jun and Vives (2004), who show that Bertrand competition with costly price adjustments leads to a steady-state price that is higher than the equilibrium price arising in the static game.

where $\delta > 0$ is the growth rate of the asset. Demand is given by $p(Q) = \max\{1 - Q, 0\}$. Firm i's problem can be written as $(i, j = 1, 2, j \neq i)$

$$\begin{cases} \max\limits_{q_i} \displaystyle\int_0^\infty e^{-rt}\pi_i dt \\ \text{s.t. } \dot{x}_i = \delta x_i - q_i, \dot{x}_j = \delta x_j - q_j(x_i, x_j), x_i(0) = x_{i0} \end{cases}$$

where $\pi_i = p(Q)q_i$ is Firm i's instantaneous profit, $r > 0$ is the discount rate, common to both firms, and $q_j(x_i, x_j)$ is Firm j's feedback strategy. Firm i's HJB equation is

$$rV_i(x_i, x_j) = \max_{q_i}\left\{\left[1 - q_i - q_j(x_i, x_j)\right]q_i + \frac{\partial V_i}{\partial x_i}(\delta x_i - q_i) + \frac{\partial V_i}{\partial x_j}\left[\delta x_j - q_j(x_i, x_j)\right]\right\}. \tag{18.28}$$

The feedback equilibrium trajectory of Firm i's quantity turns out to be

$$q_i(t) = \delta x_\infty + (2\delta - r)(x_{i0} - x_\infty)e^{-(\delta - r)t},$$

where x_∞ indicates the steady-state value of x_i. If $\delta > r$, then equilibrium strategies induce a trajectory of Firm i's output that converges asymptotically, for every possible initial conditions, to the static Cournot equilibrium output. Within the class of (stationary) linear feedback strategies, other strategies exist that stabilize the state vector for some initial conditions. Colombo and Labrecciosa (2013b) establish that there exists one which is more efficient (in that it is associated with higher stationary profits) than that converging to the static equilibrium, and that, as $r \to 0$, for some initial conditions, the feedback equilibrium price converges to a price which is above the static Cournot equilibrium price. This limiting result is in contrast with the limit game analyzed in Fershtman and Kamien (1987), in which feedback equilibrium strategies turn out to be less aggressive than static Cournot equilibrium strategies.

5 Research and Development

In this section, we review differential games of innovation.[27] We start by presenting the classical research and development (R&D) race model analyzed in Reinganum (1981, 1982), where innovation is modelled as a competitive process (race for technological breakthrough) among potential innovators that aim to be the first. Consider $n \geq 2$ identical firms competing with each other for the completion of a research project over a fixed finite time period $[0, T]$. Assume that the time of

[27] For discrete-time games of innovation, see Petit and Tolwinski (1996, 1999) and Breton et al. (2006).

the completion of the project by Firm i is a random variable τ_i with the probability distribution function $F_i(t) = \Pr\{\tau_i \leq t\}$. The completion times τ_i are mutually independent random variables. The player k with $\tau_k = \tau$ is called innovator, with $\tau = \min_{i=1,2,\ldots,n}\{\tau_i\}$. By stochastic independence, we have

$$F(t) = \Pr\{\tau \leq t\} = 1 - \prod_{i=1}^{n} [1 - F_i(t)].$$

Let $u_i(t) \geq 0$ denote the intensity of research efforts by Firm i. The hazard rate corresponding to F_i, i.e., the rate at which the discovery is made at a certain point in time by Firm i given that it has not been made before, is a linear function of $u_i(t)$,

$$h_i = \frac{\dot{F}_i}{1 - F_i} = \lambda u_i,$$

where $\lambda > 0$ measures the effectiveness of current R&D effort in making the discovery.[28] The probability of innovation depends on cumulative R&D efforts. Denote by $\overline{V}$ and by $\underline{V}$ the present value of the innovation to the innovator and the imitator, respectively, with $\overline{V} > \underline{V} \geq 0$. The idea is that the firm that makes the innovation first is awarded a patent of positive value $\overline{V}$, to be understood as the expected net present value of all future revenues from marketing the innovation net of any costs the firm incurs in doing so. If patent protection is perfect then $\underline{V} = 0$. $\overline{V}$ and $\underline{V}$ are constant, therefore independent of the instant of completion of the project. Assume a quadratic cost function of R&D, $C(u_i) = u_i^2/2$. Moreover, assume firms discount future payoffs at a rate equal to $r > 0$. The expected payoff of Firm i is given by

$$J_i = \int_0^T \left(\lambda \overline{V} u_i + \lambda \underline{V} \sum_{j \neq i}^{n} u_j - \frac{1}{2} u_i^2 e^{-rt}\right) \prod_{i=1}^{n} [1 - F_i]\, dt.$$

Assume for simplicity that $\underline{V} = 0$. Let z_i denote the accumulated research efforts (proxy for knowledge), i.e.,

$$\dot{z}_i(t) = u_i(t), \; z_i(0) = z_{i0},$$

and let $F_i(t) = 1 - \exp[-\lambda z_i]$, implying that Firm i's probability of success depends on the research efforts accumulated by Firm i by t. Given stochastic independence, we have

$$\prod_{i=1}^{n} [1 - F_i] = \exp[-\lambda Z],$$

[28] Choi (1991) and Malueg and Tsutsui (1997) assume that the hazard rate is uncertain, either zero (in which case the projet is unsolvable) or equal to $\lambda > 0$ (in which case the projet is solvable). The intensity of R&D activity is fixed in the former paper and variable in the latter. Chang and Wu (2006) consider a hazard rate that does not depend only on R&D expenditures but also on the accumulated production experiences, assumed to be proportional to cumulative output.

where $Z = \sum_{i=1}^{n} z_i$. The expected payoff of Firm i can then be written as a function of the research efforts accumulated by all the n firms by t,

$$J_i = \int_0^T \left(\lambda \overline{V} u_i - \frac{1}{2} u_i^2 e^{-rt} \right) \exp[-\lambda Z] dt.$$

Note that, although firms are assumed to observe the state vector $(z_1, z_2, \ldots, z_n)$, the only variable that is payoff relevant is Z. Call $y = \exp[-\lambda Z]$. It follows that $\dot{y} = -\lambda y U$, where $U = \sum_{i=1}^{n} u_i$. Firm i's problem can then be written as

$$\max_{u_i} J_i = \int_0^T \left(\lambda \overline{V} u_i - \frac{1}{2} u_i^2 e^{-rt} \right) y dt$$
$$s.t. \ \dot{y} = -\lambda y U.$$

Since the state variable, y, enters both the instantaneous payoff function and the equation of motion linearly, then, as is well known in the differential game literature, the open-loop equilibrium coincides with the feedback equilibrium, therefore, it is subgame perfect. The open-loop equilibrium strategy, derived in Reinganum (1982), is given by

$$u_i^* = \frac{2\lambda \overline{V} (n-1) e^{rt}}{(2n-1) - \exp[\lambda^2 (n-1) (e^{rt} - e^{rT}) \overline{V}/r]},$$

which is independent of firms' knowledge stocks.[29] The implication is that the leading firm (the one starting with a higher stock of knowledge) invests the same amount in R&D as the lagging firm. Hence, the distinction between leader and follower becomes irrelevant, and, as pointed out in Doraszelski (2003), there is no sense in which one can properly speak of one competitor being ahead of another or of the two competitors being neck and neck.

Doraszelski (2003) extends Reinganum (1982) by assuming that as a firm invests in R&D, its chances to immediately make the discovery increase and, in addition, the firm adds to its knowledge stock, which is subject to depreciation. Firm i's accumulated knowledge evolves according to

$$\dot{z}_i(t) = u_i(t) - \delta z_i(t), \quad u_i(0) = u_{i0}, \tag{18.29}$$

with $\delta \geq 0$ denoting the rate at which knowledge depreciates over time. Firm i's hazard rate of successful innovation is given by

$$h_i = \lambda u_i + \gamma z_i^{\psi},$$

[29]As shown in Mehlmann and Willing (1983) and Dockner et al. (1993), there exist also other equilibria depending on the state.

with $\lambda > 0$, $\gamma > 0$, $\psi > 0$. λ measures the effectiveness of current R&D effort in making the discovery and γ the effectiveness of past R&D efforts, with ψ determining the marginal impact of past R&D efforts.[30] The special case of an exponential distribution of success time ($\gamma = 0$) corresponds to the memoryless R&D race model analyzed in Reinganum (1981, 1982). Relaxing the assumption of an exponential distribution of success time allows to capture the fact that a firm's past experiences add to its current capability. The cost to acquire knowledge is given by $c(u_i) = u_i^\eta/\eta$, with $\eta > 1$, which generalizes the quadratic cost function considered in Reinganum (1981, 1982).

Firm i's HJB equation is

$$rV_i(z_i,z_j) = \max_{u_i} \left\{ h_i(u_i,z_i)\, G_i(z_i,z_j) + h_j(u_j,z_j) L_i(z_i,z_j) - c(u_i) + \frac{\partial V_i}{\partial z_i}\dot{z}_i + \frac{\partial V_i}{\partial z_j}\dot{z}_j \right\}, \tag{18.30}$$

where $G_i(z_i,z_j) = \overline{V} - V_i(z_i,z_j)$ is the capital gain from winning the race and $L_i(z_i,z_j) = -V_i(z_i,z_j)$ is the capital loss.

Performing the maximization indicated in (18.30) yields

$$u_i^* = \left[\lambda_i G_i(z_i,z_j) + \frac{\partial V_i}{\partial z_i} \right]^{\frac{1}{\eta-1}}.$$

Using u_i^* and focussing on a symmetric equilibrium (the only difference between firms is the initial stock of knowledge), we obtain the following system of nonlinear first-order PDEs ($i, j = 1, 2,\ j \neq i$)

$$\begin{aligned} rV_i = {} & \left\{ \lambda \left[\lambda\left(\overline{V} - V_i\right) + \frac{\partial V_i}{\partial z_i} \right]^{\frac{1}{\eta-1}} + \gamma z_i^\psi \right\} \left(\overline{V} - V_i\right) - \left\{ \lambda \left[\lambda\left(\overline{V} - V_j\right) + \frac{\partial V_j}{\partial z_j} \right]^{\frac{1}{\eta-1}} + \gamma z_j^\psi \right\} V_i \\ & - \frac{1}{\eta} \left[\lambda\left(\overline{V} - V_i\right) + \frac{\partial V_i}{\partial z_i} \right]^{\frac{\eta}{\eta-1}} + \frac{\partial V_i}{\partial z_i} \left\{ \left[\lambda\left(\overline{V} - V_i\right) + \frac{\partial V_i}{\partial z_i} \right]^{\frac{1}{\eta-1}} - \delta z_i \right\} \\ & + \frac{\partial V_i}{\partial z_j} \left\{ \left[\lambda\left(\overline{V} - V_j\right) + \frac{\partial V_j}{\partial z_j} \right]^{\frac{1}{\eta-1}} - \delta z_j \right\}, \end{aligned} \tag{18.31}$$

which, in general, cannot be solved analytically. Using numerical (approximation) methods, Doraszelski (2003) obtains a number of interesting results.[31] First, a firm has an incentive to reduce its R&D expenditures as its knowledge stock increases. In contrast with multistage models (e.g., Grossman and Shapiro 1987), in which the follower devotes less resources to R&D than the leader, the follower tries to catch

[30]Dawid et al. (2015) analyze the incentives for an incumbent firm to invest in risky R&D projects aimed to expand its own product range. They employ the same form of the hazard rate as in Doraszelski (2003), focussing on the case in which $\psi > 1$.

[31]A classical reference on numerical methods is Judd (1998).

up with the leader. This result holds true irrespective of the shape of the hazard rate, provided that the follower has sufficient knowledge itself. Also, in contrast to multistage race models, competition is not necessarily fiercest when firms are neck and neck. If the hazard rate is concave or linear, competition among firms is most intense when their knowledge stocks are of unequal size and least intense when they are of equal size, whereas this need not be the case if the hazard rate is convex. Firms can either respond aggressively or submissively to an increase in its rival's knowledge stock. A firm responds aggressively if it has a sufficiently large knowledge stock and submissively otherwise. Doraszelski (2003) also finds that the steady-state value of knowledge is increasing in the effectiveness of current R&D effort, λ, and the value of the patent, $\overline{V}$, and decreasing in the depreciation rate of knowledge, δ.

Another stream of literature dealing with differential games of innovation departs from the assumption that firms are engaged in R&D races by assuming that, at each point in time, all firms in the industry innovate, simultaneously and noncooperatively, and there are no unsuccessful firms.[32] The environment is typically deterministic. The cost of production decreases with R&D efforts, with the possibility of technological spillovers being taken into account. The first model that we consider is analyzed in Breton et al. (2004). They propose two differentiated oligopoly game models where, at each point in time $t \in [0, \infty)$, firms devote resources to cost-reducing R&D and compete in the output market either in prices (à la Bertrand) or in quantities (à la Cournot). In what follows, we focus on the case in which firms are price-setters. Firm i's demand is given by $D_i(p_i, p_j) = 1 - p_i + sp_j$, where p_i denotes the price set by Firm i for its product variety, and similarly for p_j, and s captures the degree of substitutability between the two varieties, $0 \leq s < 1$. Let z_i be Firm i's accumulated stock of knowledge, evolving over time as in (18.29). Firm i's production cost at t depends on the quantity produced by Firm i at t, the stock of knowledge accumulated by Firm i by t, and also on the stock of knowledge accumulated by Firm j by t $(i, j = 1, 2,\ j \neq i)$

$$C_i(p_i, p_j, z_i, z_j) = \left[c_i + \alpha D_i(p_i, p_j) - \psi(z_i + s\beta z_j)\right] D_i(p_i, p_j),$$

where $c_i < A$, $\alpha \geq 0$, $0 \leq \psi \leq 1$, and $0 \leq \beta \leq 1$.[33] For any given degree of knowledge spillover, the more substitute firms' products are, the more each firm benefits (from the other's accumulated stock of knowledge) in reducing its

[32] The assumption that all firms innovate is relaxed in Ben Abdelaziz et al. (2008) and Ben Brahim et al. (2016). In the former, it is assumed that not all firms in the industry pursue R&D activities. The presence of non-innovating firms (called surfers) leads to lower individual investments in R&D, a lower aggregate level of knowledge, and a higher product price. In the latter, it is shown that the presence of non-innovating firms may lead to higher welfare.

[33] A cost function with knowledge spillovers is also considered in the homogeneous product Cournot duopoly model analyzed in Colombo and Labrecciosa (2012) and in the differentiated Bertrand duopoly model analyzed in El Ouardighi et al. (2014), where it is assumed that the spillover parameter is independent of the degree of product differentiation. In both papers, costs are linear in the stock of knowledge. For a hyperbolic cost function, see Janssens and Zaccour (2014).

costs. In turn, this implies that, even if the degree of knowledge spillover is very high, if the firms' products are unrelated, then these firms will not benefit from each other's knowledge. Note that this formulation is more general than the one proposed in the seminal paper by D'Aspremont and Jacquemin (1988) and adopted in the subsequent multistage game literature on cost-reducing R&D (e.g., Qiu 1997; Symeonidis 2003). The cost associated with R&D investments is assumed to be quadratic, $F_i(u_i) = \theta_i u_i + \phi u_i^2/2$. Firm i's HJB equation is

$$rV_i(z_i, z_j) = \max_{u_i, p_i} \left\{ D_i(p_i, p_j(z_i, z_j))p_i - C_i(p_i, p_j(z_i, z_j), z_i, z_j) - F_i(u_i) + \frac{\partial V_i}{\partial z_i}(u_i - \delta z_i) + \frac{\partial V_i}{\partial z_j}\left(u_j(z_i, z_j) - \delta z_j\right) \right\}. \quad (18.32)$$

Performing the maximization indicated in (18.32) yields (assuming inner solutions exist)

$$u_i^* = \frac{1}{\phi}\left(\frac{\partial V_i}{\partial z_i} - \theta_i\right),$$

and

$$p_i^* = \frac{2\alpha(1+s) + c_i + sc_j - \psi\left[z_i + sz_j + s\beta\left(z_j + sz_i\right)\right]}{2\alpha(1-s^2)},$$

which can be used to obtain the following system of PDEs,

$$rV_i(z_i, z_j) = D_i(p_i^*, p_j^*)p_i^* - C_i(p_i^*, p_j^*, z_i, z_j) - F_i(u_i^*) + \frac{\partial V_i}{\partial z_i}\left(u_i^* - \delta z_i\right) + \frac{\partial V_i}{\partial z_j}\left(u_j^* - \delta z_j\right).$$

Consider quadratic value functions of the form,

$$V_i(z_i, z_j) = \kappa_1 z_i^2/2 + \kappa_2 z_j^2/2 + \kappa_3 z_i + \kappa_4 z_j + \kappa_5 z_i z_j + \kappa_6.$$

Equilibrium strategies can be written as

$$u_i^* = \frac{1}{\phi}\left(\kappa_3 - \theta_i + \kappa_1 z_i + \kappa_5 z_j\right),$$

with $\kappa_3 > \theta_i$, $\kappa_1 < \delta$, $|\kappa_1 - \delta| < |\kappa_5|$, and

$$p_i^* = \omega_1 + \omega_2 z_i + \omega_3 z_j,$$

with $\omega_1 > 0$ and $\omega_2, \omega_3 < 0$. Equilibrium prices and R&D investments turn out to be lower in Bertrand than in Cournot competition. Cournot competition becomes more efficient when R&D productivity is high, products are close substitutes, and R&D spillovers are not close to zero.

A different approach in modelling spillovers is taken in Colombo and Dawid (2014), where it is assumed that the unit cost of production of a firm is decreasing in its stock of accumulated knowledge and it is independent of the stocks of knowledge accumulated by its rivals. Unlike Breton et al. (2004), the evolution of the stock of knowledge of a firm depends on the stock of knowledge accumulated not only by that firm but also on the aggregate stock of knowledge accumulated by the other firms in the industry, i.e., there are knowledge spillovers. Colombo and Dawid (2014) analyze an n-firm homogeneous product Cournot oligopoly model where, at each point in time $t \in [0, \infty)$, each firm chooses its output level and the amount to invest in cost-reducing R&D. At $t = 0$, Firm $i = 1, 2, \ldots, n$ chooses also where to locate, whether in a cluster or not. Demand is given by $p = \max\{a - bQ, 0\}$, with $Q = \sum_{i=1}^{n} q_i$ denoting industry output, $a, b > 0$. Firm i's marginal cost depends on its stock of knowledge z_i in a linear way, i.e.,

$$c_i(t) = \max\{\overline{c} - \gamma z_i(t), 0\},$$

with $\overline{c}, \gamma > 0$. Knowledge accumulation depends on firms location: if Firm i locates in the cluster then its stock of knowledge is given by

$$\dot{z}_i(t) = u_i(t) + \beta \sum_{j=1}^{m-1} z_j(t) - \delta z_i(t), \tag{18.33}$$

where $m-1$ is the number of firms in the cluster except i, $\beta > 0$ captures the degree of knowledge spillovers in the cluster, and $\delta \geq 0$ the rate at which knowledge depreciates. The idea is that, by interacting with the other firms in the cluster, Firm i is able to add to its stock of knowledge even without investing in R&D. If Firm i does not locate in the cluster, then $\beta = 0$. Firms optimal location choices are determined by comparing firms' value functions for different location choices evaluated at the initial vector of states. R&D efforts are associated with quadratic costs, $F_i(u_i) = \eta_i u_i^2/2$, with $\eta_i > 0$. Colombo and Dawid (2014) make the assumption that $\eta_1 < \eta_i = \eta$, with $i = 2, 3, \ldots, n$, i.e., Firm 1 is the technological leader, in the sense that it is able to generate new knowledge at a lower cost than its competitors; all Firm 1's rivals have identical costs. The feedback equilibrium can be fully characterized by three functions: one for the technological leader; one for a technological laggard located in the cluster; and one for a technological laggard located outside the cluster. The main result of the analysis is that the optimal strategy of a firm is of the threshold type: the firm should locate in isolation only if its technological advantage relative to its competitors is sufficiently large.[34]

Kobayashi (2015) considers a symmetric duopoly ($\eta_1 = \eta_2 = \eta$) and abstracts from firms' location problem. Instead of assuming that spillovers depend on knowledge stocks, Kobayashi (2015), building on Cellini and Lambertini (2005, 2009), makes the alternative assumption that technological spillovers depend on firms' current R&D efforts. The relevant dynamics is given by[35]

$$\dot{z}_i(t) = u_i(t) + \beta u_j(t) - \delta z_i(t).$$

The rest of the model is as in Colombo and Dawid (2014). Kobayashi (2015) shows that the feedback equilibrium of the duopoly game can be characterized analytically. Attention is confined to a symmetric equilibrium ($z_1 = z_2 = z$ and $x_1 = x_2 = x$, with $x_i = u_i, q_i$). Firm i's HJB equation is ($i, j = 1, 2,\ j \neq i$)

$$rV_i(z_1, z_2) = \max_{u_i, q_i} \left\{ \{a - b\left[q_i + q_j(z_1, z_2)\right] - (\overline{c} - \gamma z_i)\} q_i - \eta \frac{u_i^2}{2} + \frac{\partial V_i}{\partial z_i}\left[u_i + \beta u_j(z_1, z_2) - \delta z_i\right] + \frac{\partial V_i}{\partial z_j}\left[u_j(z_1, z_2) + \beta u_i - \delta z_j\right] \right\}. \tag{18.34}$$

Equilibrium strategies are given by

$$q^* = \frac{a - \overline{c} + \gamma z}{3b},\ u^* = \frac{1}{\eta}\left[\left.\frac{\partial V_i}{\partial z_i}\right|_z + \beta \left.\frac{\partial V_i}{\partial z_j}\right|_z \right].$$

By comparing equilibrium strategies in the noncooperative case with those in the cooperative case, Kobayashi (2015) shows that cooperative R&D investments are larger than noncooperative investments for all possible values of spillovers.[36] Moreover, as r approaches infinity, the steady-state open-loop equilibrium converges to the subgame perfect equilibrium of the two-stage game analyzed in D'Aspremont and Jacquemin (1988).

[34] Colombo and Dawid (2014) also consider the case in which all firms have the same R&D cost parameter η, but there exists one firm which, at $t = 0$, has a larger stock of knowledge than all the other firms.

[35] The case in which knowledge is a public good ($\beta = 1$) is considered in Vencatachellum (1998). In this paper, the cost function depends both on current R&D efforts and accumulated knowledge, and firms are assumed to be price-taking.

[36] The literature on R&D cooperation is vast. Influential theoretical (static) papers include D'Aspremont and Jacquemin (1988), Choi (1993), and Goyal and Joshi (2003). Dynamic games of R&D competition vs cooperation in continuous time include Cellini and Lambertini (2009) and Dawid et al. (2013). For a discrete-time analysis, see Petit and Tolwinski (1996, 1999).

Cellini and Lambertini (2005) make the alternative assumption that the state variable is the vector of firms' unit costs, rather than the vector of firms' stocks of knowledge. They analyze a homogeneous product Cournot oligopoly with $n \geq 2$ firms. At each point in time $t \in [0, \infty)$, Firm $i = 1, 2, \ldots, n$ chooses its output level and the amount to invest in cost-reducing R&D, u_i. R&D costs are symmetric and quadratic. The dynamics of the unit cost of production of Firm i is given by

$$\dot{c}_i(t) = c_i(t)\left[\delta - u_i(t) - \beta \sum_{j \neq i}^{n} u_j(t)\right],$$

where $\delta \geq 0$ measures the instantaneous decrease in productive efficiency due to technological obsolescence and $\beta \geq 0$ is the spillover parameter. They focus on the characterization of an open-loop equilibrium, which can be justified by assuming that Firm i cannot observe the production costs of its rivals and therefore cannot condition its actions on the current realization of rivals' costs. The main result of the analysis is that the aggregate R&D effort is monotonically increasing in the number of firms. Consequently, more competition in the market turns out to be beneficial to innovation.[37]

Most of the literature on innovation focusses on process innovation. Prominent exceptions are Dawid et al. (2013, 2015), who analyze a differentiated duopoly game where firms invest in R&D aimed at the development of a new differentiated product; Dawid et al. (2009), who analyze a Cournot duopoly game where each firm can invest in cost-reducing R&D for an existing product and one of the two firms can also develop a horizontally differentiated new product; and Dawid et al. (2010), who study the incentives for a firm in a Cournot duopoly to lunch a new product that is both horizontally and vertically differentiated. Leaving aside the product proliferation problem, Cellini and Lambertini (2002) consider the case in which firms invest in R&D aimed at decreasing the degree of product substitutability between $n \geq 2$ existing products. Firm i's demand is given by $p_i = A - Bq_i - D\sum_{j \neq i}^{n} q_j$, where D denotes the degree of product substitutability between any pair of varieties. At each point in time $t \in [0, \infty)$, Firm i chooses its output level and the amount to invest in R&D, u_i. Production entails a constant marginal cost, c. The cost for R&D investments is also linear, $F_i(u_i) = u_i$. The degree of product substitutability is assumed to evolve over time according to the following differential equation

[37]Setting $n = 2$, Cellini and Lambertini (2009) compare private and social incentives toward cooperation in R&D, showing that R&D cooperation is preferable to noncooperative behavior from both a private and a social point of view. On R&D cooperation in differential games see also Navas and Kort (2007), Cellini and Lambertini (2002, 2009), and Dawid et al. (2013). On R&D cooperation in multistage games, see D'Aspremont and Jacquemin (1988), Kamien et al. (1992), Salant and Shaffer (1998), Kamien and Zang (2000), Ben Youssef et al. (2013).

$$\dot{D}(t) = -D(t) \sum_i^n u_i(t) \left[1 + \sum_i^n u_i(t) \right]^{-1}, D(0) = B.$$

Since $D(0) = B$, at the beginning of the game, firms produce the same homogeneous product. As $\sum_i^n u_i(t)$ tends to infinity, D approaches zero, meaning that products tend to become unrelated. The main result of the analysis is that an increase in the number of firms leads to a higher aggregate R&D effort and a higher degree of product differentiation. More intense competition favors innovation, which is also found in Cellini and Lambertini (2005).

6 Strategic Investments Under Uncertainty: A Real Options Approach

In this section, we review recent contributions on investment timing in oligopolistic markets whose developments are uncertain, for instance, because of uncertain demand growth. Since Dixit and Pindyk (1994), a number of papers have studied how market competition affects firms' investment decisions, thus extending the traditional single investor framework of real option models to a strategic environment where the profitability of each firm's project is affected by other firms' decision to invest.[38] These papers extend the classical literature on timing games originating from the seminal paper by Fudenberg and Tirole (1985), to shed new light on preemptive investments, strategic deterrence, dissipation of first-mover advantages, and investment patterns under irreversibility of investments and demand uncertainty.[39]

Before considering a strategic environment, we outline the standard real option investment model (see Dixit and Pindyk 1994).[40] A generic firm (a monopolist) seeks to determine the optimal timing of an irreversible investment I, knowing that the value of the investment project follows a geometric Brownian motion

$$dx(t) = \alpha x(t)dt + \sigma x(t)dw, \tag{18.35}$$

[38]Studies of investment timing and capacity determination in monopoly include Dangl (1999) and Decamps et al. (2006). For surveys on strategic real option models where competition between firms is taken into account, see Chevalier-Roignant et al. (2011), Azevedo and Paxson (2014), and Huberts et al. (2015).

[39]The idea that an incumbent has an incentive to hold excess capacity to deter entry dates back to Spence (1977, 1979).

[40]Note that the real options approach represents a fundamental departure from the rest of this survey. Indeed, the dynamic programming problems considered in this section are of the optimal-stopping time. This implies that investments go in one lump, causing a discontinuity in the corresponding stock, instead of the more incremental control behavior considered in the previous sections.

where α and σ are constants corresponding to the instantaneous drift and the instantaneous standard deviation, respectively, and dw is the standard Wiener increment. α and σ can be interpreted as the industry growth rate and the industry volatility, respectively. The riskless interest rate is given by $r > \alpha$ (for a discussion on the consequences of relaxing this assumption, see Dixit and Pindyk 1994). The threshold value of x at which the investment is made maximizes the value of the firm, V. The instantaneous profit is given by $\pi = xD_0$ if the firm has not invested and $\pi = xD_1$ if the firm has invested, with $D_1 > D_0$. The investment cost is given by δI, $\delta > 0$. In the region of x such that the firm has not invested, the Bellman equation is given by

$$rV(x) = xD_0 dt + E[dV(x)].$$

From standard real option analysis, it follows that

$$V(x) = \frac{xD_0}{r-\alpha} + Ax^{\beta} + Bx^{\lambda},$$

where A, B, β, λ are constants. The indifference level x^* can be computed by employing the value-matching and smooth-pasting conditions,

$$V(x^*) = \frac{x^* D_1}{r-\alpha} - \delta I,$$

$$\left.\frac{\partial V(x)}{\partial x}\right|_{x^*} = \frac{D_1}{r-\alpha},$$

and the boundary condition, $V(0) = 0$. These conditions yield the optimal investment threshold,

$$x^* = \frac{(r-\alpha)\,\beta\delta I}{(D_1 - D_0)\,(\beta - 1)},$$

where

$$\beta = \frac{1}{2} - \frac{\alpha}{\sigma^2} + \sqrt{\left(\frac{\alpha}{\sigma^2} - \frac{1}{2}\right)^2 + \frac{2r}{\sigma^2}} > 1. \tag{18.36}$$

Given the assumptions, x^* is strictly positive. Hence, the value of the firm is given by

$$V(x) = \begin{cases} \dfrac{xD_0}{r-\alpha} + \left[\dfrac{x\,(D_1 - D_0)}{r-\alpha} - \delta I\right]\left(\dfrac{x}{x^*}\right)^{\beta} & \text{for } x \leq x^* \\ \dfrac{xD_1}{r-\alpha} - \delta I \quad \text{for } x > x^*, & \end{cases}$$

where the term x/x^* can be interpreted as a stochastic discount factor. An increase in the volatility parameter σ, by decreasing β, leads to an increase in x^* (since x^* is decreasing in β), thus delaying the investment. The investment is also delayed when α increases and when, not surprisingly, firm becomes more patient (r decreases).

The above framework is applied by Pawlina and Kort (2006) to an asymmetric duopoly game. The instantaneous payoff of Firm $i = A, B$ can be expressed as $\pi_i(x) = xD_{00}$ if neither firm has invested, $\pi_i(x) = xD_{11}$ if both firms have invested, $\pi_i(x) = xD_{10}$ if only Firm i has invested, and $\pi_i(x) = xD_{01}$ if only Firm j has invested, with $D_{10} > D_{00} > D_{01}$ and $D_{10} > D_{11} > D_{01}$. Assume Firm A is the low-cost firm. Its investment cost is normalized to I, whereas the investment cost of Firm B is given by δI, with $\delta > 1$. It is assumed that $x(0)$ is sufficiently low to rule out the case in which it is optimal for the low-cost firm to invest at $t = 0$ (see Thijssen et al. 2012). As in the standard real option investment model, I is irreversible and exogenously given. Three types of equilibria can potentially arise in the game under consideration. First, a preemptive equilibrium. This equilibrium occurs when both firms have an incentive to become the leader, i.e., when the cost disadvantage of Firm B is relatively small. In this equilibrium, unlike in Fudenberg and Tirole (1985), there is no rent equalization between the leader and the follower: Firm A has always an incentive to take the lead, even in the presence of the slightest degree of cost asymmetry. Second, a sequential equilibrium. This equilibrium occurs when Firm B has no incentive to lead. Firm A behaves as a monopolist that is not threatened by future competition (even if Firm B investments will affect Firm A's payoff). Third, a simultaneous equilibrium. This equilibrium occurs when both firms invest at the same point in time. Pawlina and Kort (2006) show that the characterization of the equilibrium crucially depends on the first-mover advantage, defined as the ratio D_{10}/D_{11}, which can result from improved product characteristics or greater cost efficiency, and the investment cost asymmetry, δ. Specifically, there exists a threshold of δ, δ^*, depending on the first-mover advantage, the interest rate, and the parameters of the stochastic process, separating the regions of the preemptive and the sequential equilibrium. For all $\delta < \delta^*$, Firm A needs to take into account possible preemption by Firm B, whereas $\delta \geq \delta^*$ implies that firms always invest sequentially at their optimal thresholds. Moreover, there exists another threshold of δ, δ^{**}, depending on the first-mover advantage, the interest rate, and the parameters of the stochastic process, such that the resulting equilibrium is of the joint investment type for all $\delta < \delta^{**}$, and of the sequential/preemptive investment type for $\delta \geq \delta^{**}$. When $\delta > \max(\delta^*, \delta^{**})$, the firms invest sequentially and Firm A can act as a sole holder of the investment opportunity. Furthermore, a set of parameter values exists for which $\delta^{**} = 1$, meaning that simultaneous investment is never optimal.

Boyer et al. (2004, 2012) consider continuous-time duopoly models where symmetric firms add capacity in lump sums. The former paper assumes Bertrand competition, whereas the latter deals with Cournot competition. These papers build on the literature on strategic capacity investments, which includes the seminal contributions of Gilbert and Harris (1984) and Fudenberg and Tirole (1985), and the more recent works by Besanko and Doraszelski (2004), Genc et al. (2007), and

Besanko et al. (2010), and the literature on real option games (e.g., Boyer et al. 2004; Gutiérrez and Ruiz-Aliseda 2011; Huisman and Kort 2004, 2015; Pawlina and Kort 2006; Ruiz-Aliseda 2012; Weeds 2002).[41] In what follows, we get a closer look at Boyer et al. (2004). The market is described by a price inelastic unit demand

$$D(p(t)) = \begin{cases} 0 & \text{for } p > x \\ [0,1] & \text{for } p = x \\ 1 & \text{for } p < x, \end{cases}$$

where the total willingness to pay x for the commodity produced by the firms is subject to aggregate demand shocks described by a geometric Brownian motion, as in (18.35). Firm $i = 1, 2$ is risk-neutral and discount future revenues and costs at a constant risk-free rate, $r > \alpha$. Investment is irreversible and takes place in a lumpy way. Each unit of capacity allows a firm to cover at most a fraction $1/N$ of the market for some positive integer N. The cost of each unit of capacity is constant and equal to $I > 0$. Capacity does not depreciate. Within $[t, t + dt)$, the timing of the game is as follows: (i) firstly, each firm chooses how many units of capacity to invest in, given the realization of x and the existing levels of capacity; (ii) next, each firm quotes a price given its new level of capacity and that of its rival; (iii) lastly, consumers choose from which firm to purchase, and production and transfers take place. Following Boyer et al. (2004), we briefly consider two benchmarks: the optimal investment of a monopolist and the investment game when $N = 1$. The expected discounted value of a monopolist investing N units of capacity is given by

$$V(x) = \max_T E\left[\int_{t=T}^{\infty} \exp[-rt]x(t)dt - NI\exp[-rt] \,\middle|\, x(0) = x\right], \tag{18.37}$$

where T is the time at which the investment is undertaken. An optimal stopping time for (18.37) is to invest when the state of demand reaches the threshold

$$x^* = \frac{\beta NI\,(r-\alpha)}{\beta - 1},$$

where β is given in (18.36). x^* is above the threshold at which the value of the firm is nil, $\hat{x} = (r - \alpha)NI$.

In the investment game with single investment ($N = 1$), when Firm j has invested, Firm i's expected discounted value is nil; otherwise, it is given by

$$V_i(x) = E\left[\int_{t=0}^{\infty} \exp[-rt]x(t)dt - I \,\middle|\, x(0) = x\right] = \frac{x}{r-\alpha} - I.$$

[41] Genc et al. (2007), in particular, use the concept of S-adapted equilibrium of Haurie and Zaccour (2005) to study different types of investment games.

When x reaches $\hat{x}$, firms become indifferent between preempting and staying out of the market. For $x < \hat{x}$, it is a dominant strategy for both firms not to invest. There exists another threshold, $\tilde{x}$, corresponding to the monopolistic trigger, above which firms have no incentives to delay investments. By the logic of undercutting, for any $x \in (\hat{x}, \tilde{x})$, each firm wants to preempt to avoid being preempted in the future. As a consequence, when x reaches $\hat{x}$, a firm will invest, and the other firm will remain inactive forever. For both firms, the resulting expected payoff at $t = 0$ is nil. This finding is in line with the classical rent dissipation phenomenon, described, for instance, in Fudenberg and Tirole (1985). However, in the case of multiple investments ($N = 2$), Boyer et al. (2004) show that, in contrast with standard rent dissipation results, no dissipation of rents occurs in equilibrium, despite instantaneous price competition. Depending on the importance of the real option effect, different patterns of equilibria may arise. If the average growth rate of the market is close to the risk-free rate, or if the volatility of demand changes is high, then the unique equilibrium acquisition process involves joint investment at the socially optimal date. Otherwise, the equilibrium investment timing is suboptimal, and the firms' long-run capacities depend on the initial market conditions.

Huisman and Kort (2015) combine investment timing and capacity determination in a Cournot duopoly model where demand is linear and subject to stochastic shocks. The stochastic shocks admit a geometric Brownian motion process, as in the standard real option investment model. The problem of each firm consists in determining the timing of investment and the capacity to install at the time of investing. As explained in Huisman and Kort (2015), in a setting with uncertainty and competition, an interesting trade-off arises: while uncertainty generates a value of waiting with investment, the presence of competition gives firms the incentive to preempt their competitor. One would expect the preemption effect to dominate in a moderately uncertain environment. Closely related papers are Dangl (1999), who analyzes the joint determination of the timing and the size of the investment but abstracts from competition, and Yang and Zhou (2007), who consider competition, but take the incumbent decision as a given.

First, we consider the monopoly setting. The market price at t is given by

$$p(t) = x(t)[1 - Q(t)],$$

where x follows a geometric Brownian motion, as in (18.35), and Q is the industry output. The firm is risk neutral. It becomes active by investing in capacity Q at a cost δQ, $\delta > 0$. It is assumed that the firm always operates at full capacity. The interest rate is denoted by $r > \alpha$. Let $V(x)$ denote the value of the firm. The investment problem of the monopolist is given by

$$V(x) = \max_{T,Q} E\left[\int_{t=T}^{\infty} x(t)(1 - Q)Qe^{-rt}dt - \delta Q e^{-rT} \middle| x(0) = x \right].$$

Maximization w.r.t. Q gives

$$Q(x) = \frac{1}{2}\left[1 - \frac{\delta(r-\alpha)}{x}\right].$$

Solving the corresponding value-matching and smooth-pasting conditions gives x^* and Q^*,

$$x^* = \frac{\delta(\beta+1)(r-\alpha)}{\beta-1},\ Q^* = \frac{1}{\beta+1},$$

where β is given in (18.36). A comparative statics analysis reveals that increased uncertainty delays investment but increases the size of the investment. The investment timing is socially desirable, in the sense that it corresponds to the investment timing a benevolent social planner would choose. However, the monopolist chooses to install a capacity level that is half the socially optimal level.

Next, we consider the case in which there is competition. In this case, $Q = q_1 + q_2$. The investment cost of Firm $i = 1, 2$ is $\delta_i q_i$, with $\delta_1 < \delta_2$ (Firm 1 is the low-cost firm). When Firm 1 has a significant cost advantage, Firm 2's decision involves no strategic aspects. The optimal investment decisions of Firm 2 are characterized by

$$x^*(q_1) = \frac{\delta_2(\beta+1)(r-\alpha)}{(\beta-1)(1-q_1)},\ q_2^*(q_1) = \frac{1-q_1}{\beta+1}.$$

Taking into account the optimal investment decisions of Firm 2, Firm 1 can either use an entry deterrence strategy, in which case Firm 1 will act as a monopolist as long as x is low enough, or an entry accommodation strategy, in which case Firm 1 will invest just before Firm 2. In this case, since Firm 1 is the first to invest, and it is committed to operate at full capacity, Firm 1 will become the Stackelberg leader. From $x^*(q_1)$, entry by Firm 2 will be deterred when $q_1 > \hat{q}_1(x)$, with

$$\hat{q}_1(x) = 1 - \frac{\delta_2(\beta+1)(r-\alpha)}{(\beta-1)x}.$$

When instead $q_1 \leq \hat{q}_1(x)$, entry by Firm 2 will be accommodated. Huisman and Kort (2015) show that entry can only be temporarily deterred, the reason being that at one point in time the market will have grown enough to make it optimal for Firm 2 to enter. Firm 1 will overinvest not only to induce Firm 2 to invest less but also to induce Firm 2 to delay investment. With moderate uncertainty, the preemption effect prevails, implying that Firm 1 invests early in a small capacity, with Firm 2 becoming the larger firm. Once investment cost asymmetry is removed, Firm 1 invests relatively late in a larger capacity than Firm 2. In the end, after both firms have installed capacity, Firm 1 will become the larger firm in the industry.

The above papers on strategic investments under uncertainty assume that investments are irreversible. In real options models (with or without competition) in which firms decide when entering (becoming active in) a market, the possibility of exit is typically not contemplated. A few exceptions are Murto (2004), considering firms' exit in a declining duopoly; Dockner and Siyahhan (2015), where a firm can abandon an R&D project and leave the market; and Ruiz-Aliseda (2016), where firms may decide to exit a market that expands up to a random maturity date and declines thereafter. Firms' exit decisions are also considered in the seminal paper by Ericson and Pakes (1995), who provide a discrete-time framework for numerically analyzing dynamic interactions in imperfectly competitive industries. Such a framework has been applied in a variety of oligopolistic settings. We refer the interested reader to the survey in Doraszelski and Pakes (2007).

7 Concluding Remarks

In this chapter, we have provided an overview of differential games applied to industrial organization. Needless to say, although we have tried to cover as much as possible, we do not claim to have been exhaustive. Many differential games, which we have not included in this survey, address problems that are related to industrial organization but follow different approaches to the study of dynamic competition. Two examples are differential games in advertising and vertical channels, in which the focus is upon the study of optimal planning of marketing efforts, rather than market behavior of firms and consumers. For excellent references on differential games in marketing, we refer the interested reader to Jørgensen and Zaccour (2004, 2014).

Much work has been done in the field of applications of differential games to industrial organization. However, it is fair to say that a lot still needs to be done, especially in regard to uncertainty and industry dynamics. Indeed, mainly for analytical tractability, the vast majority of contributions abstract from uncertainty and focus on symmetric (linear) Markov equilibria. If one is willing to abandon analytical tractability, then it becomes possible to apply the tool box of numerical methods to the analysis of more complex dynamic (either deterministic or stochastic) models, which can deal with asymmetries, nonlinearities, and multiplicity of states and controls.

References

Azevedo A, Paxson D (2014) Developing real options game models. Eur J Oper Res 237:909–920
Başar T, Olsder GJ (1995) Dynamic noncooperative game theory. Academic Press, San Diego
Ben Abdelaziz F, Ben Brahim M, Zaccour G (2008) R&D equilibrium strategies with surfers. J Optim Theory Appl 136:1–13
Ben Brahim M, Ben Abdelaziz F, Zaccour G (2016) Strategic investments in R&D and efficiency in the presence of free riders. RAIRO Oper Res 50:611–625
Ben Youssef S, Breton M, Zaccour G (2013) Cooperating and non-cooperating firms in inventive and absorptive research. J Optim Theory Appl 157:229–251
Benchekroun H (2003a) Unilateral production restrictions in a dynamic duopoly. J Econ Theory 111:214–239

Benchekroun H (2003b) The closed-loop effect and the profitability of horizontal mergers. Can J Econ 36:546–565
Benchekroun H (2008) Comparative dynamics in a productive asset oligopoly. J Econ Theory 138:237–261
Benchekroun H, Gaudet G (2015) On the effects of mergers on equilibrium outcomes in a common property renewable asset oligopoly. J Econ Dyn Control 52:209–223
Benchekroun H, Gaudet G, Lohoues H (2014) Some effects of asymmetries in a common pool natural resource oligopoly. Strateg Behav Environ 4:213–235
Benchekroun H, Halsema A, Withagen C (2009) On nonrenewable resource oligopolies: the asymmetric case. J Econ Dyn Control 33:1867–1879
Benchekroun H, Halsema A, Withagen C (2010) When additional resource stocks reduce welfare. J Environ Econ Manag 59:109–114
Benchekroun H, Long NV (2002) Transboundary fishery: a differential game model. Economica 69:207–221
Benchekroun H, Long NV (2006) The curse of windfall gains in a non-renewable resource oligopoly. Aust Econ Pap 4:99–105
Benchekroun H, Long NV (2016) Status concern and the exploitation of common pool renewable resources. Ecol Econ 125:70–82
Benhabib J, Radner R (1992) The joint exploitation of a productive asset: a game theoretic approach. Econ Theory 2:155–190
Besanko D, Doraszelski (2004) Capacity dynamics and endogenous asymmetries in firm size. RAND J Econ 35:23–49
Besanko D, Doraszelski U, Lu LX, Satterthwaite M (2010) On the role of demand and strategic uncertainty in capacity investment and disinvestment dynamics. Int J Ind Organ 28:383–389
Boyer M, Lasserre P, Mariotti T, Moreaux M (2004) Preemption and rent dissipation under price competition. Int J Ind Organ 22:309–328
Boyer M, Lasserre P, Moreaux M (2012) A dynamic duopoly investment game without commitment under uncertain market expansion. Int J Ind Organ 30:663–681
Breton M, Turki A, Zaccour G (2004) Dynamic model of R&D, spillovers, and efficiency of Bertrand and Cournot equilibria. J Optim Theory Appl 123:1–25
Breton M, Vencatachellum D, Zaccour G (2006) Dynamic R&D with strategic behavior. Comput Oper Res 33:426–437
Bulow J, Geanakoplos J, Klemperer P (1985) Multimarket oligopoly: strategic substitutes and complements. J Polit Econ 93:488–511
Cabral L (2012) Oligopoly dynamics. Int J Ind Organ 30:278–282
Cellini R, Lambertini L (1998) A dynamic model of differentiated oligopoly with capital accumulation. J Econ Theory 83:145–155
Cellini R, Lambertini L (2002) A differential game approach to investment in product differentiation. J Econ Dyn Control 27:51–62
Cellini R, Lambertini L (2004) Dynamic oligopoly with sticky prices: closed-loop, feedback and open-loop solutions. J Dyn Control Syst 10:303–314
Cellini R, Lambertini L (2005) R&D incentives and market structure: a dynamic analysis. J Optim Theory Appl 126:85–96
Cellini R, Lambertini L (2007a) Capital accumulation, mergers, and the Ramsey golden rule. In: Quincampoix M, Vincent T, Jørgensen S (eds) Advances in dynamic game theory and applications. Annals of the international society of dynamic games, vol 8. Birkhäuser, Boston
Cellini R, Lambertini L (2007b) A differential oligopoly game with differentiated goods and sticky prices. Eur J Oper Res 176:1131–1144
Cellini R, Lambertini L (2009) Dynamic R&D with spillovers: competition vs cooperation. J Econ Dyn Control 33:568–582
Chang SC, Wu HM (2006) Production experiences and market structure in R&D competition. J Econ Dyn Control 30:163–183
Chevalier-Roignant B, Flath CM, Huchzermeier A, Trigeorgis L (2011) Strategic investment under uncertainty: a synthesis. Eur J Oper Res 215:639–650

Choi JP (1991) Dynamic R&D competition under ‘hazard rate’ uncertainty. RAND J Econ 22: 596–610

Choi JP (1993) Dynamic R&D competition, research line diversity, and intellectual property rights. J Econ Manag Strateg 2:277–297

Clemhout S, Wan H Jr (1985) Dynamic common property resources and environmental problems. J Optim Theory Appl 46:471–481

Colombo L, Dawid H (2014) Strategic location choice under dynamic oligopolistic competition and spillovers. J Econ Dyn Control 48:288–307

Colombo L, Labrecciosa P (2012) Inter-firm knowledge diffusion, market power, and welfare. J Evol Econ 22:1009–1027

Colombo L, Labrecciosa P (2013a) Oligopoly exploitation of a private property productive asset. J Econ Dyn Control 37:838–853

Colombo L, Labrecciosa P (2013b) On the convergence to the Cournot equilibrium in a productive asset oligopoly. J Math Econ 49:441–445

Colombo L, Labrecciosa P (2015) On the Markovian efficiency of Bertrand and Cournot equilibria. J Econ Theory 155:332–358

D’Aspremont C, Jacquemin A (1988) Cooperative and noncooperative R&D in duopoly with spillovers. Am Econ Rev 78:1133–1137

Dangl T (1999) Investment and capacity choice under uncertain demand. Eur J Oper Res 117: 415–428

Dawid H, Kopel M, Dangl T (2009) Trash it or sell it? A strategic analysis of the market introduction of product innovations. Int Game Theory Rev 11:321–345

Dawid H, Kopel M, Kort PM (2010) Innovation threats and strategic responses in oligopoly markets. J Econ Behav Organ 75:203–222

Dawid H, Kopel M, Kort PM (2013) R&D competition versus R&D cooperation in oligopolistic markets with evolving structure. Int J Ind Organ 31:527–537

Dawid H, Keoula MY, Kopel M, Kort PM (2015) Product innovation incentives by an incumbent firm: a dynamic analysis. J Econ Behav Organ 117:411–438

Decamps JP, Mariotti T, Villeneuve S (2006) Irreversible investments in alternative projects. Econ Theory 28:425–448

Dixit AK, Pindyk R (1994) Investment under uncertainty. Princeton University Press, Princeton

Dockner EJ (1988) On the relation between dynamic oligopolistic competition and long-run competitive equilibrium. Eur J Polit Econ 4:47–64

Dockner EJ (1992) A dynamic theory of conjectural variations. J Ind Econ 40: 377–395

Dockner EJ, Feichtinger G, Mehlmann A (1993) Dynamic R&D competition with memory. J Evol Econ 3:145–152

Dockner EJ, Gaunersdorfer A (2001) On the profitability of horizontal mergers in industries with dynamic competition. Jpn World Econ 13:195–216

Dockner EJ, Gaunersdorfer A (2002) Dynamic oligopolistic competition and quasi-competitive behavior. In: Zaccour G (ed) Optimal control and differential games – essays in honor of Steffen Jørgensen. Kluwer Academic Publishers, Boston/Dordrecht, pp 107–119

Dockner E, Haug A (1990) Tariffs and quotas under dynamic duopolistic competition. J Int Econ 29:147–159

Dockner E, Haug A (1991) The closed loop motive for voluntary export restraints. Can J Econ 3:679–685

Dockner EJ, Jørgensen S, Long NV, Sorger G (2000) Differential games in economics and management science. Cambridge University Press, Cambridge

Dockner EJ, Long NV (1993) International pollution control: cooperative versus non-cooperative strategies. J Environ Econ Manag 24:13–29

Dockner EJ, Mosburger G (2007) Capital accumulation, asset values and imperfect product market competition. J Differ Equ Appl 13:197–215

Dockner EJ, Siyahhan B (2015) Value and risk dynamics over the innovation cycle. J Econ Dyn Control 61:1–16

Dockner EJ, Sorger G (1996) Existence and properties of equilibria for a dynamic game on productive assets. J Econ Theory 171:201–227

Dockner EJ, Wagener FOO (2014) Markov perfect Nash equilibria in models with a single capital stock. Econ Theory 56:585–625

Doraszelski U (2003) An R&D race with knowledge accumulation. RAND J Econ 34:20–42

Doraszelski U, Pakes A (2007) A framework for applied dynamic analysis in IO. In: Armstrong M, Porter R (eds) Handbook of industrial organization, Elsevier, vol 3, pp 1887–1966

Driskill R (2001) Durable goods oligopoly. Int J Ind Organ 19:391–413

Driskill R, McCafferty S (1989) Dynamic duopoly with adjustment costs: a differential game approach. J Econ Theory 49:324–338

Driskill R., McCafferty S (1996) Industrial policy and duopolistic trade with dynamic demand. Rev Ind Organ 11:355–373

Dutta PK, Sundaram RK (1993a) The tragedy of the commons. Econ Theory 3:413–426

Dutta PK, Sundaram RK (1993b) How different can strategic models be? J Econ Theory 60:42–61

El Ouardighi F, Shnaiderman M, Pasin F (2014) Research and development with stock-dependent spillovers and price competition in a duopoly. J Optim Theory Appl 161:626–647

Ericson R, Pakes A (1995) Markov-perfect industry dynamics: a framework for empirical work. Rev Econ Stud 62:53–82

Fershtman C, Kamien M (1987) Dynamic duopolistic competition with sticky prices. Econometrica 55:1151–1164

Fershtman C, Kamien M (1990) Turnpike properties in a finite-horizon differential game: dynamic duopoly with sticky prices. Int Econ Rev 31:49–60

Figuières C (2009) Markov interactions in a class of dynamic games. Theor Decis 66:39–68

Fisher R, Mirman LJ (1992) Strategic dynamic interactions: fish wars. J Econ Dyn Control 16:267–287

Fisher R, Mirman LJ (1996) The complete fish wars: biological and dynamic interactions. J Environ Econ Manag 30:34–42

Friedman JW (1971) A non-cooperative equilibrium for supergames. Rev Econ Stud 28:1–12

Fudenberg D, Tirole J (1985) Preemption and rent equalization in the adoption of new technology. Rev Econ Stud 52:383–401

Fudenberg D, Maskin E (1986) The folk theorem in repeated games with discounting or with incomplete information. Econometrica 54:533–554

Fujiwara K (2008) Duopoly can be more anti-competitive than monopoly. Econ Lett 101: 217–219

Fujiwara K (2009) Gains from trade in a differential game model of asymmetric oligopoly. Rev Int Econ 17:1066–1073

Fujiwara K (2011) Losses from competition in a dynamic game model of a renewable resource oligopoly. Resour Energy Econ 33:1–11

Gaudet G, Long NV (1994) On the effects of the distribution of initial endowments in a non-renewable resource duopoly. J Econ Dyn Control 18:1189–1198

Genc T, Reynolds SS, Sen S (2007) Dynamic oligopolistic games under uncertainty: a stochastic programming approach. J Econ Dyn Control 31:55–80

Gilbert R, Harris R (1984) Competition with lumpy investment. RAND J Econ 15:197–212

Goyal S, Joshi S (2003) Networks of collaboration in oligopoly. Games Econ Behav 43:57–85

Grossman GM, Shapiro C (1987) Dynamic R&D competition. Econ J 97:372–387

Gutiérrez O, Ruiz-Aliseda F (2011) Real options with unknown-date events. Ann Finance 7:171–198

Hamermesh D, Pfann G (1996) Adjustment costs in factor demand. J Econ Lit 34:1264–1292

Haurie A, Krawczyk JB, Zaccour G (2012) Games and dynamic games. World Scientific, Singapore

Haurie A, Zaccour G (2005) S-adapted equilibria in games played over event trees: an overview. Ann Int Soc Dyn Games 7:367–400

Huberts NFD., Huisman KJM, Kort PM, Lavrutich MN (2015) Capacity choice in (strategic) real options models: a survey. Dyn Games Appl 5:424–439

Huisman K, Kort P (2004) Strategic technology adoption taking into account future technological improvements: a real options approach. Eur J Oper Res 159:705–728

Huisman K, Kort P (2015) Strategic capacity investment under uncertainty. RAND J Econ 46: 376–408

Janssens G, Zaccour G (2014) Strategic price subsidies for new technologies. Automatica 50: 1999–2006

Jørgensen S, Yeung DWK (1996) Stochastic differential game model of common property fishery. J Optim Theory Appl 90:381–403

Jørgensen S, Zaccour G (2004) Differential games in marketing. Dordrecht, Kluwer

Jørgensen S, Zaccour G (2014) A survey of game-theoretic models of cooperative advertising. Eur J Oper Res 237:1–14

Judd K (1998) Numerical methods in economics. MIT Press, Cambridge

Jun B, Vives X (2004) Strategic incentives in dynamic duopoly. J Econ Theory 116:249–281

Kamien MI, Muller E, Zang I (1992) Cooperative joint ventures and R&D cartels. Am Econ Rev 82:1293–1306

Kamien MI, Zang I (2000) Meet me half way: research joint ventures and absorptive capacity. Int J Ind Organ 18:995–1012

Karp L (1992a) Efficiency inducing tax for a common property oligopoly. Econ J 102:321–332

Karp L (1992b) Social welfare in a common property oligopoly. Int Econ Rev 33:353–372

Karp L, Perloff J (1989) Dynamic oligopoly in the rice export market. Rev Econ Stat 71:462–470

Karp L, Perloff J (1993a) Open-loop and feedback models of dynamic oligopoly. Int J Ind Organ 11:369–389

Karp L, Perloff J (1993b) A dynamic model of oligopoly in the coffee export market. Am J Agric Econ 75:448–457

Kobayashi S (2015) On a dynamic model of cooperative and noncooperative R&D oligopoly with spillovers. Dyn Games Appl 5:599–619

Lambertini L, Mantovani A (2014) Feedback equilibria in a dynamic renewable resource oligopoly: pre-emption, voracity and exhaustion. J Econ Dyn Control 47:115–122

Levhari D, Mirman L (1980) The great fish war: an example using a dynamic Cournot-Nash solution. Bell J Econ 11:322–334

Lewis T, Schmalensee R (1980) Cartel and oligopoly pricing of non-replenishable natural resources. In: Liu P (ed) Dynamic optimization and application to economics. Plenum Press, New York

Long NV (2010) A survey of dynamic games in economics. World Scientific, Singapore

Long NV (2015) Dynamic games between firms and infinitely lived consumers: a review of the literature. Dyn Games Appl 5:467–492

Loury G (1986) A theory of 'oil'igopoly: Cournot equilibrium in exhaustible resource market with fixed supplies. Int Econ Rev 27:285–301

Malueg DA, Tsutsui SO (1997) Dynamic R&D competition with learning. RAND J Econ 28:751–772

Maskin E, Tirole J (2001) Markov perfect equilibrium I. Observable actions. J Econ Theory 100:191–219

Mehlmann A, Willing R (1983) On non-unique closed-loop Nash equilibria for a class of differential games with a unique and degenerated feedback solution. J Optim Theory Appl 41:463–472

Murto P (2004) Exit in duopoly under uncertainty. RAND J Econ 35:111–127

Navas J, Kort PM (2007) Time to complete and research joint ventures: a differential game approach. J Econ Dyn Control 31:1672–1696

Okuguchi K (1973) Quasi-competitiveness and Cournot oligopoly. Rev Econ Stud 40:145–148

Pawlina G, Kort P (2006) Real options in an asymmetric duopoly: who benefits from your competitive disadvantage? J Econ Manag Strateg 15:1–35

Petit ML, Tolwinski B (1996) Technology sharing cartels and industrial structure. Int J Ind Organ 15:77–101

Petit ML, Tolwinski B (1999) R&D cooperation or competition? Eur Econ Rev 43:185–208

Qiu L (1997) On the dynamic efficiency of Bertrand and Cournot equilibria. J Econ Theory 75:213–229

Ramsey FP (1928) A mathematical theory of saving. Econ J 38:543–559

Reinganum J (1981) Dynamic games of innovation. J Econ Theory 25:21–41

Reinganum J (1982) A dynamic game of R&D: patent protection and competitive behavior. Econometrica 50:671–688

Reinganum J, Stokey N (1985) Oligopoly extraction of a common property resource: the importance of the period of commitment in dynamic games. Int Econ Rev 26:161–173

Reynolds S (1987) Capacity investment, preemption and commitment in an infinite horizon model. Int Econ Rev 28:69–88

Reynolds S (1991) Dynamic oligopoly with capacity adjustment costs. J Econ Dyn Control 15:491–514

Rincón-Zapatero J, Martinez J, Martín-Herrán G (1998) New method to characterize subgame perfect Nash equilibria in differential games. J Optim Theory Appl 96:377–395

Roos CF (1925) A mathematical theory of competition. Am J Mathem 46:163–175

Roos CF (1927) A Dynamical theory of economics. J Polit Econ 35:632–656

Ruffin RJ (1971) Cournot oligopoly and competitive behaviour. Rev Econ Stud 38:493–502

Ruiz-Aliseda F, Wu J (2012) Irreversible investments in stochastically cyclical markets. J Econ Manag Strateg 21:801–847

Ruiz-Aliseda F (2016) Preemptive investments under uncertainty, credibility, and first mover advantages. Int J Ind Organ 44:123–137

Salant SW, Shaffer G (1998) Optimal asymmetric strategies in research joint ventures. Int J Ind Organ 16:195–208

Salant SW, Switzer S, Reynolds R (1983) Losses from horizontal merger: the effects of an exogenous change in industry structure on Cournot-Nash equilibrium. Q J Econ 98:185–213

Simaan M, Takayama T (1978) Game theory applied to dynamic duopoly problems with production constraints. Automatica 14:161–166

Simon LK, Stinchcombe MB (1989) Extensive form games in continuous time: pure strategies. Econometrica 57:1171–1214

Singh N, Vives X (1984) Price and quantity competition in a differentiated duopoly. RAND J Econ 15:546–554

Symeonidis G (2003) Comparing Cournot and Bertrand equilibria in a differentiated duopoly with product R&D. Int J Ind Organ 21:39–55

Sorger G (2005) A dynamic common property resource problem with amenity value and extraction costs. Int J Econ Theory 1:3–19

Spence M (1977) Entry, capacity, investment and oligopolistic pricing. Bell J Econ 8:534–544

Spence M (1979) Investment strategy and growth in a new market. Bell J Econ 10:1–19

Thijssen J, Huisman K, Kort PM (2012) Symmetric equilibrium strategies in game theoretic real option models. J Math Econ 48:219–225

Tsutsui S (1996) Capacity constraints and voluntary output cutback in dynamic Cournot competition. J Econ Dyn Control 20:1683–1708

Tsutsui S, Mino K (1990) Nonlinear strategies in dynamic duopolistic competition with sticky prices. J Econ Theory 52:136–161

Vencatachellum D (1998) A differential R&D game: implications for knowledge-based growth models. J Optim Theory Appl 96:175–189

Vives X (1999) Oligopoly pricing: old ideas and new tools, 1st edn, vol 1. MIT Press Books/The MIT Press, Cambridge

Weeds H (2002) Strategic delay in a real options model of R&D competition. Rev Econ Stud 69:729–747

Wirl F (2010) Dynamic demand and noncompetitive intertemporal output adjustments. Int J Ind Organ 28:220–229

Wirl F, Dockner EJ (1995) Leviathan governments and carbon taxes: costs and potential benefits. Eur Econ Rev 39:1215–1236

Wirl F (1996) Dynamic voluntary provision of public goods: extension to nonlinear strategies. Eur J Polit Econ 12:555–560

Wiszniewska-Matyszkiel A, Bodnar M, Mirota F (2015) Dynamic oligopoly with sticky prices: off-steady-state analysis. Dyn Games Appl 5:568–598

Yang M, Zhou Q (2007) Real options analysis for efficiency of entry deterrence with excess capacity. Syst Eng Theory Pract (27):63–70

Dynamic Games in Finance

19

Michèle Breton

Contents

Abstract

Finance is a discipline that encompasses all the essential ingredients of dynamic games, through the involvement of investors, managers, and financial intermediaries as players who have competing interests and who interact strategically over time. This chapter presents various applications of dynamic game models used in the broad area of finance, with the objective of illustrating the scope of possibilities in this field. Both corporate and investment finance applications

M. Breton (✉)
HEC Montréal, Montréal, QC, Canada
e-mail: michele.breton@hec.ca

T. Başar, G. Zaccour (eds.), *Handbook of Dynamic Game Theory*,
https://doi.org/10.1007/978-3-319-44374-4_23

are presented. Topics covered include game options and their use as financial instruments, bankruptcy games and their association with the valuation of debt and equity, and dynamic game models used to explain empirical observations about the financial decisions made by firms, for instance, on capital structure, dividend payments, and investment choices. In each case, the presentation highlights the game's various ingredients, the choice of the equilibrium concept, and the solution approach used. The chapter's focus is on the contributions made by dynamic game models to financial theory and practice.

Keywords
Game theory · Finance · Game options · Bankruptcy · Corporate finance · Dynamic games

1 Introduction

Finance in general is concerned with the dynamics of economies, where savings are reinvested in the stock of capital, to be used as a production factor. Finance is usually divided into two main fields: investment finance, which relates to investors' decisions, and corporate finance, which deals with firms' decisions. A third area of interest is financial intermediation or disintermediation, where the focus is on the proper functioning of the transmission channel between investors and firms.

Consequently, finance is a field that has all the essential ingredients of dynamic games. Players are investors, firms, and financial intermediaries. Investors lend funds to firms in anticipation of future benefits. Firms borrow from investors in order to finance their activities, increase their value, and pay dividends. Funds are transmitted from lenders to borrowers through markets or financial intermediaries, such as banks. These players have competing interests and interact strategically over time. Moreover, time and risk play crucial parts in the way financial decisions are evaluated by these various players.

Finance game-theoretic models originated with the seminal papers of Leland and Pyle (1977), Ross (1977), and Bhattacharya (1979), in response to observed phenomena that could not be explained by classical models relying on assumptions of the homogeneity of players' beliefs and information. These early papers assume information asymmetries and strategic interactions and develop signalling models to explain corporate decisions such as capital structure and the payment of dividends. However, even if the Ross (1977) model has two periods and the Bhattacharya (1979) model is extended to a multiperiod setting, these early game-theoretic models represent special cases of dynamic games, since they require that players pre-commit to their strategies and since the players' beliefs do not adapt as information is revealed.

Surveys of game theory in finance, covering the first generation of game-theoretic models, not necessarily dynamic, can be found in Thakor (1991) and in Allen and Morris (2014). Thakor (1991) offers an in-depth discussion of the importance of the sequence of moves in signalling games and reviews selected applications in corporate finance and financial intermediation. Allen and Morris (2014) offers a review of some of the important issues in asset pricing and corporate finance that have been addressed using game theory and an analysis of the role of information and beliefs in game-theoretic models.

This chapter presents various applications of dynamic game models in the area of finance in order to illustrate the scope of possibilities in this area. Models are organized according to application subfields.

Section 2 is mainly concerned with investment finance and introduces a class of option-bearing derivative securities involving multiple holders with interacting rights. The valuation of *game options* corresponds to the solution of a dynamic game between the multiple holders. An important class of financial instruments involves game options, namely, debt instruments such as bonds and warrants. The valuation of game options is therefore also related to the valuation of corporate debt.

Section 3 discusses *bankruptcy games* as they relate to financial distress and its resolution. These games involve the various claimants to a firm's assets, and the solution indicates how these assets are distributed among them in the case of bankruptcy. Bankruptcy games are of interest in both investment and corporate finance, as their solution can be used to assess the value of debt and equity and to determine the optimal debt structure for firms.

Finally, Sect. 4 pertains to corporate finance; it deals with financial decisions made by firms, such as the choice between debt and equity when financing operations, the amount of dividends paid out to shareholders, and decisions about whether or not to invest in risky projects. These *corporate games* usually involve managers or entrepreneurs interacting strategically with investors or financial institutions (equity or debt holders). As in the first generation of finance game-theoretic models, corporate games are mainly used to explain or justify generally observed patterns in management decisions and their impact on claimholders' responses.

2 Game Options

A *contingent claim*, or derivative security, is a financial instrument whose value depends on the value of some basic underlying assets. Among derivative securities, options are contingent claims giving their holder an optional right, called exercise. Options are characterized by their maturity, payoff function, and exercise schedule: European options can be exercised only at the maturity date; Bermudan options can be exercised at a finite number of predefined dates; and American options can be exercised at any time during their life.

The main issues involved in analyzing an optional contingent claim are, first, the purpose it serves as a financial instrument and, second, the determination of its value. A contingent claim can be replicated by a portfolio of primitive assets and can be evaluated using no-arbitrage arguments: the value of the security is equal to the minimum capital required to set up a self-financing portfolio that covers the payoffs of the claim for any exercise strategy. In complete market models, the value of the contingent claim is also equal to the discounted expected value of the claim's future cash flows, under a unique martingale measure (the so-called "risk-neutral" probability measure).

Game options appear when a contingent claim gives interacting optional rights to more than one holder, that is, when the exercise of an optional right by one holder modifies the options (e.g., their payoff function or exercise schedule) of the others. Game options were introduced in Kifer (2000), where a contingent claim, called an Israeli option, is analyzed. This contract contains two optional clauses that can be exercised at any time up to a maturity date, by either of two players: Player 1 has the right to buy (call) or to sell (put) an underlying asset from Player 2 for a contractual price, while Player 2 has the right to cancel the contract and indemnify Player 1 by paying a penalty that depends on the price of the underlying asset.

2.1 A General Game Option Model

Consider a contract with an inception date $t = 0$ and maturity T, where $\mathcal{T} \subseteq (0, T]$ is the set of dates where exercise is allowed. Let $(X_t)_{0 \le t \le T}$ denote the stochastic process describing the price of the underlying asset and $(G_t^i)_{t \in \mathcal{T}} \ge 0$ denote payoff processes adapted to the filtration generated by X_t, defined for $i \in \{1, 2\}$. An Israeli option is a contract with stopping features that are introduced through a pair of stopping times (τ_1, τ_2) with respect to the filtration generated by the process $(X_t)_{0 \le t \le T}$, so that Player 1 selects $\tau_1 \in \mathcal{T}$ and Player 2 selects $\tau_2 \in \mathcal{T}$, triggering a stopping event at date $\tau_1 \wedge \tau_2 = \min\{\tau_1, \tau_2\}$ and leading to an immediate cash flow received by Player 1 from Player 2, where

$$R(\tau_1, \tau_2) \equiv G_{\tau_1}^1 \mathbb{I}_{\tau_1 \le \tau_2} + G_{\tau_2}^2 \mathbb{I}_{\tau_2 < \tau_1}$$

is the discounted value of this cash flow at $t = 0$, and $\mathbb{I}_A$ denotes the indicator function of event A. The payoff processes G_t^i thus represent the (discounted) amount paid by the seller to the buyer if Player i exercises her option first, at date t, conditional to the observed prices of the underlying asset up to t. The exercise of her option by either of the two players terminates the contract and therefore cancels the other player's option.

In the context of complete markets, Kifer (2000) shows, using a replicating portfolio argument, that the fair price of an Israeli option is equal to the value V of a zero-sum optimal stopping game of two players (Dynkin 1969),

$$V = \sup_{\tau_1 \in \mathcal{T}} \inf_{\tau_2 \in \mathcal{T}} \mathbb{E}_{\tau_1 \wedge \tau_2} [R(\tau_1, \tau_2)] = \inf_{\tau_2 \in \mathcal{T}} \sup_{\tau_1 \in \mathcal{T}} \mathbb{E}_{\tau_1 \wedge \tau_2} [R(\tau_1, \tau_2)],$$

and that there exists a unique hedging strategy for this option, where the notation $\mathbb{E}_t[\cdot]$ represents the expectation, conditional on the information observed up to t.

Kühn (2004) studies the value of game options in incomplete markets. The model assumes that each player chooses both a trading strategy and a stopping time in order to maximize the utility of her terminal wealth, where the trading strategy indicates the number of shares of the primitive assets held in the portfolio at time $t \in [0, T]$. The valuation of a game option contract can then be interpreted as determining the prices under which neither buyer nor seller can profit from trading the claim and is then equivalent to the equilibrium solution of a nonzero-sum Dynkin game. Kühn (2004) shows the existence of an equilibrium for this game in a Markovian context, when players use discrete stopping times and have exponential utility functions (i.e., constant risk aversion). Hamadène and Zhang (2010) generalizes this setting and shows the existence of a Nash equilibrium in continuous time for general stochastic processes and arbitrary utility functions.

2.2 Bonds and Embedded Options

The general setting of game options, as defined above, does not address whether such instruments exist in financial markets or what their purpose is. There is however an important class of financial instruments that can be modeled as game options: bonds with interacting embedded options. *Bonds* are debt instruments requiring the issuer to repay to the lender the amount borrowed (principal) plus interest (coupons) over a specified period of time, until maturity, at which time the principal is due. Bonds are characterized as fixed-income instruments, since the coupons and principal are known deterministic cash flows from the issuer to the lender, assuming that the issuer does not default prior to the maturity date. The most common type of embedded option is the call provision in *callable* or *redeemable* bonds, which gives the issuer the right to retire the debt before the maturity date. Bonds can also include a put provision (*putable* bonds), which allows the bondholder to sell the security back to the issuer before maturity, and/or a conversion provision (*convertible* or *exchangeable* bonds), which gives the bondholder the right to exchange the bond for a specified number of shares of another security. The call feature allows the bond issuer to replace the debt by a lower-interest one if interest rates on the market decline. The put feature allows the bondholder to invest in higher-interest bonds if market interest rates increase, while the conversion feature allows her to take advantage of movements in the price of the associated security.

Game options appear when bonds contain provisions available to both the issuer and the lender. In such cases, the exercise of an embedded option by one of the players terminates the contract, thereby eliminating both the remaining future cash flows and the embedded option owned by the other player. No-arbitrage approaches to evaluating bonds with embedded options were proposed in the literature, prior to their characterization as Dynkin games. For instance, Brennan and Schwartz (1980) evaluates a callable convertible bond, and McConnell and Schwartz (1986) looks at a callable, putable, convertible bond. The following characterization can be found in Brennan and Schwartz (1980, p. 907): "The equilibrium value of a convertible bond is defined as that value which offers the potential of arbitrage profit neither to purchaser nor to short seller, given that the bondholder pursues an optimal strategy with respect to conversion and that the firm pursues an optimal policy with respect to calling the bonds." This characterization clearly relates the value of a (callable) convertible bond to the equilibrium value of a two-player game.

2.3 Defaultable Game Options

In the Brennan and Schwartz (1980) model, the stochastic process X_t is two-dimensional, so the value of the bond depends both on the value of the issuing firm and on the interest rate; moreover, the model allows for the possibility of default by the bond issuer. The value of the firm affects the conversion value of the bond and the probability of default, while the interest rate affects the discounted value of future cash flows. It is interesting to point out that, contrary to the game option model in Kifer (2000), the evolution of the underlying stochastic process X_t is not independent of the solution of the game, since the value of the firm includes the value of its convertible bonds. Assuming that default happens when the firm value hits a boundary that depends on the debt principal, and that conversion by all bondholders is simultaneous, the authors characterize the players' optimal (equilibrium) strategies of the players. The value of the bond is governed by a differential equation with boundary conditions and can be solved numerically.

In fact, considering the possibility that one of the parties defaults on her obligation, that is, that she fails to provide her contractual payoff to the other party, amounts to adding an additional stopping time to the game option model. In the corporate finance literature, the possibility of default is typically characterized by two ingredients: the way the default event is triggered and the recovery process describing the payoffs in case of default. Three main avenues are used to characterize the default stopping time. The first consists of assuming that the default event is triggered by the underlying assets' price process, for instance, when the value of a firm's assets falls below the value of its liabilities (structural models)—as in Brennan and Schwartz (1980). A second avenue considers default to be governed by an exogenous process (intensity-based or reduced-form models). A third possibility is to assume that the default is decided by one of the strategic players (optimal default).

The value of a defaultable convertible bond is analyzed in Sîrbu and Shreve (2006) in the structural model setting, albeit with a single source of risk, where the bondholder maximizes the value of her claim, while the issuer minimizes the same. The interest rate is assumed constant, and default occurs when the value of the firm falls to zero. Accordingly, the decisions to call or to convert the bond are driven by the firm's value (the bondholder has an incentive to convert if she expects her ensuing stake in the firm to be more valuable than the coupon and principal cash flows), while the firm's value depends on the players' strategies. Sîrbu and Shreve (2006) shows that, depending on the relative values of the coupon rate, interest rate, and call price, the Dynkin game that characterizes the bond price reduces to an optimal stopping problem and a fixed point problem, admitting a unique solution.

A reduced-form model is considered in Bielecki et al. (2008) using a general market model, which is arbitrage-free but possibly incomplete, and allowing for uncertainty in the discount rate. The paper characterizes the arbitrage price of game options with dividends in terms of the solution of Dynkin games, and the results are extended to defaultable game options by considering that bankruptcy occurs at an exogenously given random moment. The specific case of defaultable convertible bonds is analyzed, allowing for commonly included features such as call and put provisions, a call notice period, and call protection and assuming that either conversion or recovery (according to some recovery process) is allowed in case of default.

In Chen et al. (2013), an optimal default model, which is where the firm's stockholders decide on the timing of bankruptcy, is used to characterize the equilibrium strategies of the holders of callable convertible bonds and of the issuing firm's stockholders. The model also accounts for tax benefits and liquidation costs, which makes the game between the two types of players nonzero-sum. *Tax benefits* refer to the fact that interest (coupon) payments are tax deductible, while *liquidation costs* refer to the losses that are incurred upon default when a firm's assets are liquidated in order to reimburse bondholders. The underlying asset is the market value of equity, and the discount process is assumed constant. Using a nonzero-sum stochastic game framework, the authors obtain the existence and uniqueness of the Nash equilibrium between the two players and characterize their equilibrium strategies. They argue that credit risk and tax benefits may provide an explanation to empirically observed call and conversion strategies.

2.4 Warrants and Dilutive Claims

Warrants are contingent claims that are similar to options in that they give their holder an optional right to buy a certain security, at a given price, according to some exercise schedule, up to their maturity. The key difference between warrants and options is that the exercise payoff of a warrant consists of a newly issued security. The exercise of a warrant therefore dilutes the equityholders' stakes and reduces their share of the issuing firm's future dividends. Accordingly, a warrant can be interpreted as a game option, in the sense that the exercise of a warrant by one holder modifies the expected payoff of all the other warrantholders that have not yet exercised their claim, along with that of the equityholders of the issuing firm. This strategic interaction is pointed out in Emanuel (1983, pp. 211–212), which states that "…when multiple warrants are outstanding, the exercise of some of the warrants leads to the creation of new shares and to concomitant changes in the dividend policy and capital structure of the firm. The value accruing to one warrantholder is therefore not independent of what the other warrantholders do."

The dilution effect is also extant in other convertible securities, namely, convertible bonds which were presented in Sect. 2.2. However, the models presented in this section use one of the two following simplifying assumptions: there is a single bond, or all conversion options are exercised at the same time. A similar approach is taken in Yagi and Sawaki (2010) to evaluate a callable warrant. When all warrants are exercised at the same time, the exercise payoff of a warrantholder is the conversion value divided by the number of shares after exercise (the initial number of shares plus the number of warrants issued by the firm).

However, Emanuel (1983) shows that it is optimal for a single warrantholder holding multiple claims to exercise her warrants sequentially, bringing into question the assumption of a single equilibrium conversion stopping time for American warrants and convertible bonds. Constantinides (1984) and Constantinides and Rosenthal (1984) propose a theory of warrants held by competitive warrantholders (i.e., who do not collude to determine their exercise strategy) who are not constrained to exercise their warrants as a single block. The model is a noncooperative dynamic game in discrete time between the warrantholders, who maximize their expected payoffs against known strategies of the firm (e.g., call, coupon, dividend, investment, and stock issuance policies are exogenous). The players use feedback strategies that depend on the observation of the vector of underlying asset prices, on the current number of shares of common stock, and on the number of outstanding warrants. Constantinides and Rosenthal (1984) demonstrates the existence of at least one competitive equilibrium, under the assumption that players are atomic, that is, that they are not aware of the impact of their decisions on the state vector, which depends on the joint decisions of all players. Constantinides (1984) shows that there exists an equilibrium (sequential) strategy that results in the same price as under the block strategy and that this price is the highest one when multiple equilibria exist.

Koziol (2006) analyzes an extension of the warrant game model where firms are assumed to issue both bonds and warrants and compares the competitive equilibrium to the optimal block exercise strategy. The model accounts for the possibility of default at bond maturity if the firm value is not sufficient to cover the principal. The author finds significant differences between the exercise strategies and warrant prices, depending on the two exercise variants, and points out that the solution of the warrant game differs from the exercise strategy of convertible bonds in a levered firm. This difference is due to the fact that the exercise proceeds from warrants are invested in the firm, while the conversion of bonds into stock does not affect the firm's value.

2.5 Numerical Approaches

Various numerical approaches have been proposed for the valuation of game options. The evaluation of call, put, and/or conversion provisions embedded in bonds, and, more specifically, the decomposition of the price of a bond into the value of an equivalent "straight" bond and the value of all its embedded options, is of particular interest since such provisions are present in most corporate bonds. Moreover, the interaction of competing provisions offered to different players in game options generally implies that the value of these provisions is not additive.

Finite difference methods were introduced in Brennan and Schwartz (1980); these methods rely on the numerical solution of a stochastic differential equation (SDE) characterizing the evolution, in continuous time, of the value of the game option, contingent to the evolution of the underlying stochastic process; for instance, in Brennan and Schwartz (1980), the two state variables are the value of the firm and the interest rate. The stabilization refinement proposed in d'Halluin et al. (2001) for callable bonds allows for the inclusion of a call notice period.

Dynamic programming approaches have also been proposed to evaluate options embedded in bonds, e.g., trinomial (Hull and White 1990) and binomial (Kifer 2000) trees. These methods rely on a recursive characterization of the value of the dynamic game and of the subgame-perfect equilibrium when the exercise schedule is discrete. A numerical approach combining dynamic programming with finite element interpolation is proposed in Ben-Ameur et al. (2007). The setting allows for continuous models of the stochastic interest rate process and evaluates interacting call and put options with advance notice.

2.6 A Dimensionality Issue and an Identity Crisis

Table 19.1 summarizes the main features of the game-option papers reviewed in this section. The literature on Israeli options is mainly concerned with existence results, in increasingly general settings for the underlying market. However, Israeli options are not commonly traded financial instruments.

On the other hand, warrants and bonds that include embedded options constitute a very important class of traded financial instruments, and their valuation, which relies on the characterization of their exercise strategy, is a fundamental issue. In general, the value of such financial instruments corresponds to the solution of a non-zero stochastic game where players use feedback strategies. In the case of bonds, one distinctive difficulty is the dimensionality of the state space, since the value of a bond depends on the term structure of interest rates (which typically depends on more than one factor) and could also depend on the value of another security (in the case of convertible bonds) or on other market risk factors (in the case of defaultable bonds). For both convertible bonds and warrants, an additional difficulty lies in the number of players involved. Most models presented in Table 19.1 use simplifying assumptions to reduce the state to a single factor, to aggregate the players into two classes, or to attenuate the players' strategic weight.

Finally, it is interesting to note that, while the general model of a game option is a straightforward stopping game between an option holder and an option issuer, the most common finance applications of game options, namely, interacting provisions in corporate debt and financing instruments, raise an interesting issue: the classification of players into two distinct categories with opposing interests. Indeed, abstracting from bankruptcy costs and tax benefits, the value of the firm is divided between the bondholders (lenders) and the equityholders (borrowers), so that exercise strategies result in wealth transfers between equity and debtholders. In reality, individual players may well be both simultaneously; moreover, exercising a conversion option changes the identity—and point of view—of the player of the corporate game, from bondholder to equityholder. We refer to Jalan and Barone-Adesi (1995) where a repeated cooperative game model is proposed to

Table 19.1 Game option papers

Paper	Context	Features	Solution concept
Kifer (2000)	Israeli option	Complete market	Dynkin game
Kühn (2004)	Israeli option	Incomplete market	Nonzero-sum Dynkin game
Hamadène and Zhang (2010)	Israeli option	General stochastic process	Nash equilibrium
Ben-Ameur et al. (2007)	Callable putable bonds	General stochastic process	Dynamic programming
Jalan and Barone-Adesi (1995)	Callable convertible bond	Equity vs bondholders	Repeated cooperative game
McConnell and Schwartz (1986)	Callable putable convertible bond	General stochastic process	SDE
Sîrbu and Shreve (2006)	Defaultable convertible bond	Structural default	Optimal stopping problem
Bielecki et al. (2008)	Defaultable convertible bond	Reduced-form default	Dynkin game
Brennan and Schwartz (1980)	Defaultable callable convertible bond	Structural default, 2 factors	SDE
Chen et al. (2013)	Defaultable callable convertible bond	Optimal default	Stochastic game
Yagi and Sawaki (2010)	Warrant	Simultaneous exercise	Binomial tree
Constantinides and Rosenthal (1984)	Warrant	Sequential exercise	Discrete-time dynamic game
Koziol (2006)	Defaultable warrant	Structural default	Non-atomic game

analyze callable convertible bonds, where the players are current equityholders and bondholders.

3 Bankruptcy Games

In game theory, the *bankruptcy problem* relates to the allocation of a given amount (the estate) among a group of claimants when this amount is not sufficient to cover the total of the individual claims. The liquidation of a bankrupt firm and the distribution of the proceeds to its creditors is an archetypal example of the bankruptcy problem. The solution of the bankruptcy problem consists of identifying rules for this allocation and analyzing its properties (see Thomson 2003, 2015 for a survey); this analysis is usually done within a framework of cooperative game theory.

We define *bankruptcy games* as models of the strategic interactions between various classes of claimants to the assets of a firm. These games are used to determine the solution of the corresponding bankruptcy problem. Unlike the classical bankruptcy problem, where claimants are assumed to differ only in the amount of their claims, bankruptcy games usually involve claims with different priorities and/or players with different leadership or precedence advantages. In particular, there exists a rule—the *absolute priority rule* (APR)—that stipulates the order in which claims will be paid in the liquidation of a bankrupt firm. Bankruptcy games are used to address important issues in both corporate and investment finance, namely, the optimal structure of a firm's debt, the value of debt and equity, and the impact of the solution of the bankruptcy problem on players' incentives to invest and to borrow.

3.1 Liquidation and Optimal Default

The use of the solution of a bankruptcy problem to evaluate defaultable claims was introduced in Merton (1974). This seminal work proposes a model that evaluates debt instruments (bonds) when there is a significant probability of default by the debt issuer (or equityholders). The Merton (1974) model is, however, not designed as a bankruptcy game, since both the moment of default and the compensation received by claimants upon default are defined exogenously. The assumption is that, at maturity, the debtholders instantly and costlessly liquidate the firm's assets and seize the totality of the proceeds if their value does not cover the principal.

Leland (1994) and Leland and Toft (1996) extend the defaultable debt valuation model by considering additional features (tax benefits, bankruptcy costs, and covenants) but, more importantly, assume that the moment of default is determined as an optimal decision by the equityholders. The default decision depends on the firm's value, which in turn depends on when bankruptcy occurs; therefore, these two components must be determined jointly. In such a case, the total value of the firm (debt plus equity) depends on the capital structure, that is, on the relative claims of both types of stakeholders.

These two papers go a step further by using the resulting corporate debt value to determine the optimal capital structure (in Leland 1994, debt is assumed to be perpetual, while in Leland and Toft (1996), the debt maturity is finite). The sequential framework used by the authors can be seen as a two-stage Stackelberg game, where the first stage is a cooperative game that maximizes the total value of the firm (i.e., equity plus debt) by taking into account the optimal equityholders' response, which consists of liquidating the firm when the value of equity reaches zero.

François and Morellec (2004) and Broadie et al. (2007) propose extensions of the optimal default model of Leland (1994) by assuming that liquidation is not immediately triggered by a default decision by the equityholders, that is, that the default or bankruptcy boundary does not coincide with the liquidation boundary. In both cases, the model's essential features are motivated by the U.S. Bankruptcy Code, which includes both the possibility of liquidation (Chap. 7) and reorganization (Chap. 11). In François and Morellec (2004), the firm is liquidated if the value of its assets stays below the bankruptcy boundary longer than a prescribed grace period, and the cash flows generated during this period are shared among the claimholders.[1] In Broadie et al. (2007), two endogenous boundaries are defined, such that the firm can be liquidated if the value of its assets either reaches the liquidation boundary or stays below the bankruptcy boundary for longer than the grace period; the firm does not pay dividends or coupons while it is in default, and it repays a fraction of the accumulated debt arrears if it exits from the default state. The authors compare the equilibrium solutions, according to the identity of the player who decides on the bankruptcy boundary, with the solution maximizing the value of the firm. Variants are proposed in Bruche and Naqvi (2010), where the equityholder acts as a Stackelberg leader, knowing that the creditor decides on the liquidation boundary, and in Bruche (2011), where creditors can eventually coordinate to determine this boundary.

Wang (2011) proposes a model involving the manager as a third player, in order to evaluate the impact of agency problems on finite-maturity defaultable debt valuation. At each stage, the manager decides on the amount of interest and dividend offered to the debtholder and equityholder, respectively, and retains the rest of the cash flow as a private rent. If the interest payment is less than the contractual coupon, the debtholder has the option to liquidate the firm's assets (which involves a liquidation cost) according to the APR. If the firm is not liquidated, the equityholder can accept the dividend payment or dismiss the manager (which involves a loss in human capital cost). The magnitude of this loss defines management's entrenchment power. The equilibrium solution is obtained numerically by backward recursion and

[1] The sharing rule is not determined according to the APR but rather as a Nash bargaining solution where the players have exogenously fixed bargaining powers.

is used to characterize the impact of management entrenchment on corporate debt and dividend policies.

3.2 Strategic Debt Service

In the models described in the previous section, even if part of the debt arrears can be forgiven in some cases, the contractual payments remain set according to the original agreement over the horizon of the game. The notion of *strategic debt service* can be ascribed to Hart and Moore (1989, 1997) where a debt contract is modeled as a multistage game between an entrepreneur and an investor, involving a possible renegotiation of its terms. In this game, the entrepreneur runs her project as long as she honors the debt contract, which requires a fixed stream of payments to the investor. If the entrepreneur defaults, the investor can seize and liquidate the project's assets. At this stage, the entrepreneur and investor can renegotiate the contract. When liquidation is costly, the investor may be better off accepting a reduced payment rather than seizing the project's assets.

Within that framework, Anderson and Sundaresan (1996) proposes a dynamic game model based on the possibility of strategic private negotiations between a firm's equityholders and bondholders. The game's equilibrium between the two types of players yields the value of the firm's risky debt. The dynamic game in Anderson and Sundaresan (1996) assumes that leadership in negotiations is taken on by the equityholder. At discrete dates (e.g., coupon dates), the equityholder can propose a reduced debt service. The bondholder then has the option of triggering bankruptcy procedures, resulting in a costly liquidation of the firm's assets to reimburse the outstanding debt. At a given date where such negotiations take place, the game can be represented by the payoff matrix in Table 19.2 (the equityholder being the row player and the bondholder, the column player), where f is the portion of the firms' assets that is periodically paid out to the shareholders; s is the proposed debt service; c is the contractual coupon; $\tau \in (0, 1]$ is the tax rate; $V^l = V_1^l + V_2^l$ is the firm's liquidation value, which is shared between the two players according to the priority of their claim (the solution of the bankruptcy problem); and E^c and D^c are the expected values at the next negotiation date for the equityholders and the debtholders, respectively. The continuation and liquidation values are contingent on the underlying process X_t.

In the Anderson and Sundaresan (1996) model, X_t is the value of the firm's assets, which are assumed to be observable up to t by both players. Moreover, it is assumed that cash flows f_t are an exogenously fixed proportion of the assets X_t and that debt service at date t can only be met out of f_t; otherwise, the firm

Table 19.2 Private negotiation payoff matrix

	Accept	Liquidate
s	$((1-\tau)(f-s)+E^c, s+D^c)$	(V_1^l, V_2^l)
c	$((1-\tau)(f-c)+E^c, c+D^c)$	–

is liquidated. This last assumption makes the evolution of X_t independent of the debt service choice; therefore, the equilibrium solution is for the equityholder to propose $\min\{V_2^l - D^c; c\}$ to the bondholder—that is, a reduced debt service just high enough to avoid liquidation—provided this is feasible. The game is then solved by backward recursion from maturity, where the bondholder's continuation value is zero. Notice that, since V_2^l, D^c, and f are contingent on the asset value, three regions can be identified according to the value of X_t at a given decision date: when $c < V_2^l - D^c \leq f$, the equityholder pays the contractual coupon; when $V_2^l - D^c > f$, the firm's assets are liquidated and divided among the claimants; otherwise, the bondholder agrees to receive an amount $V_2^l - D^c$, which is less than the contractual coupon.

The Anderson and Sundaresan (1996) model triggered a sizeable literature on strategic default, where the value of risky debt is determined from the equilibrium solution of a dynamic game between shareholders and bondholders.

Anderson et al. (1996) proposes a continuous-time representation of the Anderson and Sundaresan (1996) model, leading to a characterization of the value of risky debt through a Black- and Scholes-type partial differential equation for a general class of strategic models, where the service flow and the boundary conditions are determined by the solution of the one-stage game presented in Table 19.2.

Mella-Barral and Perraudin (1997) makes use of a continuous-time model where the underlying asset is the firm's output price and the bond contract is perpetual. By characterizing the strategies used by the players taking a leadership role, the authors obtain a closed form for the value of risky debt and for the debt service function. Two contrasting cases are studied: in the first case, equityholders make take-it-or-leave-it offers to the bondholders regarding debt service; in the second, the bondholders are the ones who take leadership.

In the abovementioned models, it is assumed that debt can only be serviced out of cash flows (available liquidity), which allows for the possibility of liquidation when these cash flows are not sufficient to cover the difference between the liquidation and the continuation value for the debtholders. The portion of cash flows that is not used for debt service is then paid out to equityholders as dividends. Models accounting for dividend policies are proposed in Fan and Sundaresan (2000) and Acharya et al. (2006).

In Fan and Sundaresan (2000), instead of assuming that one of the players takes leadership in the negotiation game and can make take-it-or-leave-it offers to the other player, a Nash bargaining solution (NBS) is used to determine the outcomes of negotiation games and liquidation, where the claimants' bargaining powers are given constants. The players can bargain over the firm's assets, giving rise to debt-equity swaps, or over the firm's value, where the equilibrium outcome is a reduced debt service. Using the NBS as the solution concept in continuous time allows for analytical solutions, and the paper characterizes the dividend rate that maximizes the value of equity, and the circumstances under which it is optimal to issue new equity to finance contractual coupon payments.

Acharya et al. (2006) proposes an extension of the (discrete-time) Anderson and Sundaresan (1996) model, where the equityholders have the additional options of setting up cash reserves (i.e., of deciding on the amount of the dividends), or of raising cash by issuing additional equity, in order to avoid liquidity defaults. The authors use numerical solutions to show that strategic debt service has an important impact on the debt value when the cost of issuing new equity is low.

3.3 Insolvency and Reorganization

In the models presented in the previous section, the implicit assumption is that of repeated private and costless negotiations (workouts) between two players: a single equityholder, or a manager aligned with the interests of the equityholders, and a single debtholder (e.g., a bank). Moreover, in strategic debt service models, while temporary relief can be negotiated, the debt structure itself is assumed fixed over time. The present section briefly reviews dynamic game models that better represent public debt and negotiations regulated by insolvency laws, possibly involving many creditors with different claim priorities, and allowing for a restructuring of the debt.

Brown (1989) analyzes the reorganization process of a distressed firm as a game with successive negotiation rounds, where the rules of the game (e.g., number of rounds, sequence, voting rules, claim priority) are formalized by the bankruptcy code. The equilibrium solution of the reorganization game is obtained by backward induction; still, this is a static solution, in the sense that the firm's liquidation and reorganization values are taken as constants.

In Berkovitch and Israel (1998), a two-period game is defined between an owner/manager and a creditor, where the creditor can decide to liquidate the firm' assets at date $t = 1$ after observing an imperfect signal of its profitability. If the creditor does not decide to liquidate, the equityholder chooses between either liquidating the firm, paying the debt obligation, and continuing operations or renegotiating the debt contract privately or under court supervision. If it is still operating at the end of the second period, the firm is liquidated. The outcome of the game in this model is shown to be related to asymmetric information and to the manager's decision to over- or underinvest during the first period.

Mella-Barral (1999) develops a pricing model with dynamic debt restructuring, where debt is assumed perpetual, with continuous coupons. The model takes the form of a differential game with a Stackelberg information structure, where cohesive players are characterized by their bargaining power, and negotiation is costless. Debtors are the Stackelberg leaders: they decide on the default event by maximizing the value of equity, taking into account the optimal reaction of the creditors, who can choose, in the event of a default, between forcing a liquidation, decreasing the coupon, and changing the claimants' shares if the assets are liquidated. The equilibrium solution yields closed-form pricing formulas for equity and debt. This model is used to explain debt restructuring and departures from the APR upon liquidation, both of which can be observed in practice. More specifically, the author shows that when the firm's leverage is relatively high, debtors will tend to default early, and creditors will offer to reduce the coupon in order to defer liquidation; while, when the firm's leverage is relatively low, debtors will tend to default late so creditors will offer them a larger share of the proceeds than is stipulated by the APR, in order to precipitate liquidation.

Hege and Mella-Barral (2005) proposes an extension of the model in Mella-Barral (1999), allowing for multiple creditors who are not coordinating their actions. Renegotiation is assumed costless (out-of-court). The debtor is the leader in a Stackelberg differential game; her strategy is to offer a series of new debt contracts, over time, to a limited number of creditors, in exchange for the voluntary surrender of old contracts. These new contracts offer more liquidation rights in exchange for debt service concessions. In each offer, the debtor commits to defaulting on debt service payments if the offer is rejected, thereby triggering the (costly) liquidation of the firms' assets. It is assumed that there is a very large number of creditors, so that each creditor neglects her impact on the success or failure of a debt-restructuring proposal. The authors derive closed-form solutions for the value of equity and defaultable debt. They find that, with respect to single-creditor debt, creditor dispersion limits the size of concessions that can be obtained by an opportunistic debtor and typically results in greater optimal leverage for the firm.

Moraux and Silaghi (2014) also extend the model of Mella-Barral (1999) by incorporating fixed renegotiation costs in a model of multiple renegotiations. As a result, the model predicts a finite number of renegotiations instead of infinitesimal coupon reductions and makes it possible to compute the optimal number of renegotiations. Two polar cases are considered, where either the equityholders or the creditors are leaders and have all the bargaining power. The leader decides on the renegotiation thresholds, the coupon reductions, and the optimal number of renegotiations and is assumed to pay the renegotiation costs.

In Noe and Wang (2000), a three-stage game characterizes strategic behavior during debt renegotiations involving a manager and two creditors. The latter differ according to the size of their claim and the amount they can recover from liquidation. The manager negotiates sequentially with each creditor; she can choose the sequence and the nature of her restructuring offer, while creditors can accept, pass, or reject offers. Game dynamics only pertain to the sequence of decisions during the negotiation process, since cash flows and liquidation values are given constants. The model is solved by backward induction, yielding a subgame-perfect Nash equilibrium. The authors show that strategic flexibility is valuable to distressed firms: they use the claims of the creditor who is in a better position (e.g., with a smaller loan or a larger recovery) to extract greater concessions from the other creditor.

Annabi et al. (2012a,b) consider legal bankruptcy procedures and model the resolution of financial distress as a noncooperative game between claimants under the supervision of a bankruptcy judge, who is a non-strategic player. The game is played in discrete time and consists of costly successive negotiation rounds involving the debtor and two creditors with different seniorities. In each negotiation round, one of the claimants acts as leader and proposes a debt-restructuring plan, that is, a vector of new perpetual coupons replacing the existing contractual coupon pair. The two other players act as followers and play a unanimity game: the plan is implemented, and the reorganized firm emerges from bankruptcy procedures if both agree on the proposed restructuring. There is a positive probability that the bankruptcy judge intervenes and imposes a "fair and equitable" plan (i.e., an NBS solution) if the players do not agree. Liquidation occurs when cash flows cannot cover the cost of negotiation. The game is solved by backward induction, and the equilibrium solution is obtained numerically. In Annabi et al. (2012b), the context is Chap. 11 of the U.S. Bankruptcy Code, while Annabi et al. (2012a) consider a more general setting allowing for various provisions in the legal procedure. These papers find that the identity of the class of claimants proposing the first reorganization plan is a key determinant of the time spent under bankruptcy, the likelihood of liquidation, and the renegotiated value of claims.

Christensen et al. (2014) proposes a continuous-time model where the capital structure is dynamic, with a single class of callable perpetual debt. Equityholders continuously decide whether to continue servicing the existing debt or to restructure the firm's capital. Restructuring consists of either calling existing debt (and subsequently issuing new debt) or making a take-it-or-leave-it offer to the debtholders to reduce the existing coupon. The model departs from the usual strategic debt service literature by assuming that debtholders will not accept offers that are not credible (for instance, threatening to liquidate the firm, since this would normally leave the equityholders with nothing). Consequently, at equilibrium, the benefits of renegotiation are shared among the claimants according to their exogenously given bargaining power. The model also considers an additional state variable by assuming that the equityholders have a limited number of renegotiation options. Note that the introduction of the possibility of calling and reissuing the debt when the firm is in a good position makes the model richer and can be related to the literature on defaultable game options presented in Sect. 2.3; however, while defaultable callable bond models consider a liquidation event, in Christensen et al. (2014), renegotiation ensures that liquidation never occurs. The equilibrium solution is obtained by backward induction and is used to show that violations of the APR are to be expected.

3.4 Bridging Investment and Corporate Finance

Table 19.3 presents the main features of the bankruptcy game papers reviewed in this section. These bankruptcy games typically model the strategic interactions between creditors and debtors, often aggregated into two representative players, and sometimes involve management as a third player. Strategies vary in complexity, from simple decisions, such as defaulting on debt payments or triggering bankruptcy procedures, to more intricate compromises involving temporary relief of debt service or changes in the debt contract. In many cases, special consideration is devoted to the impact of leadership, according to the identity of the player who moves first.

One important issue in this literature is the possibility that the firm emerge (or not) from bankruptcy and the circumstances leading to its liquidation. The interactions between the players are used to define the solution of the bankruptcy

Table 19.3 Bankruptcy game papers

Paper	Strategies (player)	Leader	Liquidation	Other features
Leland (1994); Leland and Toft (1996)	D(d)		Default	
François and Morellec (2004)	D(d)		Sojourn	
Broadie et al. (2007)	B,E(d or c)		Sojourn or barrier	
Bruche and Naqvi (2010)	D(d) ; E(c)	d	Barrier	
Bruche (2011)	D(d) ; E(c)	d	Barrier	Multiple creditors
Wang (2011)	S,Q(m) ; B(c) ; X(d)	m	Creditor	
Hart and Moore (1989)	S(d) ; B(c)	d	Creditor	Multistage game
Anderson and Sundaresan (1996)	S(d) ; B(c)	d	Cash-flows	
Anderson et al. (1996)	S(d) ; B(c)	d	Cash-flows	
Mella-Barral and Perraudin (1997)	S(d or c)	d or c	Follower	
Fan and Sundaresan (2000)	S		Bargaining solution	Bargaining solution
Acharya et al. (2006)	S,Q,R(d) ; B(c)	d	Cash-flows	
Berkovitch and Israel (1998)	B(c) ; R(d)	c	Creditor or debtor	Two-stage imperfect information
Mella-Barral (1999)	D(d) ; B or S(c)	d	Follower	Bargaining power
Hege and Mella-Barral (2005)	R or D(d)	d	Follower	Multiple non-atomic creditors
Moraux and Silaghi (2014)	E,R(d or c)	d or c	Follower	Costly negotiation
Noe and Wang (2000)	R(m)	m	Negotiation failure	Three stages; asymmetric creditors
Annabi et al. (2012a,b)	R(d or c)	turns	Cash-flows or judge	Asymmetric creditors; costly negotiation
Christensen et al. (2014)	R (d)	d		Bargaining power

Legend: *D* default on debt payment, *B* trigger bankruptcy, *E* set endogenous barrier, *S* temporary debt service, *R* debt restructuring, *Q* dividends, *X* dismiss manager, *d* debtor, *c* creditor, *m* manager problem, that is, the amount that various claimants expect to recover, whether bankruptcy leads to emergence or to the liquidation and redistribution of a firm's assets. In some cases, the equilibrium solution can also be used to define how the cash flows will be shared among players (e.g., coupon vs. dividends) during the life of the firm.

The equilibrium value of bankruptcy games can be interpreted as the valuation of debt and equity, expressed as a function of the underlying state variables. For this reason, bankruptcy games provide an important contribution to investment finance. Indeed, the possibility of financial distress should be taken into account when valuing debt instruments, such as bonds, when deciding whether or not to invest in a given entrepreneurial project, or when determining the market value of a stock.

The solution of bankruptcy games can also be used at a higher level to decide on the values of the parameters of these games, such as, for instance, the capital structure or the dividend payout rate. As will be seen in the next section, bankruptcy costs and the way financial distress is resolved play a significant role in many corporate finance models.

4 Corporate Games

A number of theoretical models in corporate finance were developed to explain or predict the financial decisions firms make, including the following:

1. The choice between debt and equity to finance operations
2. The amount of dividends paid out to shareholders
3. Decisions on whether or not to invest in projects.

In many cases, the predictions made by theoretical models are not in line with what is actually observed in the corporate world: firms use less debt than expected given the relative weight of tax advantages vs. bankruptcy costs, and they pay more dividends than is optimal for shareholders given that personal capital gains are usually taxed less than dividend revenues.

Many corporate game models have been proposed in the literature to interpret some puzzling empirical findings in the light of strategic interactions and asymmetric information, the largest proportion being devoted to capital structure issues. The following subsections present a selection of dynamic corporate game models addressing capital structure, dividend payout, and investment policies.

4.1 Capital Structure

There are essentially two ways for a firm to obtain the funds required to finance its activities: *debt contracting* involves committing to deterministic payments to the lenders over a given horizon (debt service), while *equity financing* implies selling part of the firm's ownership, and, therefore, its claims on future profits, to outside investors. The models presented in this section address the debt vs. equity mix for financing a firm's operations, the choice between bank (private) and bond (public) debt, the inclusion of contractual rights granted to debtors, and, in some cases, the choice between different types of investors.

The following sections depict the various classes of games that have been proposed in the literature to explain such capital structure decisions.

4.1.1 Signalling Games

Myers and Majluf (1984) proposes a two-period game between a firm's management and potential investors when the firm needs financing to take advantage of an investment opportunity. The game is one of asymmetric information about both asset and project values: At $t = 0$, the probability distributions for the value of the firm's assets and the value of the investment opportunity are known by all players. At $t = 1$, management observes the realization of both uncertain variables and decides whether or not to issue new shares to finance the investment opportunity. Uncertainty is resolved for the investors at $t = 2$.

By assuming that management's interests are aligned with those of the existing shareholders, the authors show that, in some cases, it is in management's best interest not to issue shares for profitable investment opportunities, so that its decision to not issue shares sends a positive signal to investors and thereby affects the firm's market value. Equilibrium results are used to explain empirical observations, such as the facts that stock prices often fall when an equity issue is announced and that firms prefer to finance their operations using internal funds or debt, rather than equity.

Noe (1988) refines the setting in Myers and Majluf (1984) by explicitly modeling the firm's financing decision as a sequential signalling game of two periods. Management strategies are messages (issue debt, issue equity, or request no financing). Investors respond by refusing to finance or by offering an amount of either debt or equity. Beliefs about the firm's quality are revised using Bayes' rule. As in Myers and Majluf (1984), debt is preferred to equity financing when management has perfect information. The author shows that this result does not necessarily hold when management has imperfect information. However, the choice of financing still has an informational impact, since higher-quality firms are more likely to choose debt over equity financing.

Similar two-period sequential signalling games, where management is informed, while the investors are not and revise their beliefs based on debt-choice messages sent by management, are also proposed in Constantinides and Grundy (1989), who investigate the role of straight and convertible debt and of stock repurchasing as signals, as well as in Kale and Noe (1990), where signals are debt maturity choices (long or short term).

A multiperiod signalling game is proposed in Gomes (2000) to address the agency problem between managing shareholders and minority shareholders. The model is a stochastic dynamic game played over a finite number of periods, where the management signal consists of its level of effort and the number of shares traded.

Outsiders use this information to price shares according to their belief about the type of management. The author uses the features of the signalling equilibrium to explain why, even when there are no mechanisms to protect minority shareholders, management will not expropriate them in order to maintain its reputation, which results in higher stock prices and a gradual divestment of its shares over time.

4.1.2 Bargaining Games

Building on the strategic debt service model of Hart and Moore (1989, 1997), some dynamic games relate capital structure decisions to the anticipated outcome of negotiations between the firm and its investors in cases of a default, which can be triggered under two distinct circumstances (liquidity and strategic defaults) and which gives creditors the right to liquidate the company's assets. Bargaining game models aim to characterize optimal debt contracts that deter strategic defaults and avoid costly liquidation. Capital structure bargaining games are generally played in two stages: at $t = 1$, the parties agree to a financing contract; at $t = 2$, the firm decides whether or not to default on its contractual obligations, where, in the case of default, a bankruptcy game is played.

Berglöf and Von Thadden (1994) proposes a bargaining game where the firm has all the bargaining power and makes a take-it-or-leave-it offer to investors. The authors show that choosing a capital structure with multiple investors with different maturity and seniority claims is better than choosing only one type of claim. They argue that having a variety of claims reduces the incentive for the firm manager to renegotiate the initial debt contract. Bolton and Scharfstein (1996) analyzes a bargaining game under a general contract—specifying the conditions under which creditors have the right to liquidate a fraction of the assets with a given probability—in order to look at specific aspects of debt structure. The authors find that firms with a low credit quality should maximize liquidation values, which translates to having a single creditor and covenants facilitating asset sales, while firms with a high credit quality should make strategic default less attractive, which translates to having multiple creditors and voting rules that allow some creditors to block asset sales. Park (2000) considers the case where moral hazard is severe and allows for monitoring by the lenders. The author finds that, under the optimal debt contract, monitoring is performed by a single senior lender. The model provides a rationale for prioritizing debt contracts and explains why bank debt is usually short-term and senior to widely held long-term debt. Berglöf (2000) finds that the optimal debt contract involves multiple creditors: this simultaneously increases the entrepreneur's capacity to raise funds and increases the occurrence of strategic default.

4.1.3 Sequential Games

The papers reviewed in this section use a sequential game framework to analyze the financing decisions of entrepreneurial firms, where a contract is proposed and an investment is made at $t = 0$. At $t = 1$, the entrepreneur chooses whether or not to default on her obligations after privately observing output or after deciding on her level of effort. The lender can then take action at $t = 2$ (e.g., according to the contract's specifications, monitor the project, liquidate the firm, or accept a payment offer from the entrepreneur). Returns are distributed at $t = 3$. Variations of this general framework are used to propose various determinants of capital structure related to monitoring, control, and enforcement parameters.

Repullo and Suarez (1998) considers the financing mix (bank loans vs. bonds and equity) by characterizing lenders according to their monitoring capability. Entrepreneurs decide on their level of (costly) effort, which determines the probability of positive returns. They are characterized by their wealth and by the deterministic liquidation value of their project. The solution of the game between the entrepreneur and the lender is a contract specifying the amount invested by the lender and each player's share of the returns and liquidation proceeds. When lenders are uninformed, they do not act at $t = 2$; feasible contracts are then characterized by the Nash equilibrium of a matrix game, with each player having two pure strategies (to participate or not). When lenders are informed, they can decide to liquidate the firm at $t = 2$; feasible contracts are then characterized by the subgame-perfect equilibrium of the sequential game, where the lender's liquidation decision depends on the observed level of effort by the entrepreneur.

The authors also analyze the optimal contract under mixed finance, which is characterized by the solution of a sequential game between the entrepreneur and two different types of lenders, and the possibility of a renegotiation of the original contract, where the entrepreneur and the informed lender may collude to change their share of the returns after the effort decision has been made. The analysis yields a characterization of the circumstances, in the space of entrepreneur wealth and project liquidation value, under which the various configurations for the financing mix are optimal. The sequential game model is able to explain a number of empirical observations, such as the fact that many firms are funded by a mix of informed lenders (such as banks) and uninformed lenders (such as bondholders) and that informed debt is senior to uninformed debt in case of liquidation. The model also predicts that investments involving liquid assets are more likely to be funded exclusively by informed lenders.

Krasa and Villamil (2000) relates the characteristics of an optimal contract to the time consistency of enforcement strategies. At $t = 0$, both players have common beliefs about the distribution of returns. At $t = 1$, the entrepreneur privately observes the outcome and decides on a voluntary payment to the investor. This payment is used to update the investor's belief at $t = 2$, when the investor can decide to enforce a final payment or not. Enforcement is provided by an outside agent (e.g., a bankruptcy court) and is costly for both players. The optimal contract, specifying the final payment as a function of the realized outcome, is obtained by solving for a perfect Bayesian equilibrium maximizing the agents' returns. The authors show that simple debt is the optimal contract when players cannot commit to open-loop strategies and renegotiation is possible, whereas a stochastic contract is optimal otherwise.

Krasa et al. (2008) uses a similar framework to analyze the role of bankruptcy parameters on firm finance and more precisely on loan rate, default probability, and welfare. The legal enforcement system is characterized by the enforcement costs and the level of debtor protection. At $t = 1$, the entrepreneur decides to default or not on her contractual payments. At $t = 2$, the lender can decide to request enforcement of the contract or not. The authors solve for Pareto-efficient equilibria and show that the legal enforcement system may have an important impact on financing and bankruptcy decisions.

Hvide and Leite (2010) also relates financial structure and default behavior to the cost of enforcement. The model assumes that the entrepreneur can finance her project using debt, equity, or a mix of both. At $t = 1$, the entrepreneur decides on the amount offered to the investors, who can decide to request contract enforcement or not, at $t = 2$. Under the assumption that intervention costs are higher for debt than for equity, the authors show that debt is the preferred security and that equity is issued in combination with debt when the funding requirements are high.

In von Thadden et al. (2010), strategic interactions between lenders are called upon to analyze the interrelationship between bankruptcy rights, corporate debt structure, and debt contracts. The sequential game model involves one firm and two lenders, and the debt contract is assumed to contain individual foreclosure rights that may differ from the claimants' share in bankruptcy proceeds. At $t = 1$, according to the cash flow realization, the firm decides on the amount offered to each creditor separately. At $t = 2$, if what she is offered is less than the contractual amount, a creditor chooses between accepting the offer or foreclosing on the firm's assets. The outcome of the game depends on the creditors' joint decision, since bankruptcy is triggered when both creditors reject the firm's offer. Contracts are chosen to maximize the firm's expected payoff, subject to the investors' participation at $t = 0$, where the players' strategies are obtained by solving for the subgame-perfect Nash equilibrium corresponding to a given contract. The solution of the game yields a number of properties for the optimal contract; in particular, the optimal contract involves more than one creditor and cannot result in unilateral foreclosure. The game model is able to generate instances of strategic default or violations of the APR.

Gennaioli and Rossi (2013) suggests that the level of investor protection is a characteristic of the legal system, which is a determinant of both the capital structure and the terms of the lending contract achieving the first-best solution for the firm. The model considers three possible outcomes for cash flow realizations over time (good, bad, and temporary financial distress) and assumes that liquidation and reorganization values are known by all players at $t = 1$. The terms of the contract specify the amount loaned, the identity of the player controlling liquidation and reorganization decisions, and the amounts received by claimants according to the identity of the controlling player. In the single-creditor case, the authors show that the optimal contract depends on the level of investor protection. They then show that the optimal capital structure consists of two classes of creditors, where one large creditor controls the reorganization/liquidation decision and the other class contains many dispersed small creditors who have no claim in the reorganized firm.

Buehlmaier (2014) proposes that the debt vs. equity choice depends on the characteristics of the investment project (expected return and risk). The model is an extension of the one in Krasa and Villamil (2000), offering a larger strategy space to the investor at $t = 2$. After updating her beliefs according to the voluntary payment offered by the entrepreneur, the investor can decide to exercise monitoring rights or not. The investor can then either accept the voluntary payment, request enforcement of the contractual final payment, or initiate a series of renegotiation rounds. A renegotiation round consists of a payout request from the investor, which can be accepted or rejected by the entrepreneur, and where the entrepreneur's response is used to update the investor's beliefs.

The optimal contract, specifying the enforceable final payment and whether or not the investor has monitoring rights, is obtained by solving the sequential game. The author shows that when expected returns and risk are relatively low, as compared to the monitoring costs, the optimal contract has the characteristics of simple debt, whereas, in the opposite case, it has the characteristics of equity.

Meneghetti (2012) proposes a different outlook, linking capital structure choices to managerial incentive compensation. As in the previous models, the manager's compensation depends on the outcome of the project, but it is also assumed to have an incentive component corresponding to a fraction of the firm's value. At $t = 1$, the manager decides whether to invest in a safe or a risky project, where the safe project has lower expected returns. It is assumed that banks can (imperfectly) monitor the manager's action and liquidate the firm when they receive a signal that the manager has invested in the risky project. The lenders decide on the parameters of the contract (interest rate and collateral), while the manager chooses between bonds and bank debt.

The game's equilibrium solution depends on the level of the incentive component in the manager's compensation. For low values, the manager chooses bonds and invests in the safe project. As the incentive component increases, the manager chooses bank debt and invests in the safe project. For high values of incentive compensation, the manager chooses bank debt when the signal is precise and bonds when the signal is imprecise. The author uses a sample of bank loans and bond issues in the USA between 1993 and 2005 to show that model predictions are supported empirically.

4.1.4 Repeated Games

Fluck (1998) considers a repeated game between an entrepreneur-manager and an investor, where the investor can hold either debt or equity. The project requires investment in equipment that has to be replaced every two periods and produces uncertain (high or low) cash flows, which are observable at the end of the two periods. When the investment is made by equityholders, the manager decides on the amount of dividends and the amount of a depreciation allowance used to replace the equipment. Equityholders can decide to liquidate the company or to replace or keep the manager at any time. When the investors are debtholders, the manager decides whether or not to default on her contractual payments, which triggers a bargaining game as in Hart and Moore (1989).

The author shows that the only sustainable equity contracts have an unlimited lifespan, while the maturity of equilibrium debt contracts matches the lifespan of the assets and then characterizes projects that can raise debt, equity, or both, according to the variability of cash flows. The results are consistent with empirical evidence, where lower debt-equity ratios are observed in industries with a higher cash flow risk.

A similar setting is used in Anderson and Nyborg (2011) where, however, an additional exogenous event is considered in order to examine how corporate ownership affects growth. The model assumes that, from some given date, outside management becomes more efficient than the original entrepreneur. The equilibrium solution confirms that leverage is inversely related to growth and profitability, as predicted by empirical evidence.

Another repeated game of corporate ownership and control between management and equityholders is presented in Fluck (1999), where management decides on the amount of dividends, while, at each round, dispersed equityholders can decide to retain or to fire management. In the latter case, the probability of success of a control challenge depends on the number of outsiders and their strategic value, as computed using the Shapley value of a majority voting game. Accordingly, management, existing shareholders, or outsiders can strategically decide to purchase or sell shares in the company.

The author finds that corporate ownership is related to the cost of capital or to the investors' time preference. When the cost of capital is low, management holds a negligible stake in the company, while it accumulates shares when it is high.

4.1.5 Differential Games

In Hilli et al. (2013), a differential game setting is proposed to explain the observed tendency of firms to evolve from a concentrated to a dispersed ownership. The model involves one manager, as well as one large and many atomistic shareholders. As different types of projects become available over time, the manager decides on her level of effort, which allows her to obtain information on project types and states of nature. The large shareholder simultaneously decides on her monitoring level, which allows her to learn what the manager knows.

The authors show that the large investor should divest her shares over time, and compare the prior commitment strategy, which is not time consistent, with the Markov-perfect equilibrium. They find that the divestment rate is related to the degree of divergence between the interests of the manager and shareholders. When the divergence is mild, all shares are sold immediately, but when divergence is high, divestment is gradual, to prevent a fall in the share price.

4.1.6 Financing Decisions

Capital structure papers presented in this section mainly cover two distinct themes. A first stream of papers is concerned with firm's ownership. Table 19.4 summarizes the features of the papers analyzing the decision to issue equity to finance a firm's activities. Questions covered in this literature include the decision to issue additional shares, the choice between debt and equity, and the characteristics of a firm's shareholders. A second stream of papers, presented in Table 19.5, is concerned with debt contracting. The issues analyzed in these papers include the choice between private and public debt, the concentration of creditors, and the various possible characteristics of debt contracts, including control and bankruptcy rights.

Table 19.4 Financial mix papers

Paper	Capital structure decision	Strategies (player)	Results
Myers and Majluf (1984)	Equity issue	Signal (d)	Impact of share issue on market value
Gomes (2000)	Equity issue	Signal (d)	Impact of divestment on market value
Noe (1988)	Debt vs equity	Signal (d)	Firms of higher-quality choose debt over equity
Buehlmaier (2014)	Debt vs equity	P(d); M, B, R (c)	High monitoring costs lead to debt
Hvide and Leite (2010)	Debt, equity or mix	P (d); B (c)	Impact of funding requirements
Fluck (1998)	Debt, equity or mix	Q, D (m); B, X (c)	Mix depends on the variability of cash flows
Anderson and Nyborg (2011)	Debt, equity or mix	Q, D (m); B, X (c)	Impact of growth and profitability on leverage
Fluck (1999)	Corporate ownership	Q (m); X (d)	Corporate ownership depends on cost of capital
Hilli et al. (2013)	Ownership concentration	E (m); M (d)	Concentrated to disperse ownership

Legend: *D* default on debt payment, *B* trigger bankruptcy, *E* level of effort, *R* debt restructuring, *Q* dividends, *X* dismiss manager, *P* payment amount, *M* monitor, *d* debtor, *c* creditor, *m* manager

Table 19.5 Debt structure papers

Paper	Debt structure decision	Strategies (player)	Results
Berglöf (2000)	No	P (d); F, B (c)	Multiple creditors
Berglöf and Von Thadden (1994)	No, maturity, seniority	D (d); B (c)	Asymmetric claims reduce renegotiation
Bolton and Scharfstein (1996)	No, liquidation rights	D (d); B (c)	Impact of credit quality
von Thadden et al. (2010)	No, foreclosure rights	P (d); F (c)	Multiple creditors
Gennaioli and Rossi (2013)	No, control rights	D (d); B, R (c)	Multiple creditors in two classes
Krasa and Villamil (2000)	Simple debt vs conditional contract	P (d); B (c)	Simple debt when renegotiation is possible
Krasa et al. (2008)	Debtor protection	D (d); B(c)	Impact of the legal enforcement system
Park (2000)	Bank vs bond, priority	E (d); M, B, R (c)	Bank is senior to long-term bond debt
Repullo and Suarez (1998)	Bank vs bond, priority	E, D (d); B, R(c)	Bank senior to bond; asset liquidity impact
Meneghetti (2012)	Bank vs bond, managerial ownership	E (m); B (c)	Impact of incentive compensation level

Legend: *No* number of creditors, *D* default on debt payment, *B* trigger bankruptcy, *E* level of effort, *R* debt restructuring, *P* payment amount, *M* monitor, *F* foreclose, *d* debtor, *c* creditor, *m* manager

As in bankruptcy games, the players involved are creditors and debtors. It is interesting to note that the possibility of default and bankruptcy is a very important ingredient of many capital structure games and that most debt contract characteristics modeled in the debt structure papers relate to the rights that are granted to the creditor in the case of default and the priority of claims in the case of bankruptcy.

4.2 Dividend Policies

There are mainly two ways available to corporations for distributing surplus to stockholders: dividend payments and share repurchases. While the impact for the firm of the two instruments is equivalent, in many countries, dividends are taxed at a higher rate than are capital gains, and therefore, share repurchases should dominate dividend payouts. However, empirical evidence shows that dividends are consistently more popular than share repurchases in the corporate world. This is one of the puzzles in corporate finance (see Black 1976).

One of the popular rationalizations for the widespread distribution of dividends is that they are used by firms to communicate private information to investors, namely, the expected value of future earnings. Under that assumption, a second puzzling fact emerges: dividends are often smoothed over time, sticky, and imperfectly correlated with earnings.

Bhattacharya (1979) is one of the first papers to propose a signalling model to derive an optimal dividend policy, where the objectives of investors and managers are aligned. The author also discusses the issue of multiperiod planning horizons, pointing out the difficulty of accounting for dynamics and learning in dividend policy signalling models.

In response to this paper, various signalling game models of dividend distribution under asymmetrical information emerged in the literature, usually in the form of sequential games of incomplete information over two periods, possibly repeated: At date $t = 1$, the manager of the firm decides on the allocation of earnings to investments and dividends, on the basis of private information (e.g., current earnings or productivity). At date $t = 2$, the investor observes the dividend paid and uses this information to revise her belief about the firm's value and, in some cases, to decide whether or not to invest additional funds into it. The total earnings are then distributed and the game is terminated.

In many of these models, the choice of a dividend level determines the amount invested in the second period, such that the firm's dividend and investment policies are interdependent. This is notably the case in signalling models where the investor does not act in the second period, as in Miller and Rock (1985), Guttman et al. (2010), and Baker et al. (2015). In these models, investors use the information conveyed by the dividend to valuate the stock, and the equilibrium is characterized by the manager's choice of a dividend level and the corresponding investors' belief regarding the firm's value. Miller and Rock (1985) establishes the existence of a time-consistent equilibrium that leads to lower levels of investment than would be optimal under full information. Guttman et al. (2010) shows that dividend stickiness can be explained by a partially pooling equilibrium where dividend policy is constant over a range of current earnings. The authors show that underinvestment is lower when investors and management are able to coordinate to select such an equilibrium. Baker et al. (2015) proposes a multiperiod model where investors evaluate current dividends against a reference point established by past dividends and are averse to dividend cuts. This model is able to explain observed patterns of managers' dividend policies.

Signalling models where investors are active players include Kumar (1988), Kumar and Lee (2001), and Allen et al. (2000). Kumar (1988) models a game between an entrepreneur-manager and a representative shareholder, who differ in their attitudes toward risk, and where the shareholder does not know the manager's productivity. The author shows that, because the two players' objectives are not perfectly aligned, no signalling equilibrium exists in pure strategies. Conditions are derived under which a coarse signalling equilibrium can be found, in the sense that dividends are expressed as a step function of managerial productivity. This result is offered as an explanation for dividend smoothing. Kumar and Lee (2001) proposes an extension of this game to a multiperiod setting, where the manager maximizes her terminal wealth while the representative investor maximizes her consumption stream.

Allen et al. (2000) proposes a sequential game between the managers of two types of firms and two types of investors (institutional and retail investors). Managers have inside information on the value of their firm and choose a dividend policy maximizing its expected share price. Investors differ in their risk aversion, tax rate, and ability to monitor management. Investors trade together and allocate their wealth between the two types of firms in order to maximize their expected utility. The authors solve the game by backward induction and obtain conditions for a separating equilibrium to exist. They also consider an extension to a three-period game by adding a cooperative stage where institutional investors jointly decide on the amount of monitoring they will apply to firms, where monitoring is costly but increases the value of earnings. The authors conclude that the dividend policy can be related to a clientele effect: more productive firms pay higher dividends and attract more institutional investors, who are more likely than dispersed retail investors to monitor management.

The models presented above share the shortcoming identified in Bhattacharya (1979), whereby dynamics reduce to a sequence of decisions in what are essentially one-shot games. Kaya (2009) proposes a general model for repeated signalling, using state variables that can be identified with reputation, and a recursive formulation. The salient characteristics of this model are the persistent use of an informed player and the observability of the decision history. As pointed out by the author, this general model can be used to represent the dynamic dividend policy problem. The main features of the dividend signalling models reviewed in this section are summarized in Table 19.6.

Table 19.6 Dividend signalling papers

Paper	Features	Results
Bhattacharya (1979)	Aligned interests	Optimal dividend policy
Miller and Rock (1985)	Passive investor	Underinvestment
Guttman et al. (2010)	Equilibrium selection	Dividend policy depends on earnings
Baker et al. (2015)	Multiperiod	Past dividends provide reference
Kumar (1988)	Asymmetric risk preferences	Dividend smoothing
Kumar and Lee (2001)	Multiperiod	Dividend smoothing
Allen et al. (2000)	Multiple asymmetric investors, monitoring	Clientele effect
Kaya (2009)	Dynamic state, decision history	General repeated signalling model

4.3 Investment Policy

One of the important issues of corporate governance is the agency problems that arise between equityholders (and/or management) and debtholders in the selection of risky projects. *Asset substitution* refers to the incentive of equityholders of a levered firm to increase the risk level of the firm's investments due to their limited liability. *Underinvestment* arises when equityholders do not undertake some profitable projects because these would mostly benefit debtholders, due to the priority of their claims in the case of bankruptcy. This section presents a selection of dynamic game models focusing specifically on agency problems in corporate investment choices.

John and Nachman (1985) proposes a two-period game between a manager and bondholders and shows that the underinvestment problem is attenuated when a sequential game of imperfect information is played. In the model considered by the authors, underinvestment occurs in the static case because the manager has no incentive to invest in profitable projects having a net present value that is lower than the promised debt payments. In a dynamic setting, it is assumed that the profitability of projects undertaken in the second period is correlated with the realization in the first period, which is the private information of the manager. The investment, financing, and dividend decisions made by the manager are used by bondholders as a signal of the firm's future prospects, which determines the amount allocated to debt payments. The authors interpret the equilibrium strategy as an endogenous reputation effect, giving managers an incentive to invest more when they expect the profitability of future projects to be high.

Jørgensen et al. (1989) proposes an open-loop Stackelberg differential game over a finite horizon, between a manager, who acts as a leader, and two shareholders, who, as followers, play a Nash game. The manager decides on investment and financing, while the majority shareholder decides on the dividend rate. The model is used to study the impact of separating ownership and management on corporate decisions and firm dynamics.

Hirshleifer and Thakor (1992) uses a two-period game to address agency problems in investment policies and their impact on capital structure. The model involves two types of managers (good or bad) and two types of projects (good but risky or bad). Good managers can obtain better results from good risky projects. The manager knows her type and chooses the project, while investors determine the stock price. The two-period setup enables an analysis of the impact of managerial reputation building on investment choices. The authors compare the two contrasting cases of a firm with and without outstanding debt and study the impact of possible takeovers. They find that, in an unlevered firm, reputation building can cause excessive conservatism in terms of the investment policy, while the reverse is obtained in the case of a levered firm, leading to an increase in the debt-equity ratio.

Heinkel and Zechner (1990) analyzes the relationship between capital structure and investment incentives. The game involves equityholders and potential investors who have asymmetrical information about the quality of an investment project, and it is played in two periods. At $t = 0$, equityholders choose the capital structure. At $t = 1$, a risky investment opportunity arises, for which cash flows differ in a good and a bad state. Equityholders privately observe the probability of a good state and decide to invest, which requires additional financing from the market, or not. At $t = 2$, cash flows are distributed and the game ends. The authors show that an all-equity firm will overinvest, while issuing debt causes underinvestment. They argue that issuing the "right" mix of securities avoids adverse incentives.

Zwiebel (1996) also links capital structure to corporate agency conflicts. The model is a finite-horizon multiperiod game between the manager and the equityholders of a firm. In each period, there is an investment opportunity that can be either good or bad, with a known constant probability, identified as the manager's type. The equityholders know the manager's type, but the quality of the investment opportunity is only known by the manager. In each period, the manager decides on capital structure (i.e., the debt level) and dividend payout. The equityholders can decide to take control of the firm after observing the manager's decision. If still in control, the manager then decides whether or not to invest in the project, after observing its quality. Bankruptcy occurs if the firm is not able to service the debt. Otherwise, the game moves to the next period. The manager's objective is to stay in control and undertake projects, while the equityholders care about the firm's equity value. The game is solved by backward recursion. The author argues that capital structure and dividend policy are linked to outside investment opportunities, so that various equilibrium strategies are obtained depending on the manager's type.

Almazan and Suarez (2003) proposes a hierarchical game between small shareholders and a manager who controls both investment and financing decisions. The model involves an investment project that can be good or bad, with a known probability. The manager chooses the type of financing (bank or public) and her level of effort (high, low) after privately observing the type of investment project. The probability of success depends on both decisions, and it is assumed that bank monitoring induces a high level of effort from managers. The equityholders choose the contract offered to the manager, taking her reaction into account. The authors solve the game to obtain the optimal incentive compensation contract and find that, when the probability of a project being good is sufficiently high, managers will choose bank monitoring for these good projects.

François et al. (2011) analyzes the role of convertible debt in mitigating the asset substitution problem. The model assumes the occurrence of an asset substitution possibility at some random date posterior to the determination of the capital structure. A dynamic game is played between holders of convertible bonds, who can decide to exercise their conversion option, and equityholders, who can decide to increase asset risk. Two possible equilibrium solutions are identified, according to the identity of the player who has first-mover advantage. The authors conclude that, in a multiperiod setting, the issuance of convertible debt does not eliminate the possibility of asset substitution.

Table 19.7 summarizes the features of papers reviewed in this section and shows that, because of the dynamic agency problems between debt and equity holders, corporate investment decisions are often closely related to capital structure decisions and to dividend policies.

Table 19.7 Investment policy papers

Paper	Decisions	Features	Results
John and Nachman (1985)	I, C, Q	Reputation building	Signal impacts on the market value of bonds
Jørgensen et al. (1989)	I, C, Q	Majority shareholder chooses Q	Impact of separating ownership and management
Hirshleifer and Thakor (1992)	I	Reputation building	Impact of leverage on investment conservatism
Heinkel and Zechner (1990)	I, C	–	Impact of leverage on investment conservatism
Zwiebel (1996)	I, C, Q, X	Equityholders can take control	Impact of agent's type on capital structure and dividends
Almazan and Suarez (2003)	E, C	Equityholders choose the contract	Optimal compensation contract
François et al. (2011)	I, C	I occurs after contract design	Convertible debt does not eliminate asset substitution

4.4 Interdependency of Corporate Financial Decisions

A very large body of literature is concerned with corporate finance decisions, and various theories have been proposed to explain payout policies, capital structure, and the investment decisions made by firms. Game-theoretic models arise naturally whenever ownership and control are in separate hands and usually involve asymmetric information, moral hazard, or imperfectly aligned utility functions.

The models presented in this section show that a wide variety of dynamic game models have been used to explain corporate finance decisions, ranging from simple hierarchical models involving active and passive players to stochastic dynamic and differential games. Most of these models focus on a few aspects or features to explain empirical observations of corporate behavior. Corporate financial decisions are, however, closely interrelated, and comprehensive dynamic game models should continue to be of considerable interest to corporate finance theory.

In a recent paper, Lambrecht and Myers (2016) proposes a general framework to analyze the dynamics of investment, borrowing, and payout decisions by public corporations in the context of agency problems in corporate finance (tax and other financial distortions are not considered). The model is a repeated game between a manager and dispersed shareholders, where the manager maximizes the expected net present value of her future rents, while shareholders maximize the market value of the firm, and the state variable is the (stochastic) periodic operating income. In each period, the manager proposes a payout to the shareholders and retains the remaining part of the operating income as rent. Shareholders can either accept this proposal or reject it and take over the firm. The cost of intervention by the shareholders is interpreted as a measure of the effectiveness of corporate governance. This general model is used to analyze contrasting cases where managers are risk neutral or risk averse, and where shareholders have full or partial information on the realized cash flows, and to address several problems in corporate finance, including capital structure and debt policy, investment and abandonment decisions, dividend smoothing, and takeovers and other external governance mechanisms.

5 Conclusion

This chapter presents a selection of models illustrating the contribution of dynamic game theory to investment and corporate finance. The topics covered are restricted to pricing, bankruptcy, and corporate decision issues, where dynamic game theory plays a prominent role. The examination of the literature covered in this chapter shows that dynamic game theory has had a meaningful contribution to the appreciation of issues and puzzles in the area of investment and corporate finance over the last forty years. It is impossible not to notice the predominance of the financial distress event as an ingredient of the various game models used in this literature. The recent years have seen the emergence of default risk-transferring instruments, such as *credit default swaps,* for instance, whereby the player affected by a default event is not necessarily the one negotiating the debt contract. Many of the classical themes covered in this chapter could be revisited in a game-theoretic framework in order to assess the impact of the availability of these risk-transferring instruments. Additional corporate finance decisions involving many players, such as mergers and acquisitions and initial public offerings, could be analyzed in a dynamic framework.

Other areas in finance can also benefit from the contribution of dynamic game models, such as intermediation, financial market microstructure, and corporate real option theory, to name only a few. Intermediation models capture the relationship between financial intermediaries and their clients, for instance, between banks and corporations or private investors or among syndicates of lenders. Financial microstructure models aim at explaining the dynamics of market price formation from the individual actions of informed and uninformed investors. Finally, real option models apply to the investment, divestment, and abandonment decisions of interacting corporate players. Dynamic game models in these areas are less conspicuous, and these represent interesting avenues that may gain importance in the future.

References

Acharya V, Huang JZ, Subrahmanyam M, Sundaram RK (2006) When does strategic debt-service matter? Econ Theory 29(2):363–378

Allen F, Morris S (2014) Game theory models in finance. In: Chatterjee K, Samuelson WF (eds) Game theory and business applications. Springer U.S. international series in operations research & management science, vol 194, 2nd edn. Kluwer Academic Publishers, Dortrecht, pp 17–41

Allen F, Bernardo AE, Welch I (2000) A theory of dividends based on tax clienteles. J Finance 55(6):2499–2536

Almazan A, Suarez J (2003) Managerial compensation and the market reaction to bank loans. Rev Financ Stud 16(1):237–261

Anderson RW, Nyborg KG (2011) Financing and corporate growth under repeated moral hazard. J Financ Intermed 20(1):1–24

Anderson RW, Sundaresan S (1996) Design and valuation of debt contracts. Rev Financ Stud 9(1):37–68

Anderson RW, Sundaresan S, Tychon P (1996) Strategic analysis of contingent claims. Eur Econ Rev 40(3):871–881

Annabi A, Breton M, François P (2012a) Game theoretic analysis of negotiations under bankruptcy. Eur J Oper Res 221(3):603–613

Annabi A, Breton M, François P (2012b) Resolution of financial distress under, Chapter 11. J Econ Dyn Control 36(12):1867–1887

Baker M, Mendel B, Wurgler J (2015) Dividends as reference points: a behavioral signaling approach. Rev Financ Stud 29(3):697–738

Ben-Ameur H, Breton M, Karoui L, L'Ecuyer P (2007) A dynamic programming approach for pricing options embedded in bonds. J Econ Dyn Control 31(7):2212–2233

Berglöf E, Von Thadden EL (1994) Short-term versus long-term interests: capital structure with multiple investors. Q J Econ 109(4):1055–1084

Berglöf E, Roland G, Von Thadden EL (2000) An incomplete contracts approach to corporate bankruptcy. Université de Lausanne, Cahiers de Recherches Economiques du DEEP, 00.12

Berkovitch E, Israel R (1998) The bankruptcy decision and debt contract renegotiations. Eur Finance Rev 2(1):1–27

Bhattacharya S (1979) Imperfect information, dividend policy, and "the bird in the hand" fallacy. Bell J Econ 10(1):259–270

Bielecki TR, Crépey S, Jeanblanc M, Rutkowski M (2008) Arbitrage pricing of defaultable game options with applications to convertible bonds. Quant Finance 8(8):795–810

Black F (1976) The dividend puzzle. J Portf Manag 2(2):5–8

Bolton P, Scharfstein DS (1996) Optimal debt structure and the number of creditors. J Polit Econ 104(1):1–25

Brennan MJ, Schwartz ES (1980) Analyzing convertible bonds. J Financ Quant Anal 15(04): 907–929

Broadie M, Chernov M, Sundaresan S (2007) Optimal debt and equity values in the presence of Chapter 7 and Chapter 11. J Finance 62(3):1341–1377

Brown DT (1989) Claimholder incentive conflicts in reorganization: the role of bankruptcy law. Rev Financ Stud 2(1):109–123

Bruche M (2011) Creditor coordination, liquidation timing, and debt valuation. J Financ Quant Anal 46(05):1407–1436

Bruche M, Naqvi H (2010) A structural model of debt pricing with creditor-determined liquidation. J Econ Dyn Control 34(5):951–967

Buehlmaier MM (2014) Debt, equity, and information. J Math Econ 50:54–62

Chen N, Dai M, Wan X (2013) A nonzero-sum game approach to convertible bonds: tax benefit, bankruptcy cost, and early/late calls. Math Finance 23(1):57–93

Christensen PO, Flor CR, Lando D, Miltersen KR (2014) Dynamic capital structure with callable debt and debt renegotiations. J Corp Finance 29:644–661

Constantinides GM (1984) Warrant exercise and bond conversion in competitive markets. J Financ Econ 13(3):371–397

Constantinides GM, Grundy BD (1989) Optimal investment with stock repurchase and financing as signals. Rev Financ Stud 2(4):445–465

Constantinides GM, Rosenthal RW (1984) Strategic analysis of the competitive exercise of certain financial options. J Econ Theory 32(1):128–138

d'Halluin Y, Forsyth PA, Vetzal KR, Labahn G (2001) A numerical PDE approach for pricing callable bonds. Appl Math Finance 8(1):49–77

Dynkin E (1969) Game variant of a problem on optimal stopping. Sov Math Dokl 10:270–274

Emanuel DC (1983) Warrant valuation and exercise strategy. J Financ Econ 12(2):211–235

Fan H, Sundaresan SM (2000) Debt valuation, renegotiation, and optimal dividend policy. Rev Financ Stud 13(4):1057–1099

Fluck Z (1998) Optimal financial contracting: debt versus outside equity. Rev Financ Stud 11(2):383–418

Fluck Z (1999) The dynamics of the management-shareholder conflict. Rev Financ Stud 12(2):379–404

François P, Morellec E (2004) Capital structure and asset prices: some effects of bankruptcy procedures. J Bus 77(2):387–411

François P, Hübner G, Papageorgiou N (2011) Strategic analysis of risk-shifting incentives with convertible debt. Q J Finance 1(02):293–321

Gennaioli N, Rossi S (2013) Contractual resolutions of financial distress. Rev Financ Stud 26(3):602–634

Gomes A (2000) Going public without governance: managerial reputation effects. J Finance 55(2):615–646

Guttman I, Kadan O, Kandel E (2010) Dividend stickiness and strategic pooling. Rev Financ Stud 23(12):4455–4495

Hamadène S, Zhang J (2010) The continuous time nonzero-sum Dynkin game problem and application in game options. SIAM J Control Optim 48(5):3659–3669

Hart O, Moore J (1989) Default and renegotiation: a dynamic model of debt. MIT Working Paper No. 520, revised as University of Edinburgh Discussion Paper

Hart O, Moore J (1997) Default and renegotiation: a dynamic model of debt (No. w5907). National Bureau of Economic Research

Hege U, Mella-Barral P (2005) Repeated dilution of diffusely held debt. J Bus 78(3):737–786
Heinkel R, Zechner J (1990) The role of debt and preferred stock as a solution to adverse investment incentives. J Financ Quant Anal 25(1):1–24
Hilli A, Laussel D, Van Long N (2013) Large shareholders, monitoring, and ownership dynamics: toward pure managerial firms? J Econ Dyn Control 37(3):666–679
Hirshleifer D, Thakor AV (1992) Managerial conservatism, project choice, and debt. Rev Financ Stud 5(3):437–470
Hull J, White A (1990) Pricing interest-rate-derivative securities. Rev Financ Stud 3(4):573–592
Hvide HK, Leite TE (2010) Optimal debt contracts under costly enforcement. Econ Theory 44(1):149–165
Jalan P, Barone-Adesi G (1995) Equity financing and corporate convertible bond policy. J Bank Finance 19(2):187–206
John K, Nachman DC (1985) Risky debt, investment incentives, and reputation in a sequential equilibrium. J Finance 40(3):863–878
Jørgensen S, Kort PM, Geert-Jan CT, Schijndel V (1989) Optimal investment, financing, and dividends: a Stackelberg differential game. J Econ Dyn Control 13(3):339–377
Kale JR, Noe TH (1990) Risky debt maturity choice in a sequential game equilibrium. J Financ Res 13(2):155–166
Kaya A (2009) Repeated signaling games. Games Econ Behav 66(2):841–854
Kifer Y (2000) Game options. Finance Stochast 4(4):443–463
Koziol C (2006) Optimal exercise strategies for corporate warrants. Quant Finance 6(1):37–54
Krasa S, Villamil AP (2000) Optimal contracts when enforcement is a decision variable. Econometrica 68(1):119–134
Krasa S, Sharma T, Villamil AP (2008) Bankruptcy and firm finance. Econ Theory 36(2):239–266
Kühn C (2004) Game contingent claims in complete and incomplete markets. J Math Econ 40(8):889–902
Kumar P (1988) Shareholder-manager conflict and the information content of dividends. Rev Financ Stud 1(2):111–136
Kumar P, Lee BS (2001) Discrete dividend policy with permanent earnings. Financ Manag 30:55–76
Lambrecht BM, Myers SC (2016) Agency dynamics in corporate finance. Ann Rev Financ Econ 8:53–80
Leland HE (1994) Corporate debt value, bond covenants, and optimal capital structure. J Finance 49(4):1213–1252
Leland HE, Pyle DH (1977) Informational asymmetries, financial structure, and financial intermediation. J Finance 32(2):371–387
Leland HE, Toft KB (1996) Optimal capital structure, endogenous bankruptcy, and the term structure of credit spreads. J Finance 51(3):987–1019
McConnell JJ, Schwartz ES (1986) LYON taming. J Finance 41(3):561–577
Mella-Barral P (1999) The dynamics of default and debt reorganization. Rev Financ Stud 12(3):535–578
Mella-Barral P, Perraudin W (1997) Strategic debt service. J Finance 52(2):531–556
Meneghetti C (2012) Managerial incentives and the choice between public and bank debt. J Corp Finance 18(1):65–91
Merton RC (1974) On the pricing of corporate debt: the risk structure of interest rates. J Finance 29(2):449–470
Miller MH, Rock K (1985) Dividend policy under asymmetric information. J Finance 40(4): 1031–1051
Moraux F, Silaghi F (2014) Inside debt renegotiation: optimal debt reduction, timing, and the number of rounds. J Corp Finance 27:269–295
Myers SC, Majluf NS (1984) Corporate financing and investment decisions when firms have information that investors do not have. J Financ Econ 13(2):187–221
Noe TH (1988) Capital structure and signaling game equilibria. Rev Financ Stud 1(4):331–355
Noe TH, Wang J (2000) Strategic debt restructuring. Rev Financ Stud 13(4):985–1015

Park C (2000) Monitoring and structure of debt contracts. J Finance 55(5):2157–2195
Repullo R, Suarez J (1998) Monitoring, liquidation, and security design. Rev Financ Stud 11(1):163–187
Ross SA (1977) The determination of financial structure: the incentive-signalling approach. Bell J Econ 8(1):23–40
Sîrbu M, Shreve SE (2006) A two-person game for pricing convertible bonds. SIAM J Control Optim 45(4):1508–1539
Thakor AV (1991) Game theory in finance. Financ Manag 20(1):71–94
Thomson W (2003) Axiomatic and game-theoretic analysis of bankruptcy and taxation problems: a survey. Math Soc Sci 45(3):249–297
Thomson W (2015) Axiomatic and game-theoretic analysis of bankruptcy and taxation problems: an update. Math Soc Sci 74:41–59
von Thadden EL, Berglöf E, Roland G (2010) The design of corporate debt structure and bankruptcy. Rev Financ Stud 23(7):2648–2679
Wang H (2011) Managerial entrenchment, equity payout and capital structure. J Bank Finance 35(1):36–50
Yagi K, Sawaki K (2010) The pricing and optimal strategies of callable warrants. Eur J Oper Res 206(1):123–130
Zwiebel J (1996) Dynamic capital structure under managerial entrenchment. Am Econ Rev 86(5):1197–1215

Marketing

20

Steffen Jørgensen

Contents

S. Jørgensen (✉)
Department of Business and Economics, University of Southern Denmark, Odense, Denmark
e-mail: steffenjorgensen2510@gmail.com

T. Başar, G. Zaccour (eds.), *Handbook of Dynamic Game Theory*,
https://doi.org/10.1007/978-3-319-44374-4_22

Abstract

Marketing is a functional area within a business firm. It includes all the activities that the firm has at its disposal to sell products or services to other firms (wholesalers, retailers) or directly to the final consumers. A marketing manager needs to decide strategies for pricing (toward consumers and middlemen), consumer promotions (discounts, coupons, in-store displays), retail promotions (trade deals), support of retailer activities (advertising allowances), advertising (television, internet, cinemas, newspapers), personal selling efforts, product strategy (quality, brand name), and distribution channels. Our objective is to demonstrate that differential games have proved to be useful for the study of a variety of problems in marketing, recognizing that most marketing decision problems are *dynamic* and involve *strategic* considerations. Marketing activities have impacts not only now but also in the future; they affect the sales and profits of competitors and are carried out in environments that change.

Keywords

Marketing · Differential games · Advertising · Goodwill · Pricing · Demand learning · Cost learning · Marketing channels · Channel coordination · Leadership

1 Introduction

Marketing is a functional area like, e.g., production and finance, in a business firm. It includes the activities that the firm has at its disposal to sell products or services to other firms (wholesalers, retailers) or directly to the final consumers. The main purpose of devoting resources to marketing is to increase sales in order to generate more revenues. The toolbox of a marketing manager includes a variety of activities: pricing (consumers and middlemen), consumer promotions (discounts, coupons, in-store displays), retailer promotions (trade deals), support of retailer activities (support of retailers' advertising expenditures), advertising (television, internet, cinemas, newspapers, outdoor displays), personal selling efforts, product strategy (quality, brand name), and distribution channels.

The objective of the chapter is to demonstrate that *differential games* have proved to be a useful methodology with which one can study decision problems in marketing in a formal way. Such problems are *dynamic* and involve *strategic considerations*. Marketing activities have impacts now and in the future; they affect sales, revenues, costs, and profits of competitors and are most often carried out in environments that change over time. It seems, therefore, to be less advisable to derive recommendations for a firm's marketing activities by using a static or a simple period-by-period framework or assuming a monopolistic market and disregard existing and/or potential competition.

1.1 A Brief Assessment of the Literature

The use of differential game theory to study marketing decision problems in a dynamic, competitive environment started in the 1970s. We will refer to seminal work in the areas that are covered and give the reader an account of how research has progressed. There is a literature using optimal control theory that started earlier than the dynamic game research stream. The control literature deals with marketing decision-making in dynamic, monopolistic markets. Not surprisingly, many of these works have formed the basis from which game theoretic studies emanated.

Dynamic game literature in marketing has predominantly been normative, focusing on the characterization of equilibrium marketing effort strategies. Not much emphasis has been put on empirical work. The main part of the literature aims at deriving predictions of what equilibrium behavior would be and to report these predictions as recommendations and decision support for marketing managers. Technically speaking, two options have been explored in order to derive game theoretic equilibria:

- *Analytical* methods can be used to obtain closed-form characterizations of equilibrium actions. Analytical methods may provide results of some generality when they leave parameter values and functional forms – at least partly – unspecified. A major problem here is that these methods normally work only in models of low complexity and/or one assumes a "mathematically convenient" structure. This puts a limit to the applicability of results.
- *Numerical* methods have their strength when models are complex. Their disadvantages are that results cannot be generalized beyond the scenarios considered in the numerical calculations, functional forms in a model must be specified, and one may be unable to obtain real-life data for parameter estimations. It is, however, worthwhile noticing that since the 1990s, a number of empirical studies in dynamic marketing competition have been published, particularly in the area of advertising. This research has typically used estimated differential game models to (a) test the empirical validity of the prescriptions of normative models or (b) to assess the advantages of using information about the evolution of the state (e.g., market shares or sales volumes) of a dynamic system.

There are not many textbooks dealing with marketing applications of differential games. Case (1979) was probably the first who considered many-firm advertising problems cast as differential games. The book by Jørgensen and Zaccour (2004) is devoted to differential games in marketing. Dockner et al. (2000) offered some examples from marketing. See also Haurie et al. (2012). The focus of these books is *noncooperative* dynamic games. The reader should be aware that in the last 10–15 years, there has been a growing interest in applying the theory of dynamic *cooperative* games to problems in marketing. Recent surveys of this area are Aust and Buscher (2014) and Jørgensen and Zaccour (2014).

At the beginning of the 1980s, most applications of differential games in marketing were in *advertising*. Jørgensen (1982a) summarized the early developments in this area. A few papers covered *pricing*, e.g., Jørgensen (1986a). Taking stock, Jørgensen and Zaccour (2004) noted that the study of competitive advertising and pricing strategies had been continued. Moreover, new areas of marketing have been approached using differential games. As we shall see, research activity has increased considerably in marketing channels/supply chains where researchers have approached problems of, e.g., coordinated (cooperative) decision-making and leadership.[1]

The following works, covering the period from 1982 to 2014, give an account of the current state of the art in dynamic games in marketing:

- *General*: Moorthy (1993), Eliashberg and Chatterjee (1985), Rao (1990), and Jørgensen and Zaccour (2004)
- *Pricing*: Jørgensen (1986a), Rao (1988), Kalish (1988), and Chatterjee (2009)
- *New-product diffusion models*: Dolan et al. (1986), Mahajan et al. (1990, 1993), Mahajan et al. (2000), and Chatterjee et al. (2000)
- *The production-marketing interface*: Eliashberg and Steinberg (1993), and Gaimon (1998)
- *Advertising*: Jørgensen (1982a), Erickson (1991, 1995a), Moorthy (1993), Feichtinger et al. (1994), Huang et al. (2012), Jørgensen and Zaccour (2014), and Aust and Buscher (2014)
- *Marketing channels*: He et al. (2007) and Sudhir and Datta (2009).

1.2 Notes for the Reader

1. An overwhelming majority of the literature in the area "dynamic games in marketing" has used *deterministic differential games*, and this chapter will focus on such games. A section deals with the (rather sparse) literature that has used *stochastic games* to study marketing problems.
2. There is a literature in economic applications of game theory/microeconomics/industrial organization that employs *repeated games* to study oligopolistic markets. Much work has been devoted to problems of intertemporal competition/collusion on prices, but also issues concerning, for instance, industrial structure (entry, exit, and number of firms) are treated.[2] The focus and the approach in this literature are in many respects different from that in literature on differential games in marketing. In the latter a main emphasis is on providing prescriptions for the decisions that individual firms should take in a market.

[1] Aust and Buscher (2014) present a figure showing the number of publications on cooperative advertising per year, from 1970 to 2013.

[2] Examples of textbooks that deal with economic applications of repeated games are Tirole (1988), Friedman (1991), Vives (1999), and Vega-Redondo (2003).

3. To benefit fully from the chapter, the reader should be familiar with the basics of nonzero-sum, noncooperative differential game theory.
4. Many of the models we present can be/have been stated for a market with an arbitrary number of firms. To simplify the presentation, we focus on markets having two firms only (i.e., duopolistic markets).
5. Our coverage of the literature is not intended to be exhaustive.
6. To save space and avoid redundancies, we only present – for a particular contribution – the *dynamics* (the differential equation(s)) being employed. The dynamics are a key element of any differential game model. Another element is the objective functionals. For a specific firm, say, firm i, the objective functional typically is of the form

$$J_i = \int_{t_0}^{T} \exp\{-\rho_i t\}\,[R_i(t) - C_i(t)]\,dt + \exp\{-\rho_i T\}\,\Phi_t(T)$$

   where $[t_0, T]$ is the planning period which can be finite or infinite. The term $R_i(t)$ is a revenue and $C_i(t)$ is a cost. Function Φ_t represents a salvage value to be earned at the horizon date T. If the planning horizon is infinite, a salvage value makes no sense and we have $\Phi_t = 0$. In games with a finite and short horizon, discounting may be omitted, i.e., $\rho_i = 0$.[3]

2 Advertising Games

2.1 Introduction

Advertising is an important element of a firm's communication mix. This mix also includes direct marketing (phone or e-mail), public relations, and personal selling. The planning of a firm's advertising activities involves many issues, and differential game studies have studied a subset of these only. What has gained most interest is the amount and timing of advertising expenditures, the sharing of the costs of advertising between a manufacturer and its retailers, and the interaction between advertising and other marketing activities (typically, pricing).

The literature has employed four main categories of models that are characterized by their dynamics (state equations in differential game terminology): *Lanchester, Vidale-Wolfe, new-product diffusion,* and *Nerlove-Arrow* models. The state variables are market shares or sales rates in the Lanchester and Vidale-Wolfe models, cumulative sales in new-product diffusion models, and advertising goodwill levels in Nerlove-Arrow models.

[3]Examples are short-term price promotions or advertising campaigns.

2.2 Lanchester Models

A century ago, Lanchester (1916) suggested a series of models to describe and analyze military combat. One of these models can readily be reformulated to become a simple and intuitive representation of the "battles for market share" that we observe in real-life markets (Kimball 1957).

To present the Lanchester model of advertising competition, let $X_i(t) \in [0,1]$ be market share at time t for a specific product, manufactured and sold by firm i. Denote by $a_i(t) \geq 0$ the advertising effort (or expenditure) rate of the firm. The dynamics are, in general,

$$\dot{X}_i(t) = f_i(a_i(t))X_j(t) - f_j(a_j(t))X_i(t);\ i,j \in \{1,2\}, i \neq j$$

where function $f(a)$, called the attraction rate, is increasing and takes positive values. The right-hand side of the dynamics shows that a firm's advertising has one aim only, i.e., to steal market share from the competitor. Because of this purpose, advertising effort a_i is referred to as *offensive* advertising. In the basic Lanchester advertising model, and in many differential games using this model, function $f_i(a_i)$ is linear. Some authors assume diminishing returns to advertising efforts by using for $f_i(a_i)$ the power function $\beta_i a_i^{\alpha_i}, \alpha_i \in (0,1]$ (e.g., Chintagunta and Jain 1995; Jarrar et al. 2004).

A seminal contribution is Case (1979) who studied a differential game with the basic Lanchester dynamics and found a feedback Nash equilibrium (FNE). Many extensions of the model have been suggested:

1. *Multiple types of advertising* (e.g., Erickson 1993)
2. *Industry sales expansion* (e.g., Bass et al. 2005a,b)
3. *Multiple advertising media* (e.g., Fruchter and Kalish 1998)
4. *Multiproduct* firms (e.g., Fruchter 2001)

We confine our interest to items 1 and 2. Erickson (1993) suggested that firms use two kinds of effort, offensive and *defensive* advertising, denoted $d_i(t)$. Defensive advertising aims at defending a firm's market share against attacks from the rival firm. Market share dynamics for firm i are

$$\dot{X}_i(t) = \beta_i \frac{a_i(t)}{d_j(t)} X_j(t) - \beta_j \frac{a_j(t)}{d_i(t)} X_i(t)$$

which shows that offensive efforts of firm i can be mitigated by defensive efforts of firm j. See also Martín-Herrán et al. (2012).

Industry sales are not necessarily constant (as they are in the basic Lanchester model). They might expand due to *generic* advertising, paid for by the firms in the industry. (Industry sales could also grow from other reasons, for instance, increased consumer incomes or changed consumer preferences.) Generic advertising promotes the *product category* (e.g., agricultural products), not the individual brands.

Technically, when industry sales expand, we replace market shares with sales rates in the Lanchester dynamics. Therefore, let $S_i(t)$ denote the sales rate of firm i. The following example presents two models that incorporate generic advertising.

Example 1. The dynamics used in Bass et al. (2005a,b) are

$$\dot{S}_i(t) = \beta_i a_i(t)\sqrt{S_j(t)} - \beta_j a_j(t)\sqrt{S_i(t)} + \theta_i\left[k_1 g_1(t) + k_2 g_2(t)\right]$$

where $g_i(t)$ is the rate of generic advertising, paid for by firm i.[4] The third term on the right-hand side is the share of increased industry sales that accrues to firm i. The authors identified FNE with finite as well as infinite time horizons. Jørgensen and Sigué (2015) included offensive, defensive, and generic advertising efforts in the dynamics:

$$\dot{S}_i(t) = \left[\beta a_i(t) - \lambda d_j(t)\right]\sqrt{S_j(t)} - \left[\beta a_j(t) - \lambda d_i(t)\right]\sqrt{S_i(t)} + \frac{k\left[g_i(t) + g_j(t)\right]}{2}$$

and identified FNE with a finite time horizon. Since players in this game are symmetric (same parameter values) it seems plausible to allocate one half to each player of the increase in industry sales.

Industry sales may decline as well. This happens if consumers stop using the product category as such, for example, because of technical obsolescence or changed preferences. (Examples of the latter are carbonated soft drinks and beer in Western countries.) A simple way to model decreasing industry sales is to include a decay term on the right-hand side of the dynamics.

Empirical studies of advertising competition with Lanchester dynamics appeared in Erickson (1985, 1992, 1996, 1997), Chintagunta and Vilcassim (1992, 1994), Mesak and Darrat (1993), Chintagunta and Jain (1995), Mesak and Calloway (1995), Fruchter and Kalish (1997, 1998), and Fruchter et al. (2001). Dynamics are estimated from empirical data and open-loop and feedback advertising trajectories determined. A comparison of the costs incurred by following each of these trajectories, using observed advertising expenditures, gives a clue of which type of strategy provides the better fit (Chintagunta and Vilcassim 1992). An alternative approach assumes that observed advertising expenditures and market shares result from decisions made by rational players in a game. The question then is whether prescriptions derived from the game are consistent with actual advertising behavior. Some researchers found that feedback strategies provide the better fit to

[4] Letting sales appear as $\sqrt{S}$ instead of linearly turns out to be mathematically expedient, cf. Sorger (1989).

observed advertising expenditures. This suggests that firms – not unlikely – act upon information conveyed by observed market shares when deciding their advertising expenditures.

Remark 1. The following dynamics are a pricing variation on the Lanchester model:

$$\dot{X}_i(t) = g_j(p_j(t))X_j(t) - g_i(p_i(t))X_i(t)$$

where $p_i(t)$ is the consumer price charged by firm i. Function $g_i(p_i)$ takes positive values and is convex increasing for $p_i > \bar{p}_i$. If $p_i \le \bar{p}_i$ (i.e., price p_i is "low"), then $g_i(p_i) = 0$ and firm i does not lose customers to firm j. If p_i exceeds $\bar{p}_i$, firm i starts losing customers to firm j. Note that it is a high price of firm i which drives away its customers. It is not a low price that attracts customers from the competitor. This hypothesis is the "opposite" of that in the Lanchester advertising model where advertising efforts of a firm attract customers from the rival firms: They do not affect the firm's own customers. See Feichtinger and Dockner (1985).

2.3 Vidale-Wolfe (V-W) Models

The model of Vidale and Wolfe (1957) describes the evolution of sales of a monopolistic firm. It was extended by Deal (1979) to a duopoly:

$$\dot{S}_i(t) = \beta_i a_i(t)[m - S_1(t) - S_2(t)] - \delta_i S_i(t)$$

in which m is a fixed market potential (upper limit of industry sales). The first term on the right-hand side reflects that advertising of firm i increases its sales by inducing some potential buyers, represented by the term $m - S_1(t) - S_2(t)$, to purchase its brand. The decay term $\delta_i S_i(t)$ models a loss of sales which occurs when some of the firm's customers stop buying its product. These customers leave the market for good, i.e., they do not switch to firm j. Deal characterized an open-loop Nash equilibrium (OLNE).

In the V-W model, sales of firm i are not directly affected by advertising efforts of firm j (in contrast to the Lanchester model). This suggests that a V-W model would be most appropriate in markets where firms – through advertising effort – can increase their sales by capturing a part of untapped market potential. It may happen that the market potential grows over time; this can be modeled by assuming a time-dependent market potential, $m(t)$, cf. Erickson (1995b).

Wang and Wu (2001) suggested a straightforward combination of the model of Deal (1979) and the basic Lanchester model.[5] Then the dynamics incorporate explicitly the rival firm's advertising effort. The authors made an empirical study

[5] An early contribution that combined elements of Lanchester and V-W models is Leitmann and Schmitendorf (1978); see also Jørgensen et al. (2010).

in which sales dynamics are estimated and compared to state trajectories generated by the Lanchester model. Data from the US cigarette and beer industries suggest that the Lanchester model may apply to a market which has reached a steady state. (Recall that the Lanchester model assumes a fixed market potential.) The combined model of Wang and Wu (2001) seems to be more appropriate during the transient market stages.

2.4 New-Product Diffusion Models

Models of this type describe the adoption process of a new durable product in a group of potential buyers. A good example is electronic products. The market potential may be fixed or could be influenced by marketing efforts and/or exogenous factors (e.g., growing consumer incomes). When the process starts out, the majority of potential consumers are unaware of the new product, but gradually awareness is created and some consumers purchase the product. These "early adopters" (or "innovators") are supposed to communicate their consumption experiences and recommendations to non-adopters, a phenomenon known as "*word of mouth*." With the rapidly increasing popularity of social media, word-of-mouth communication is likely to become an important driver of the adoption process of new products and services.

The seminal new-product diffusion model is Bass (1969) who was concerned primarily with modeling the adoption process. Suppose that the seller of a product is a monopolist and the number of potential buyers, m, is fixed. Each adopter buys one and only one unit of the product. Let $Y(t) \in [0, m]$ represent the *cumulative sales* volume by time t. Bass suggested the following dynamics for a new durable product, introduced at time $t = t_0$:

$$\dot{Y}(t) = \phi[m - Y(t)] + \eta \frac{Y(t)}{m}[m - Y(t)], \; Y(t_0) = 0.$$

The first term on the right-hand side represents adoptions made by "innovators." The second term reflects adoptions by "imitators" who communicate with innovators. The model features important dynamic demand effects such as *innovation*, *imitation*, and *saturation*. The latter means that the fixed market potential is gradually exhausted.

In the Bass model, the firm cannot influence the diffusion process as the model is purely descriptive. Later on, new-product diffusion models have assumed that the adoption process is influenced by one or more marketing instruments, typically advertising and price. For a monopolist firm, Horsky and Simon (1983) suggested that innovators are influenced by advertising:

$$\dot{Y}(t) = [\phi + \beta \ln a(t)]\,[m - Y(t)] + \eta \frac{Y(t)}{m}\,[m - Y(t)]$$

where the effect of advertising is represented by the term $\beta \ln a(t)$. Using the logarithm implies that advertising is subject to decreasing returns to scale.

Later on, the Bass model has been extended to multi-firm competition and studied in a series of papers using differential games. Early examples of this literature are Teng and Thompson (1983, 1985). Dockner and Jørgensen (1992) characterized an OLNE in a game with alternative specifications of the dynamics. One example is

$$\dot{Y}(t) = [\phi + \beta \ln a_i(t) + \mu Z(t)] [m - Z(t)]$$

where $Z(t) \triangleq Y_1(t) + Y_2(t)$ denotes industry cumulative sales. Equilibrium advertising rates decrease over time, a result which carries over from the monopoly case (cf. Horsky and Simon 1983). The reader should note that decreasing advertising rates often are driven by *(i)* the saturation effect – due to the fixed and finite market potential – and/or *(ii)* the absence of salvage values in the objective functionals.

New-product diffusion models have been studied with prices instead of advertising efforts as decision variables. An example is Breton et al. (1997) who studied a discrete-time dynamic game.

2.5 Advertising Goodwill Models

Most brands enjoy some goodwill among consumers – although some brands may suffer from negative goodwill (badwill). Goodwill can be created, for example, through good product or service quality, a fair price or brand advertising. Our focus here is on advertising.

The seminal work in the area is Nerlove and Arrow (1962) who studied a single firm's dynamic optimization problem. Goodwill is modeled in a simple way, by defining a state variable termed "advertising goodwill." This variable represents a stock and summarizes the effects of current and past advertising efforts. Goodwill, denoted $G(t)$, evolves over time according to the simple dynamics

$$\dot{G}(t) = \gamma a(t) - \delta G(t)$$

that have a straightforward interpretation. The first term on the right-hand side is the firm's gross investment in goodwill; the second term reflects decay of goodwill. Hence the left-hand side of the dynamics represents net investment in goodwill.

Note that omitting the decay term $\delta G(t)$ has an interesting strategic implication. When the firm has accumulated goodwill to a certain level, it is locked in and has two choices only: it can continue to build up goodwill or keep the goodwill level constant. The latter happens if the advertising rate is chosen as $a(t) = \delta G(t)/\gamma$, i.e., the advertising rate at any instant of time is proportional to the current goodwill level.

It is a simple matter to extend the Nerlove-Arrow (N-A) dynamics to an oligopolistic market. The goodwill stock of firm i, $G_i(t)$, then evolves according to the dynamics $\dot{G}_i(t) = a_i(t) - \delta G_i(t)$. Fershtman (1984) identified an OLNE where – not surprisingly – equilibrium advertising strategies mimic those in Nerlove and Arrow (1962).

The N-A model has been employed in various contexts. We present two illustrations.

Example 2. Chintagunta (1993) posed the following question: How sensitive are a firm's profits to deviations from equilibrium advertising strategies? After having characterized an OLNE, goodwill dynamics were estimated using data for a prescription drug manufactured by two firms in the US pharmaceutical industry. Then estimates of equilibrium advertising paths can be made. Numerical simulations suggested, for a quite wide range of parameter values, that profits are not particularly sensitive to deviations from equilibrium advertising levels.[6]

Example 3. Buratto and Zaccour (2009) analyzed a Stackelberg game played between a licensor (the leader) and a licensee (the follower), operating in the fashion industry. The dynamics are variations of the N-A model and a *time-consistent* open-loop Stackelberg equilibrium (OLSE) was characterized. Time-consistency means that the leader has no incentive to deviate from its announced policy: This makes the announcement credible.[7] Coordination of advertising efforts can be achieved if the licensor uses an *incentive strategy* which is designed in such a way that if the licensee sticks to its part of the cooperative agreement, the licensor will do the same (and vice versa).

As we shall see in Sect. 4, the N-A dynamics have been popular in differential games played by the firms forming a marketing channel (supply chain).

3 Pricing Games

3.1 Introduction

In the area of pricing, a variety of topics have been studied, for example, pricing of new products and the effects of cost experience and demand learning (word of mouth) on pricing strategies. In Sect. 4 we look at the determination of wholesale and retail prices in a marketing channel.

[6]This phenomenon, that a profit function is "flat," has been observed elsewhere. One example is Wilson's formula in inventory planning.

[7]Most often, an OLSE is not time-consistent and the leader's announcement is incredible. The problem can be fixed if the leader precommits to its announced strategy.

Cost experience (or cost learning) is a feature of a firm's production process that should be taken into consideration when a firm decides its pricing strategy. In particular, cost experience has been incorporated in models of new-product adoption processes. The meaning of cost experience is that the unit production cost of a product decreases with *cumulative* output, i.e., there is a learning-by-doing effect. This applies both to the production of a product and the provision of a service. Formally, let $Y_i(t)$ represent cumulative output by time t and $c_i(Y_i)$ be a function the values of which determine the unit production cost. Cost experience means that $c_i'(Y_i) < 0$, i.e., the unit cost decreases as cumulative output increases. This phenomenon has a long history in manufacturing and has been observed in many industries.

3.2 Demand Learning

The demand learning phenomenon (diffusion effect, word of mouth) was defined in Sect. 2. Let $p_i(t)$ be the consumer price charged by firm i. Dockner and Jørgensen (1988a) studied various specifications of sales rate dynamics, for example the following:

$$\dot{Y}_i(t) = \left[\alpha_i - \beta_i p_i(t) + \gamma \left(p_j(t) - p_i(t)\right)\right] f(Y_i(t) + Y_j(t)) \tag{20.1}$$

where $Y_i(t)$ denotes cumulative sales and f is a function taking positive values. The "price differential" term $p_j(t) - p_i(t)$ is somewhat extreme: if the price $p_j(t)$ is one cent higher than $p_i(t)$, sales of firm i increase.[8] Dockner and Jørgensen identified an OLNE. It was shown that if $f' < 0$, and under a certain assumption, equilibrium prices are *decreasing* for all t. A similar result was obtained by Kalish (1983) in the case of a monopoly which illustrates the lack of strategic interaction in an OLNE. Dockner and Gaunersdorfer (1996) used the dynamics in (20.1) with $f(Y_i(t) + Y_j(t)) = m - [Y_1(t) + Y_2(t)]$.

A strategy of decreasing prices is known as "skimming." The idea is to exploit the segment of customers who are first in the market because these customers often have the highest willingness to pay. Later on, as firms benefit from cost learning, prices can be lowered to reach new customer segments. A good example is the marketing of electronic goods and household appliances where we see that new products often are introduced at a high price.

Models that incorporate demand learning have been used in various contexts. Here are two examples:

Example 4. Minowa and Choi (1996) analyzed a model in which each firm sells two products, one primary and one secondary (or contingent) product. The latter is useful only if the consumer has the primary product. This is a multiproduct problem

[8] See also Eliashberg and Jeuland (1986).

with a diffusion process (including demand learning) of a contingent product that depends on the diffusion process of the primary product.

Example 5. Breton et al. (1997) studied a Stackelberg differential game. Let the leader and follower be represented by subscripts l and f, respectively, such that $i \in \{l, f\}$. The dynamics are (cf. Dockner and Jørgensen 1988a)

$$\dot{Y}_i(t) = \left[\alpha_i - \beta_i p_i(t) + \gamma(p_j(t) - p_i(t))\right][A_i + B_i Y_i(t)]\left[m - Y_l(t) - Y_f(t)\right].$$

Cost experience effects are included and the end result is that the model is quite complex. An FSE is not analytically tractable, but numerical simulations suggest that prices decline over time and the firm with the lowest cost charges the lowest price. The shape of price trajectories seems to be driven mainly by cost experience.

3.3 Government Subsidies of New Technologies

As is well known, governments spend money on a multitude of activities. In some cases, government funds are granted with the purpose of influencing the behavior of citizens in one way or another. A specific situation is where a government offers a *price subsidy* to consumers if they are willing to adopt a new technology. A good example is "green energy" such as natural gas, wind power, and solar energy. In Denmark, for instance, all three types of green energy are subsidized. There are two main reasons why a government wishes to accelerate the adoption of such technologies:

1. Increased usage of the technology is believed to benefit of the environment.
2. With increased usage, manufacturers of the technology should benefit from lower unit production costs, due to cost experience effects. This should, in turn induce manufacturers to lower their consumer prices.

Note that consumers stand to benefit twofold: They pay a lower price to the manufacturer and receive a government subsidy.

A seminal paper in the area is Kalish and Lilien (1983) who posed the government's problem as one of optimal control (manufacturers of the new technology are not decision-makers in a game). Building on this work, Zaccour (1996) considered an initial stage of the life cycle of a new technology where there still is only one firm in the market. Firm and government play a differential game in which government chooses a subsidy rate $s(t)$, while the firm decides a consumer price $p(t)$. Government commits to a subsidy scheme, specified by the time path $s(\cdot)$, and has allocated a fixed budget to the program. Its objective is to maximize the total number of units, denoted $Y(T)$, that have been sold by time T. The sales rate dynamics are

$$\dot{Y}(t) = \exp\{-\alpha(1 - s(t))p(t)\}\, h(Y(t))$$

in which h is a function taking positive values. If $h' < 0$ (i.e., there are saturation effects), the subsidy increases, while the consumer price decreases over time. Both are beneficial for the diffusion of the new technology. If $h' > 0$ (i.e., there are demand learning effects), the subsidy is decreasing. This makes sense because less support is needed to stimulate sales and hence production.

With a view to what happens in real life, it may be more plausible to assume that government is a Stackelberg leader who acts first and announces its subsidy program. Then the firm makes its decision. This was done in Dockner et al. (1996) using the sales dynamics $\dot{Y}(t) = f(p(t))[m - Y(t)]$ in which $f'(p) < 0$. In an FSE, the consumer price is nonincreasing over time if players are sufficiently farsighted, a result which is driven by the saturation effect.

The reader may have noticed some shortcomings of the above models. For example, they disregard that existing technologies are available to consumers, government has only one instrument (the subsidy) at its disposal, firm and government have same time horizon, and the government's budget is fixed. Attempts have been made to remedy these shortcomings. We give two illustrations:

Example 6. In Jørgensen and Zaccour (1999b), government is a Stackelberg leader who offers a price subsidy to consumers and commits to buying from the manufacturer a number of units of the new technology for its own use. The option of consumers buying an existing technology is also accounted for. It turns out, in equilibrium, that the subsidy and the consumer price decrease at the same rate. This implies that the net consumer price remains the same over time. The program thus is fair, in the sense that a customer pays the same net price no matter when the product is bought. Janssens and Zaccour (2014) assumed that the firm and government have different time horizons and the government's budget is seen as a flexible instrument (rather than a fixed constraint) with which government can attain a desired target level of the consumer price.

3.4 Entry of Competitors

In problems where an incumbent firm faces potential entry of competitors, the latter should be viewed as rational opponents and not as exogenous decision-makers who implement predetermined actions. A first attempt to do this in a dynamic setup was Lieber and Barnea (1977) in which an incumbent firm sets the price of its product while competitors invest in productive capacity. The authors viewed, however, the problem as an optimal control problem of the incumbent, supposing that potential entrants believe that the incumbent's price remains constant. Jørgensen (1982b) and Dockner (1985) solved the problem as a differential game, and Dockner and Jørgensen (1984) extended the analysis to Stackelberg and cooperative differential games. These analyses concluded that – in most cases – the incumbent should discourage entry by decreasing its price over time.

A more "intuitive" representation of the entry problem was suggested in Eliashberg and Jeuland (1986). A monopolist (firm 1) introduces a new durable product, anticipating the entry of a competitor (firm 2). The planning period of firm 1 is $[t_0, T_2]$ and that of firm 2 is $[T_1, T_2]$ where T_1 is the known date of entry. (An obvious modification here would be to suppose that T_1 is a random variable.) Sales dynamics in the two time periods are, indexing firms by $i = 1, 2$:

$$\dot{Y}_1(t) = [\alpha_1 - \beta_1 p_1(t)]\,[m - Y_1(t)]\,, \quad Y_1(t_0) = Y_2(t_0) = 0,\ t \in [t_0, T_1]$$
$$\dot{Y}_i(t) = \left[\alpha_i - \beta_i p_i(t) + \gamma(p_j(t) - p_i(t))\right][m - Z(t)]\,, \quad Y_2(T_1) = 0,\ t \in [T_1, T_2]$$

where $Z(t) \triangleq Y_1(t) + Y_2(t)$. Three types of incumbents are considered:

- The *non-myopic* monopolist correctly predicts entry at T_1 and has planning horizon $T_2 > T_1$.
- The *myopic* monopolist disregards the duopoly period and has planning horizon T_1.
- The *surprised* monopolist has planning horizon T_2 but does not foresee entry at $T_1 < T_2$.

It turns out that in an OLNE, the myopic monopolist sets a higher price than a non-myopic monopolist. This is the price paid for being myopic because a higher price decreases sales and hence leaves a higher untapped market open to the entrant. The surprised monopolist charges a price that lies between these two prices.

The next two papers consider entry problems where firms control both advertising efforts and consumer prices. Fershtman et al. (1990) used the N-A dynamics and assumed that one firm has an advantage from being the first in a market. Later on, a competitor enters but starts out with a higher production cost. However, production costs of the two firms eventually will be the same. Steady-state market shares were shown to be affected by the order of entry, the size of the initial cost advantage, and the duration of the monopoly period. Chintagunta et al. (1993) demonstrated that in an OLNE, the relationship between advertising and price can be expressed as a generalization of the "Dorfman-Steiner formula" (see, e.g., Jørgensen and Zaccour 2004, pp. 137–138). Prices do not appear in N-A dynamics which make their role less significant.

Remark 2. Xie and Sirbu (1995) studied entry in a type of market not considered elsewhere. The market is one in which there are positive network externalities which means that a consumer's utility of using a product increases with the number of other users. Such externalities typically occur in telecommunication and internet services. To model sales dynamics, a diffusion model was used, assuming that market potential is a function of consumer prices as well as the size of the installed bases of the duopolists' products. The authors asked questions like: Will network externalities give market power to the incumbent firm? Should an incumbent permit or prevent compatibility when a competitor enters? How should firms set their prices?

3.5 Other Models of Pricing

Some models assume that both *price* and *advertising* are control variables of a firm. A seminal paper, Teng and Thompson (1985), considered a game with new-product diffusion dynamics. For technical reasons, the authors focused on price leadership which means that there is only one price in the market, determined by the price leader (typically, the largest firm). OLNE price and advertising strategies were characterized by numerical methods. See also Fershtman et al. (1990) and Chintagunta et al. (1993).

Krishnamoorthy et al. (2010) studied a durable-product oligopoly. Dynamics were of the V-W type, without a decay term:

$$\dot{Y}_i(t) = \gamma_i a_i(t) D_i(p_i(t)) \sqrt{m - (Y_1(t) + Y_2(t))} \tag{20.2}$$

where the demand function D_i is linear or isoelastic. FNE advertising and pricing strategies were determined. The latter turn out to be constant over time. In Helmes and Schlosser (2015), firms operating in a durable-product market decide on prices and advertising efforts. The dynamics are closely related to those in Krishnamoorthy et al. (2010). Helmes and Schlosser specified the demand function as $D_i(p_i) = p_i^{-\varepsilon_i}$ where ε_i is the price elasticity of product i. The authors looked for an FNE, replacing the square root term in the dynamics in (20.2) by general functional forms.

One of the rare applications of *stochastic* differential games is Chintagunta and Rao (1996) who supposed that consumers derive value (or utility) from the consumption of a particular brand. Value is determined by (a function of) price, the aggregate level of consumer preference for the brand, and a random component. The authors looked for an OLNE and focused on steady-state equilibrium prices. Given identical production costs, a brand with high steady-state consumer preferences should set a high price which is quite intuitive. The authors used data from two brands of yogurt in the USA to estimate model parameters. Then steady-state prices can be calculated and compared to actual (average) prices, to see how much firms "deviated" from steady-state equilibrium prices prescribed by the model.

Gallego and Hu (2014) studied a two-firm game in which each firm has a fixed initial stock of perishable goods. Firms compete on prices. Letting $I_i(t)$ denote the inventory of firm i, the dynamics are

$$\dot{I}_i(t) = -d_i(t, p(t)), \quad I_i(t_0) = C$$

where d_i is a demand function, $p(t)$ the vector of prices, and $C > 0$ a fixed initial inventory. OLNE and FNE strategies were identified. Although this game is deterministic, it may shed light on situations where demand is random. For example, given that demand and supply are sufficiently large, equilibria of the deterministic game can be used to construct heuristic policies that would be asymptotic equilibria in the stochastic counterpart of the model.

4 Marketing Channels

4.1 Introduction

A marketing channel (or supply chain) is a system formed by firms that most often are independent businesses: suppliers, a manufacturer, wholesalers, and retailers. The last decades have seen a considerable interest in supply chain management that advocates an integrated view of the system. Major issues in a supply chain are how to increase efficiency by coordinating decision-making (e.g., on prices, advertising, and inventories) and information sharing (e.g., on consumer demand and inventories). These ideas are not new: The study of supply chains has a long tradition in marketing literature.

Most differential game studies of marketing channels assume a simple structure having one manufacturer and one retailer, although extensions have been made to account for vertical and horizontal competition or cooperation. Examples are a channel consisting of a single manufacturer and multiple (typically competing) retailers or competition between two channels, viewed as separate entities.

Decision-making in a channel may be classified as *uncoordinated* (decentralized, noncooperative) or *coordinated* (centralized, cooperative). Coordination means that channel members make decisions (e.g., on prices, advertising, inventories) that will be in the best interest for the channel as an entity.[9] There are various degrees of coordination, ranging from coordinating on a single decision variable (e.g., advertising) to full coordination of all decisions. The reader should be aware that formal – or informal – agreements to cooperate by making joint decisions (in fact, creating a cartel) are illegal in many countries. More recent literature in the area has studied if coordination can be achieved without making such agreements.

A seminal paper by Spengler (1950) identified "*the double marginalization problem*" which illustrates lack of coordination. Suppose that a manufacturer sets the transfer (wholesale) price by adding a profit margin to the unit production cost. Subsequently, a retailer sets the consumer price by adding a margin to the transfer price. It is readily shown that the consumer price will be higher and hence consumer demand lower, than if channel members had coordinated their price decisions. The reason is that in the latter case, one margin only would be added.

In the sequel, unless explicitly stated, state equations are the N-A advertising goodwill dynamics (or straightforward variations on that model). We consider the most popular setup, a one-manufacturer, one-retailer channel, unless otherwise indicated.

[9] It may happen that, say, a manufacturer has a strong bargaining position and can induce the other channel members to make decisions such that the payoff of the manufacturer (not that of the channel) is maximized.

4.2 Channel Coordination

4.2.1 Full Cooperation

Full cooperation is understood to mean that channel members wish to maximize channel profits. Firms remain independent, but act as one firm.[10] The literature on this issue mainly deals with situations in which channel members decide on advertising efforts, the wholesale, and the consumer prices. The problem is to induce channel members to make decisions that lead to maximization of the joint profits.

Models with Advertising and Pricing

Chintagunta and Jain (1992) considered a channel in which both manufacturer and retailer advertise. Firms play OLNE strategies if they do not coordinate their advertising efforts. It is shown that firms use more advertising effort if they cooperate and this generates higher channel profits. These results are intuitive and have often been noted in the literature. In Jørgensen and Zaccour (1999a) firms decide their respective advertising efforts. The consumer price is set by the retailer, the transfer price by the manufacturer. In an FNE, the consumer price is higher than the one which is charged under full cooperation. This is an illustration of the double marginalization phenomenon. It turns out that for all t, instantaneous channel profits under cooperation are higher than those without cooperation which gives an incentive to cooperate *at any time* during the play of the game.

In the pricing models that we have presented above, the consumer price has an *instantaneous* influence on the dynamics and the profit functionals. To allow for longer-term ("carry-over") effects of retail pricing, Zhang et al. (2012) included a consumer *reference price*, denoted $R(t)$.[11] The hypothesis is that the reference price can be (positively) affected by channel member advertising. The consumer price p is assumed constant. Reference price dynamics are given by

$$\dot{R}(t) = \beta(p - R(t)) + \mu_m A(t) + \mu_r a(t) \tag{20.3}$$

where $A(t)$ and $a(t)$ are advertising efforts of manufacturer and retailer, respectively. (N-A goodwill dynamics are also a part of the model.) The manufacturer is a Stackelberg leader who determines the rate of advertising support to be given to the retailer. The paper proposes a "nonstandard" coordination mechanism: Each channel member subsidizes advertising efforts of the other member.

In Martín-Herrán and Taboubi (2015) the consumer reference price is an exponentially weighted average of past values of the actual price $p(t)$, as in (20.3), but without the advertising terms on the right-hand side. Two scenarios were

[10] An extreme case is vertical integration where channel members merge and become one firm.

[11] The meaning of a reference price is the following. Suppose that the probability that a consumer chooses brand X instead of brand Y depends not only on the current prices of the brands but also on their *relative* values when compared to historic prices of the brands. These relative values, as perceived by the consumer, are called reference prices. See, e.g., Rao (2009).

treated: vertical integration (i.e., joint maximization) and a Stackelberg game with manufacturer as the leader. It turns out that – at least under certain circumstances – there exists an initial time period during which cooperation is not beneficial.

Models with Pricing Only

Chiang (2012) focused on a durable product. Retail sales dynamics are of the "new-product diffusion" variety, with a price-dependent market potential and saturation effects only:

$$\dot{Y}(t) = \alpha\,[m - p(t) - Y(t)]$$

where $Y(t)$ represents cumulative sales and $p(t)$ the consumer price, set by the retailer. The term $m - p(t)$ is the market potential being linearly decreasing in price. The manufacturer sets the wholesale price. OLNE, FNE, and *myopic* equilibria are compared to the full cooperation solution.[12] A main result is that both channel members are better off, at least in the long run, if they ignore the impact of current prices on future demand and focus on intermediate-term profits.

Zaccour (2008) asked the question whether a two-part tariff can coordinate a channel. A two-part tariff determines the wholesale price w as $w = c + k/Q(t)$ where c is the manufacturer's unit production cost and $Q(t)$ the quantity ordered by the retailer. The idea of the scheme clearly is to induce the retailer to order larger quantities since this will diminish the effect of the fixed ordering (processing) cost k. In a Stackelberg setup with manufacturer as leader, the answer to the question is negative: A two-part tariff will not enable the leader to impose the full cooperation solution even if she precommits to her part of that solution.

4.2.2 Partial Cooperation

In situations where full cooperation (joint profit maximization) is not an option, an obvious alternative is to cooperate partially. For such cooperation to be acceptable, it should at least be profit improving for both channel members. A pertinent research question then is whether there exist coordination schemes that fulfill this requirement. The answer is yes.

An example is *cooperative advertising* which is an arrangement where a manufacturer pays some (or all) of the costs incurred by the retailer who advertises locally to promote the manufacturer's product. In addition, the manufacturer may advertise nationally (at her own expense). The terms "local" and "national" refer to the type of media that are used for advertising. Local advertising most often has short-term, promotional purposes, while national advertising has a longer-term objective: to maintain or enhance brand loyalty.

[12]Myopic behavior means that a decision-maker disregards the dynamics when solving her dynamic optimization problem.

We give below some examples of problems that have been addressed in the area of cooperative advertising. In the first example, the manufacturer has the option of supporting the retailer by paying some (or all) of her advertising costs.

Example 7. In Jørgensen et al. (2000) each channel member can use long-term (typically, national) and/or short-term (typically, local) advertising efforts. The manufacturer is a Stackelberg leader who chooses the shares that she will pay of the retailer's costs of the two types of advertising. Four scenarios are studied: (i) no support at all, (ii) support of both types of retailer advertising, support of (iii) long-term advertising only, and (iv) support of short-term advertising only. It turned out that both firms prefer full support to partial support which, in turn, is preferred to no support.

Most of the literature assumes that any kind of advertising has a favorable impact on goodwill. The next example deals with a situation in which excessive retailer promotions may harm the brand goodwill. The question then is: Should a manufacturer support retailer advertising?

Example 8. Jørgensen et al. (2003) imagined that frequent retailer promotions might *damage* goodwill, for instance, if consumers believe that promotions are a cover-up for poor product quality. Given this, does it make sense that the manufacturer supports retailer promotion costs? An FNE is identified if no advertising support is provided, an FSE otherwise. An advertising support program is favorable in terms of profits if (a) the initial goodwill level is "low" or (b) if this level is at an "intermediate" level and promotions are not "too damaging" to goodwill.

The last example extends our two-firm channel setup to a setting with two independent and competing retailers.

Example 9. Chutani and Sethi (2012a) assumed that a manufacturer is a Stackelberg leader who sets the transfer price and supports retailers' local advertising. Retailers determine their respective consumer prices and advertising efforts as a Nash equilibrium outcome in a two-player game. It turns out that a cooperative advertising program, with certain exceptions, does not benefit the channel as such. Chutani and Sethi (2012b) studied the same problem as the (2012a) paper. Dynamics now are a nonlinear variant of those in Deal (1979), and the manufacturer has the option of offering different support rates to retailers.[13]

[13] In practice it may be illegal to discriminate retailers, for example, by offering different advertising allowances or quoting different wholesale prices.

4.3 Channel Leadership

Most of the older literature assumed that the manufacturer is the channel leader but we note that the last decades have seen the emergence of huge retail chains having substantial bargaining power. Such retail businesses may very well serve as channel leaders.

Leadership of a marketing channel could act as a coordinating device when a leader takes actions that induce the follower to decide in the best interest of the channel.[14] A leader can emerge from exogenous reasons, typically because it is the largest firm in the channel. It may also happen that a firm can gain leadership endogenously, for the simple reason that such an outcome is desired by all channel members. In situations like this, some questions arise: Given that leadership is profitable at all, who will emerge as the leader? Who stands to benefit the most from leadership? The following example addresses these questions.

Example 10. Jørgensen et al. (2001) studied a differential game with pricing and advertising. The benchmark outcome, which obtains if no leader can be found, is an FNE. Two leadership games were studied: one having the manufacturer as leader and another with the retailer as leader. These games are played a la Stackelberg and in each of them an FSE was identified. The manufacturer controls his profit margin and national advertising efforts, while the retailer controls her margin and local advertising efforts. In the retailer leadership game, the retailer's margin turns out be constant (and hence its announcement is credible). The manufacturer has the lowest margin which might support a conjecture that a leader secures itself the higher margin. In the manufacturer leadership game, the manufacturer does *not* have the higher margin. Taking channel profits, consumer welfare (measured by the level of the retail price), and individual profits into consideration, the answers to the questions above are – in this particular model – that (i) the channel should have a leader and (ii) the leader should be the manufacturer.

4.4 Incentive Strategies

Incentive strategies can be implemented in games with a leader as well as in games without a leader. The idea probably originated in Stackelberg games where, if a leader provides the right incentive to the follower, the latter can be induced to behave in accordance with the *leader's* objective. This illustrates the old adage: *'If you wish other people to behave in your own interest, then make them see things your way'* (Başar and Olsder 1995, p. 396). First we look at the original setup, i.e., games with a leader.

[14] In less altruistic environments, a leader may wish to induce the follower to make decisions that are in the best interest of the leader.

4.4.1 Games with Leadership

Suppose that the manufacturer is the channel leader who wishes to induce the retailer to act, not in the best interest of the leader, but in the best interest of the channel. Then the retailer should be induced to choose the decisions prescribed for her by a cooperative solution. We present two examples that illustrate how this can be done.

Example 11. Jørgensen and Zaccour (2003a) considered a game where aggregate retailer promotions have negative effects on goodwill, cf. Jørgensen et al. (2003), but in contrast to the latter reference, retailer promotions now have carry-over effects. Retailers – who are symmetric – determine local promotional efforts, while the manufacturer determines national advertising, denoted $A(t)$. The manufacturer wishes to induce retailers to make promotional decisions according to the fully cooperative solution and announces the incentive strategy

$$a(P)(t) = A_c(t) + \phi(t)\left[P(t) - P_c(t)\right] \tag{20.4}$$

where $P(t)$ is aggregate retailer promotional effort and the subscript c refers to cooperation. The strategy in (20.4) works as follows: if retailers do as they are supposed to do, then we shall have $P(t) = P_c(t)$ and the manufacturer will choose $A(t) = A_c(t)$. Technically, to find an equilibrium of the incentive game, retailers determine their optimal promotion rates under the belief that the manufacturer will act in accordance with (20.4). This provides $P(\phi(t))$ and the manufacturer selects the incentive coefficient $\phi(t)$ such that the optimal retailer promotion rate equals the desired one, i.e., $P_c(t)$.

Example 12. In Jørgensen et al. (2006) a manufacturer is a Stackelberg leader who wishes to induce a retailer to increase her local advertising. This can be accomplished by implementing a promotion allowance scheme such that the manufacturer pays an amount, denoted $D(P)$, per unit of retailer promotional effort $P(t)$. The incentive strategy is

$$D(P)(t) = \theta P(t)$$

where θ is a constant to be determined by the manufacturer before the start of the game. Two scenarios were considered. In the first, the manufacturer wishes the retailer to behave in accordance with the full cooperation outcome. In the second, the manufacturer is selfish and selects an incentive that will make the retailer maximize the manufacturer's profits.

4.4.2 Games Without Leadership

Even if there is no leader of the channel, incentive strategies may still be implemented. The idea is as follows. Suppose that channel members agree upon a desired outcome (e.g., the full cooperation solution) for the channel. Side payments (e.g., advertising allowances paid to the retailer) are infeasible. Nevertheless, the desired outcome could be realized if channel members play according to *equilibrium incentive strategies*. An incentive equilibrium has the following best-response property: If firm j sticks to its incentive strategy, the best choice of firm $i \neq j$ is to use its own incentive strategy. Unilateral deviations are not payoff improving and the desired outcome is in equilibrium.

The following example shows how the fully cooperative solution can be implemented as an incentive equilibrium. Other solutions than full cooperation can be implemented as equilibrium outcomes provided that channel members can agree that such a solution is the desired one.

Example 13. In Jørgensen and Zaccour (2003b) a manufacturer controls the transfer price $p_M(t)$ and national advertising effort $a_M(t)$, while the retailer controls the consumer price $p_R(t)$ and local advertising effort $a_R(t)$. As we have seen, an incentive strategy makes the manufacturer's decision dependent upon the retailer's decision and vice versa. Considering the advertising part, this can be done by using the strategies

$$\gamma_M(a_R)(t) = a_M^d(t) + \mu_M(t)\left[a_R(t) - a_R^d(t)\right]$$
$$\gamma_R(a_M)(t) = a_R^d(t) + \mu_R(t)\left[a_M(t) - a_M^d(t)\right]$$

where $\mu_M(t), \mu_R(t)$ are incentive parameters and the superscript d refers to the desired solution. To find the incentive parameters, one needs to solve *two* optimal control problems. (A similar procedure is used to determine the price incentive strategies.) A main message of the paper is that channel coordination can be achieved without a leader, using the above-incentive strategies. A related paper is De Giovanni et al. (2015) who considered a supply chain where manufacturer and retailer invest in a product recovery program.

4.5 Franchise Systems

A franchise system is a business model that may be thought of as a mixture of a centralized and a decentralized channel. Typically, the *franchisor* offers a trademark and supplies a standardized business format to be used by the *franchisees*. The latter are obliged to follow the business procedures outlined in their franchise contract. For instance, all franchisees must charge the same consumer price for a specific product. The consumer prices essentially are set by the franchisor.

We provide two examples of franchising, dealing with advertising and service provision in cooperative and noncooperative situations.

Example 14. Sigué (2002) studied a system with a franchisor and two franchisees. The former takes care of national advertising efforts, while franchisees are in charge of local advertising and service provision. Franchisees benefit from a high level of system goodwill, built up through national advertising and local service.[15] Two scenarios were examined. The first is a benchmark where a noncooperative game is played by all three players and an FNE is identified. In the second, franchisees coordinate their advertising and service decisions and act as one player (a coalition) in a game with the franchisor. Also here an FNE is identified. It turns out that cooperation among franchisees leads to higher local advertising efforts (which is expected). If franchisees do not cooperate, free riding will lead to an outcome in which less service than desirable is provided.

Example 15. Sigué and Chintagunta (2009) asked the following question: Who should do promotional and brand-image advertising, respectively, if franchisor and franchisees wish to maximize their individual profits? The problem was cast as a two-stage game. First, the franchisor selects one out of three different advertising arrangements, each characterized by its degree of centralization of advertising decisions. Second, given the franchisor's choice, franchisees decide if they should cooperate or not. Martín-Herrán et al. (2011) considered a similar setup and investigated, among other things, the effects of price competition among franchisees.

Jørgensen (2011) applied an important element of a franchise system, viz., the *contract*, to the standard two-member marketing channel (not a franchise system). A franchise contract includes as special cases the classic wholesale-price contract, a two-part tariff (see above), and a revenue-sharing contract.[16] The retailer determines an ordering rate and the retail price, while the manufacturer decides a production rate and the parameters of the contract that is offered to the retailer. Contract parameters are time dependent and the contract period is finite. Jørgensen characterized the fully cooperative solution and the selfish one in which the manufacturer wishes to maximize her own profits.

4.6 National Brands, Store Brands, and Shelf Space

Competition among national and store brands, the latter also known as private labels, is becoming more widespread. A store brand carries a name chosen by a *retail chain* with the aim of creating loyalty to the chain. Store brands have existed for many

[15] Note that there may be a problem of *free-riding* if franchisees can decide independently on service. The reason is that a franchisee who spends little effort on service gets the full benefit of her cost savings but shares the damage to goodwill with the other franchisee.

[16] A revenue-sharing contract is one in which a retailer pays a certain percentage of her sales revenue to the supplier who reciprocates by lowering the wholesale price. This enables the retailer to order more from the supplier and be better prepared to meet demand. A classic example here is *Blockbuster*, operating in the video rental industry, that signed revenue-sharing contracts with movie distributors. These arrangements were very successful (although not among Blockbuster's competitors).

years and recent years, as a consequence of the increasing power of retailer chains, have seen the emergence of many new store brands.

Example 16. Amrouche et al. (2008a,b) considered a channel where a retailer sells two brands: one produced by a national manufacturer and a private label. Firms set their respective prices and advertise to build up goodwill. The authors identified an FSE in the first paper and an FNE in the second. In the latter it was shown that creating goodwill for both brands mitigates price competition between the two types of brands. Karray and Martín-Herrán (2009) used a similar setup and studied complementary and competitive effects of advertising.

A line of research has addressed the problem of how to allocate shelf space to brands in a retail outlet. Shelf space is a scarce resource of a retailer who must allocate space between brands. With the increasing use of private labels, the retailer also faces a trade-off between space for national and for store brands. Martín-Herrán et al. (2005) studied a game between two competing manufacturers (who act as Stackelberg leaders) and a single retailer (the follower). The authors identified a time-consistent OLSE. See also Martín-Herrán and Taboubi (2005), Amrouche and Zaccour (2007).

4.7 The Marketing-Manufacturing Interface

Problems that involve multiple functional areas within a business firm are worthwhile addressing from a research point of view and from that of a manager. Some differential game studies have studied problems lying in the intersection between the two functional areas "marketing" and "manufacturing." First we look at some studies in which pricing and/or advertising games (as those encountered above) are extended with the determination of production levels, expansion and composition of manufacturing capacity, and inventories.

4.7.1 Production Capacity

A seminal paper by Thépot (1983) studied pricing, advertising goodwill creation, and productive capacity expansion. Capacity dynamics follow the standard capital accumulation equations

$$\dot{K}_i(t) = u_i(t) - \delta K_i(t)$$

where $K_i(t)$ is capacity, $u_i(t)$ the gross investment rate, and δ a depreciation factor. The author identified a series of advertising-investment-pricing regimes that may occur as OLNE. Gaimon (1989) considered two firms that determine prices: production rates and the levels and composition of productive capacity. Three games were analyzed: one with feedback strategies, one with open-loop strategies, and one in which firm i uses an open loop and firm j a feedback strategy. The type of strategy critically impacts the outcome of the game, and numerical simulations suggest that feedback strategies are superior in terms of profit.

Next we consider a branch of research devoted to the two-firm marketing channel (supply chain). The manufacturer chooses a production rate and the transfer price, while the retailer decides the consumer price and her ordering rate. Manufacturer and retailer inventories play a key role.

4.7.2 Production, Pricing, and Ordering

The following example focuses on decisions on transfer and consumer prices, production, procurement (ordering), and inventories.

Example 17. Jørgensen (1986b) studied a supply chain where a retailer determines the consumer price and the purchase rate from a manufacturer. The latter decides a production rate and the transfer price. Both firms carry inventories, and the manufacturer may choose to backlog retailer's orders. OLNE strategies were characterized. Eliashberg and Steinberg (1987) considered a similar setup and looked for an OLSE with the manufacturer as leader.

Desai (1996) studied a supply chain in which a manufacturer controls the production rate and the transfer price, while a retailer decides its processing (or ordering) rate and the consumer price. It was shown that centralized decision-making leads to a lower consumer price and higher production and processing rates, results that are expected. The author was right in stressing that channel coordination and contracting have implications, not only for marketing decisions but also for production and ordering decisions.

Jørgensen and Kort (2002) studied a system with two serial inventories, one located at a central warehouse and another at a retail outlet. A special feature is that the retail inventory is on display in the outlet and the hypothesis is that having a large displayed inventory will stimulate demand. The retail store manager determines the consumer price and the ordering rate from the central warehouse. The central warehouse manager orders from an outside supplier. The authors first analyzed a noncooperative game played by the two inventory managers, i.e., a situation in which inventory decisions are decentralized. Next, a centralized system was studied, cast as the standard problem of joint profit maximization.

Eliashberg and Steinberg (1991) investigated the strategic implications of manufacturers having different types of production and holding cost functions. The production cost of firm 1, called a "production smoother," is convex, inventory holding costs are linear, and the firm determines production, inventory, and pricing strategies. Production smoothing means that the firm aims at keeping the production rate close to the ordering rate of the retailer, trying to escape the unfortunate consequences of the convex production cost function. The production cost of firm 2, called an "order taker," is linear. Since this firm produces to order, it does not hold inventory. The firm determines production and pricing strategies.

4.7.3 Quality

Good *design quality* means that a product performs well in terms of durability and ease of use and "delivers what it promises." In Mukhopadhyay and Kouvelis (1997), firm i controls the rate of change of the design quality, denoted $u_i(t)$, of its product as well as the product price $p_i(t)$. Sales dynamics are given by a modified V-W model:

$$\dot{S}_i(t) = \alpha_i[m(t) - S_i(t)] - \delta_i S_i(t) + \gamma_i \dot{m}(t)$$

in which $m(t)$ is the market potential and the term $\gamma_i \dot{m}(t)$ is the fraction of new customers who will purchase from firm i. A firm incurs *(i)* a cost of changing the design quality as well as *(ii)* a cost of providing a certain quality level. In an OLNE, two things happen during an initial stage of the game: firms use substantial efforts to increase their quality levels as fast as possible and sales grow fast as prices are decreased. Later on, the firms can reap the benefits of having products of higher quality.

Conformance quality measures the extent to which a product conforms to design specifications. A simple measure is the proportion of non-defective units produced during some interval of time. In El Ouardighi et al. (2013), a manufacturer decides the production rate, the rate of conformance quality improvement efforts, and advertising. Two retailers compete on prices, and channel transactions follow a wholesale price or a revenue-sharing contract.

Teng and Thompson (1998) considered price and quality decisions for a new product, taking cost experience effects into account. Nair and Narasimhan (2006) suggested that product quality, in addition to advertising, affects the creation of goodwill in a duopoly.

5 Stochastic Games

As in other areas, the literature on stochastic games in marketing was preceded by optimal control studies of monopolistic firms. Early work on such problems includes Tapiero (1975), Sethi (1983), and Rishel (1985). See also the book by Tapiero (1988).

The interest in games of marketing problems played under uncertainty has been moderate. Among the rather few contributions are the following three:

Horsky (1988) studied a stochastic model of advertising competition with new-product diffusion dynamics represented by a Markov decision process (discrete state, continuous time). Firms can influence the transition probabilities through their advertising efforts. The author identified an FNE with stationary advertising strategies $a_i(Y_1, Y_2)$, and numerical simulations suggested that advertising rates are decreasing in Y_1, Y_2.

Chintagunta and Rao (1996) considered a stochastic game in which the dynamics (representing consumer preferences) are deterministic, but the value of the firm's brand depends on the aggregate consumer preference level, product price, and a random term. They identified steady-state equilibrium prices.

Prasad and Sethi (2004) studied a stochastic differential game using a modified version of the Lanchester advertising dynamics. Market share dynamics are given by the stochastic differential equation

$$dX_i = \left(\beta_i a_i \sqrt{X_j} - \beta_j a_j \sqrt{X_i} - \delta(2X_i - 1)\right) dt + \sigma\left(X_i, X_j\right) d\omega_i$$

in which the term $\delta(2X_i - 1)$ is supposed to model decay of individual market shares. The fourth term on the right-hand side is white noise.

6 Doing the Calculations: An Example

The purpose of this section is, using a simple problem of advertising goodwill accumulation, to show the reader how a differential game model can be formulated and, in particular, analyzed: The example appears in Jørgensen and Gromova (2016). Key elements are the dynamics (state equations) and the objective functionals, as well as the identification of the relevant constraints. Major tasks of the analysis are the characterization of equilibrium strategies and the associated state trajectories as well as the optimal profits to be earned by the players.

Consider an oligopolistic market with three firms, each selling its own particular brand. For simplicity of exposition, we restrict our analysis to the case of symmetric firms. The firms play a noncooperative game with an infinite horizon. Let $a_i(t) \geq 0$ be the rate of advertising effort of firm $i = 1, 2, 3$ and suppose that the stock of advertising goodwill of firm i, denoted $G_i(t)$, evolves according to the following Nerlove-Arrow-type dynamics

$$\dot{G}_i(t) = \kappa a_i(t); \quad G_i(0) \triangleq g_0 > 0 \tag{20.5}$$

Due to symmetry, advertising efforts of all firms are equally efficient (same value of κ), and we normalize κ to one. Note, in contrast to the N-A model in Sect. 2.5, that the dynamics in (20.5) imply that goodwill cannot decrease: Once a firm has accumulated goodwill up to a certain level, it is locked in. A firm's only options are to stay at this level (by refraining from advertising) or to continue to advertise (which will increase its stock of goodwill).

The dynamics in (20.5) could approximate a situation in which goodwill stocks decays rather slowly. As in the standard N-A model, the evolution of a firm's advertising goodwill depends on its own effort only. Thus the model relies on a hypothesis that competitors' advertising has no – or only negligible – impact on the goodwill level of a firm.[17]

[17] A modification would be to include competitors' advertising efforts on the right-hand side of (20.5), affecting negatively the goodwill stock of firm i. See Nair and Narasimhan (2006).

The following assumption is made for technical reasons. Under the assumption we can restrict attention to strategies for which value functions are continuously differentiable, and moreover, it will be unnecessary to explore the full state space of the system.

Assumption 1. *Initial values of the goodwill stocks satisfy* $g_0 < \beta/6$ *where* β *is a constant to be defined below.*

Let $s_i(t)$ denote the sales rate, required to be nonnegative, of firm i. The sales rate is supposed to depend on all three stocks of advertising goodwill:

$$s_i(t) = f_i(G(t)) \tag{20.6}$$

where $G(t) \triangleq (G_1(t), G_2(t), G_3(t))$. As we have seen, advertising of a firm does not affect the evolution of goodwill of its rivals, but advertising of a firm affects its own goodwill and therefore, through (20.6), the sales rates of all the firms in the industry. As in, e.g., Reynolds (1991) we specify function f_i as

$$f_i(G) = \left[\beta - \sum_{h=1}^{3} G_h\right] G_i \tag{20.7}$$

where $\beta > 0$ is the parameter referred to in Assumption 1. All firms must, for any $t > 0$, satisfy the path constraints

$$\sum_{h=1}^{3} G_h(t) \leq \beta \tag{20.8}$$

$$a_i(t) \geq 0.$$

Partial derivatives of function f_i are as follows:

$$\frac{\partial f_i}{\partial G_i} = \beta - 2G_i - G_j - G_k$$

$$\frac{\partial f_i}{\partial G_k} = -G_i, \; k \neq i.$$

Sales of a firm should increase when its goodwill increases, that is, we must require

$$\beta - 2G_i - G_j - G_k > 0 \text{ for any } t > 0. \tag{20.9}$$

Negativity of $\partial f_i/\partial G_k$ means that the sales rate of firm i decreases if goodwill of firm k increases.

Having described the evolution of goodwill levels and sales rates, it remains to introduce the economic components of the model. Let $\pi > 0$ be the constant profit margin of firm i. The cost of advertising effort a_i is

$$C(a_i) = \frac{c}{2}a_i^2$$

in which $c > 0$ is a parameter determining the curvature of the cost function.[18] Let $\rho > 0$ be a discounting rate, employed by all firms. The profit of firm i then is

$$J_i(a_i) = \int_0^\infty e^{-\rho t}\left\{\pi\left[\beta - \sum_{h=1}^{3} G_h(t)\right]G_i(t) - \frac{c}{2}a_i^2(t)\right\}dt. \tag{20.10}$$

Control and state constraints, to be satisfied by firm i for all $t \in [0,\infty)$, are as follows:

$$a_i(t) \geq 0; \quad \beta > 2G_i(t) + G_j(t) + G_k(t)$$

where we note that the second inequality implies $G_i(t)+G_j(t)+G_k(t) < \beta$. Thus, if sales f_i increase as goodwill G_i increases, sales cannot be negative.

The model parameters are time independent and the planning horizon is infinite. In such a problem, a standard approach in the literature has been to look for *stationary equilibria.* Stationarity means that advertising strategies and value functions will not depend explicitly on time.

We suppose that firms cannot (or will not) cooperate when determining their advertising efforts and hence consider a noncooperative game. In this game we shall identify an FNE and let $V_i(G)$ be a continuously differentiable value function of firm i. This function must solve the following Hamilton-Jacobi-Bellman (HJB) equation:

$$\rho V_i(G) = \max_{a_i \geq 0}\left\{\pi\left[\beta - \sum_{j=1}^{3} G_j\right]G_i - \frac{c}{2}a_i^2 + \frac{\partial V_i}{\partial G_i}a_i\right\}. \tag{20.11}$$

Performing the maximization on the right-hand side of (20.11) generates candidates for equilibrium advertising strategies:

$$a_i(G) = \begin{cases} \frac{1}{c}\frac{\partial V_i}{\partial G_i} > 0 & \text{if } \frac{\partial V_i}{\partial G_i} > 0 \\ 0 & \text{if } \frac{\partial V_i}{\partial G_i} \leq 0 \end{cases}$$

[18] A model with a linear advertising cost, say, ca_i and where a_i on the right-hand side of the dynamics is replace by, say $\sqrt{a_i}$, will provide the same results, qualitatively speaking, as the present model. To demonstrate this is left as an exercise for the reader.

which shows that identical firms use the same type of strategy. Inserting the candidate strategies into (20.11) provides

$$V_i(G) = \frac{1}{2\rho c}\left(\frac{\partial V_i}{\partial G_i}\right)^2 + \frac{\pi}{\rho}\left[\beta - \sum_{j=1}^{3} G_j\right] G_i \text{ if } a_i > 0 \tag{20.12}$$
$$V_i(G) = \frac{\pi}{\rho}\left[\beta - \sum_{j=1}^{3} G_j\right] G_i \text{ if } a_i = 0$$

from which it follows that

$$V_i(G)\mid_{a_i>0} - V_i(G)\mid_{a_i=0} = \frac{1}{2\rho c}\left(\frac{\partial V_i}{\partial G_i}\right)^2 \geq 0.$$

If the inequality is strict, strategy $a_i > 0$ payoff dominates strategy $a_i = 0$, and all firms will have a positive advertising rate throughout the game. If the equality sign holds, any firm is indifferent between advertising and no advertising. This occurs iff $\partial V_i/\partial G_i = 0$, a highly unusual situation in which a firm essentially has nothing to decide.

We conjecture that the equation in the first line in (20.12) has the solution

$$V_i(G) = \alpha + \gamma_A G_i + \frac{\epsilon_A}{2} G_i^2 + (\eta_A G_i + \gamma_B)(G_j + G_k) + \frac{\epsilon_B}{2}(G_j^2 + G_k^2) + \eta_B G_j G_k \tag{20.13}$$

in which $\alpha, \gamma_A, \gamma_B, \epsilon_A, \epsilon_B, \eta_A, \eta_B$ are constants to be determined. From (20.13) we get the partial derivatives

$$\frac{\partial V_i}{\partial G_i} = \gamma_A + \epsilon_A G_i + \eta_A(G_j + G_k). \tag{20.14}$$

Consider the HJB equation in the first line of (20.12) and calculate the left-hand side of this equation by inserting the value function given by (20.13). This provides

$$\rho\Big(\alpha + \gamma_A G_i + \gamma_B(G_j + G_k) + \frac{\epsilon_A}{2} G_i^2 + \frac{\epsilon_B}{2}(G_j^2 + G_k^2)$$
$$+ \eta_A(G_i G_j + G_i G_k) + \eta_B G_j G_k\Big) \tag{20.15}$$
$$= \frac{1}{2}\rho\epsilon_A G_i^2 + \rho\eta_A G_i G_j + \rho\eta_A G_i G_k + \rho\gamma_A G_i + \frac{1}{2}\rho\epsilon_B G_j^2$$
$$+ \rho\eta_B G_j G_k + \rho\gamma_B G_j + \frac{1}{2}\rho\epsilon_B G_k^2 + \rho\gamma_B G_k + \alpha\rho$$

Calculate the two terms on the right-hand side of the HJB equation:

$$\frac{1}{2\rho c}\left(\frac{\partial V_i}{\partial G_i}\right)^2 + \frac{\pi}{\rho}\left[\beta - \sum_{j=1}^{3} G_j\right] G_i \tag{20.16}$$

$$= \frac{1}{2c}\left(\gamma_A + \epsilon_A G_i + \eta_A(G_j + G_k)\right)^2 + \pi\left(\beta - \left(G_i + G_j + G_k\right)\right) G_i$$

$$= \frac{1}{2c}\gamma_A^2 + \frac{1}{2c}\epsilon_A^2 G_i^2 + \frac{1}{2c}\eta_A^2 G_j^2 + \frac{1}{2c}\eta_A^2 G_k^2 + \frac{1}{c}\gamma_A\epsilon_A G_i + \frac{1}{c}\gamma_A\eta_A G_j + \frac{1}{c}\gamma_A\eta_A G_k$$

$$+ \frac{1}{c}\eta_A^2 G_j G_k + \frac{1}{c}\epsilon_A\eta_A G_i G_j + \frac{1}{c}\epsilon_A\eta_A G_i G_k$$

$$+ \pi\beta G_i - \pi G_i^2 - \pi G_i G_j - \pi G_i G_k.$$

Using (20.15) and (20.16) the HJB equation becomes

$$\frac{1}{2}\rho\epsilon_A G_i^2 + \rho\eta_A G_i G_j + \rho\eta_A G_i G_k + \rho\gamma_A G_i + \frac{1}{2}\rho\epsilon_B G_j^2$$

$$+ \rho\eta_B G_j G_k + \rho\gamma_B G_j + \frac{1}{2}\rho\epsilon_B G_k^2 + \rho\gamma_B G_k + \alpha\rho$$

$$= \frac{1}{2c}\gamma_A^2 + \frac{1}{2c}\epsilon_A^2 G_i^2 + \frac{1}{2c}\eta_A^2 G_j^2 + \frac{1}{2c}\eta_A^2 G_k^2 + \frac{1}{c}\gamma_A\epsilon_A G_i + \frac{1}{c}\gamma_A\eta_A G_j$$

$$+ \frac{1}{c}\gamma_A\eta_A G_k + \frac{1}{c}\eta_A^2 G_j G_k + \frac{1}{c}\epsilon_A\eta_A G_i G_j + \frac{1}{c}\epsilon_A\eta_A G_i G_k$$

$$+ \pi\beta G_i - \pi G_i^2 - \pi G_i G_j - \pi G_i G_k$$

To satisfy this equation for any triple $\left(G_i, G_j, G_k\right)$, we must have

$$\begin{aligned}
\text{Constant term} : \alpha c\rho &= \frac{1}{2}\gamma_A^2 \\
G_i - \text{terms} : \rho c\gamma_A &= \gamma_A\epsilon_A + \pi c\beta \\
G_j G_k - \text{terms} : \rho c\gamma_B &= \gamma_A\eta_A \\
G_i^2 - \text{terms} : c\rho\epsilon_A &= \epsilon_A^2 - 2c\pi \\
G_j^2 \text{ and } G_k^2 - \text{terms} : \rho c\epsilon_B &= \eta_A^2 \\
G_i G_j \text{ and } G_i G_k - \text{terms} : \rho c\eta_A &= \epsilon_A\eta_A - c\pi \\
G_j G_k - \text{terms} : \rho c\eta_B &= \eta_A^2.
\end{aligned}$$

Disregarding α (which follows from γ_A), the remaining equations admit a unique solution

$$\gamma_A = \frac{\beta\left(\sqrt{c^2\rho^2 + 8c\pi} - c\rho\right)}{4} > 0; \quad \gamma_B = -\frac{\beta\left(c\rho - \sqrt{c^2\rho^2 + 8c\pi}\right)^2}{16c\rho} < 0$$

$$\epsilon_A = \frac{c\rho - \sqrt{c^2\rho^2 + 8c\pi}}{2} < 0; \quad \epsilon_B = \eta_B = \frac{\left(c\rho - \sqrt{c^2\rho^2 + 8c\pi}\right)^2}{16c\rho} > 0$$

$$\eta_A = \frac{c\rho - \sqrt{c^2\rho^2 + 8c\pi}}{4} < 0. \tag{20.17}$$

It is easy to see that the solution passes the following test of feasibility (cf. Bass et al. 2005a). If the profit margin π is zero, the value function should be zero because the firm has no revenue and does not advertise.

Knowing that $\epsilon_B = \eta_B$, the value function can be rewritten as

$$V_i(G) = \alpha + \gamma_A G_i + \frac{\epsilon_A}{2} G_i^2 + (\eta_A G_i + \gamma_B)(G_j + G_k) + \frac{\epsilon_B}{2}(G_j + G_k)^2.$$

Using (20.17) yields $\gamma_A = -\beta\eta_A$ and $\epsilon_A = 2\eta_A$, and the value function has the partial derivative

$$\frac{\partial V_i}{\partial G_i} = \gamma_A + \epsilon_A G_i + \eta_A(G_j + G_k) = \eta_A\left(2G_i + G_j + G_k - \beta\right)$$

which is positive because η_A is negative and $2G_i + G_j + G_k < \beta$ is imposed as a state constraint.

It remains to determine the time paths generated by the equilibrium advertising strategy and the associated goodwill levels. The advertising strategy is given by $a_i^*(G) = c^{-1}\partial V_i / \partial G_i$ and therefore

$$a_i^*(G) = \frac{\eta_A}{c}\left(2G_i + G_j + G_k - \beta\right). \tag{20.18}$$

Recalling the state dynamics $\dot{G}_i(t) = a_i(t)$, we substitute $a_i^*(G)$ into the dynamics and obtain a system of linear inhomogeneous differential equations:

$$\dot{G}_i(t) = \frac{2\eta_A}{c} G_i(t) + \frac{\eta_A}{c}\left(G_j(t) + G_k(t)\right) - \frac{\beta\eta_A}{c}. \tag{20.19}$$

Solving the equation in (20.19) yields

$$G_i(t) = \exp\left\{\frac{4\eta_A}{c} t\right\}\left(g_0 - \frac{\beta}{4}\right) + \frac{\beta}{4} \tag{20.20}$$

which shows that the goodwill stock of firm i converges to $\beta/4$ for $t \to \infty$. Finally, inserting $G_i(t)$ from (20.20) into (20.18) yields the time path of the equilibrium advertising strategy:

$$a_i^*(t) = \exp\left\{\frac{4\,\eta_A}{c}t\right\}\frac{\eta_A(4g_0-\beta)}{c} > 0.$$

Remark 3. Note that our calculations have assumed that the conditions $g_0 < \beta/6$ and, for all t, $2G_i(t) + G_j(t) + G_k(t) < \beta$ are satisfied. Using (20.20) it is easy to see that the second condition is satisfied.

7 Conclusions

This section offers some comments on (i) the modeling approach in differential games of marketing problems as well as on (ii) the use of numerical methods/algorithms to identify equilibria. Finally, we provide some avenues for future work.

Re (i): Differential game models are stylized representations of real-life, multi-player decision problems, and they rely – necessarily – on a number of simplifying assumptions. Preferably, assumptions should be justified theoretically as well as empirically. A problem here is that most models have not been validated empirically. Moreover, it is not always clear if the problem being modeled is likely to be a decision problem that real-life decision-makers might encounter. We give some examples of model assumptions that may be critical.

1. Almost all of the models that have been surveyed assume *deterministic* consumer demand. It would be interesting to see the implications of assuming *stochastic* demand, at least for the most important classes of dynamics. *Hint*: In marketing channel research, it could be fruitful to turn to the literature in operations and supply chain management that addresses supply chain incentives, contracting, and coordination under random consumer demand.
2. In *marketing channel research*, almost all contributions employed *Nerlove-Arrow* dynamics (or straightforward modifications of these). The reason for this choice most likely is the simplicity (and hence tractability) of the dynamics. It remains, however, to be seen what would be the conclusions if other dynamics were used. Moreover, most models consider a simple two-firm setup. A setup with more resemblance to real life would be one in which a manufacturer sells to multiple retailers, but this has been considered in a minority of models. Extending the setup raises pertinent questions, for example, how a channel leader can align the actions of multiple, asymmetric firms and if coordination requires tempering retailer competition. Another avenue for new research would be to study situations in which groups of retailers form coalitions, to stand united vis-a-vis a manufacturer. The literature on *cooperative differential games* should be useful here.
3. Many decisions in marketing affect, and are affected by, actions taken in *other functional areas*: procurement, production, quality management, capacity

investment and utilization, finance, and logistics. Only a small number of studies have dealt with these intersections

4. Quite many works study games in which firms have an *infinite planning horizon*. If, in addition, model parameters are assumed not to vary over time, analytical tractability is considerably enhanced. Such assumptions may be appropriate in theoretical work in mathematical economics but seem to be less useful if we insist that our recommendations should be relevant to real-life managers – who must face finite planning horizons and need to know how to operate in nonstationary environments.

Re (ii): In dynamic games it is most often the case that when models become more complex, the likelihood of obtaining complete analytical solutions (i.e., a full characterization of equilibrium strategies, the associated state trajectories, and the optimal profits of players) becomes smaller. In such situations one may resort to the use of numerical methods. This gives rise to two problems: Is there an appropriate method/algorithm and are (meaningful) data available.[19]

There exist numerical methods/algorithms to compute, given the data, equilibrium strategies and their time paths, the associated state trajectories, and profit/value functions. For stochastic control problems in continuous time, various numerical methods are available, see e.g., Kushner and Dupuis (2001), Miranda and Fackler (2002), and Falcone (2006). Numerical methods designed for specific problems are treated in Cardaliaguet et al. (2001, 2002) (pursuit games, state constraints). Numerical methods for noncooperative as well as cooperative differential games are dealt with in Engwerda (2005).

The reader should note that new results on numerical methods in dynamic games regularly appear in *Annals of the International Society of Dynamic Games*, published by Birkhäuser.

Example 18. Some studies resorted to numerical methods/algorithms to analyze differential games with Lanchester dynamics. For instance, Breton et al. (2006) identified an FSE (using data from the cola industry), while Jarrar et al. (2004) characterized an FNE. Both papers employed the Lanchester advertising model. See also Sadigh et al. (2012) who studied a problem of coordination of a two-member marketing channel. See also Jørgensen and Zaccour (2007).

The differential game literature has left many problems of marketing strategy untouched. In addition to what has already been said, we point to three:

1. The *service sector* is undoubtedly gaining increasing importance. However, the provision of services has received little attention in the literature on differential

[19] We shall disregard "numerical examples" where solutions are characterized using more or less randomly chosen data (typically, the values of model parameters).

games in marketing. An introduction to service marketing and management can be found in the handbook edited by Swartz and Iacobucci (2000).
2. *E-Business* shows substantial growth rates, but the area still has attracted little attention from researchers working with differential games in marketing. An introduction to marketing channels in the E-business era is provided by the handbook edited by Simchi-Levi et al. (2004).
3. A broad *range of instruments* are available to the marketing manager who tries to affect the decisions made by wholesalers, retailers, and consumers. The literature has focused on a relatively small number of such instruments. For example, the area of *pricing* is much broader than what our survey has shown. In real life we see various types of price promotions toward consumers and retailers, and there are interesting issues in, e.g., product line pricing, nonlinear pricing, advance selling, and price differentiation. An introduction to pricing research in marketing can be found in the handbook edited by Rao (2009).

The good news, therefore, are that there seems to be no shortage of marketing problems awaiting to be investigated by dynamic game methods.

References

Amrouche N, Zaccour G (2007) Shelf-space allocation of national and private brands. Eur J Oper Res 180:648–663

Amrouche N, Martín-Herrán G, Zaccour G (2008a) Feedback Stackelberg equilibrium strategies when the private label competes with the national brand. Ann Operat Res 164:79–95

Amrouche N, Martín-Herrán G, Zaccour G (2008b) Pricing and advertising of private and national brands in a dynamic marketing channel. J Optim Theory Appl 137:465–483

Aust G, Buscher U (2014) Cooperative advertising models in supply chain management. Eur J Oper Res 234:1–14

Başar T, Olsder GJ (1995) Dynamic noncooperative game theory, 2nd edn. Academic Press, London

Bass FM (1969) New product growth model for consumer durables. Manag Sci 15:215–227

Bass FM, Krishnamoorthy A, Prasad A, Sethi SP (2005a) Generic and brand advertising strategies in a dynamic duopoly. Mark Sci 24(4):556–568

Bass FM, Krishnamoorthy A, Prasad A, Sethi SP (2005b) Advertising competition with market expansion for finite horizon firms. J Ind Manag Optim 1(1):1–19

Breton M, Chauny F, Zaccour G (1997) Leader-follower dynamic game of new product diffusion. J Optim Theory Appl 92(1):77–98

Breton M, Jarrar R, Zaccour G (2006) A note on feedback sequential equilibria in a Lanchester Model with empirical application. Manag Sci 52(5):804–811

Buratto A, Zaccour G (2009) Coordination of advertising strategies in a fashion licensing contract. J Optim Theory Appl 142:31–53

Cardaliaguet P, Quincampoix M, Saint-Pierre P (2001) Pursuit differential games with state constraints. SIAM J Control Optim 39:1615–1632

Cardaliaguet P, Quincampoix M, Saint-Pierre P (2002) Differential games with state constraints. In: Proceedings of the tenth international symposium on dynamic games and applications, St. Petersburg

Case J (1979) Economics and the competitive process. New York University Press, New York

Chatterjee R (2009) Strategic pricing of new products and services. In: Rao VR (ed) Handbook of pricing research in marketing. E. Elgar, Cheltenham, pp 169–215
Chatterjee R, Eliashberg J, Rao VR (2000) Dynamic models incorporating competition. In: Mahajan V, Muller E, Wind Y (eds) New product diffusion models. Kluwer Academic Publishers, Boston
Chiang WK (2012) Supply chain dynamics and channel efficiency in durable product pricing and distribution. Manuf Serv Oper Manag 14(2):327–343
Chintagunta PK (1993) Investigating the sensitivity of equilibrium profits to advertising dynamics and competitive effects. Manag Sci 39:1146–1162
Chintagunta PK, Jain D (1992) A dynamic model of channel member strategies for marketing expenditures. Mark Sci 11:168–188
Chintagunta PK, Jain D (1995) Empirical analysis of a dynamic duopoly model of competition. J Econ Manag Strategy 4:109–131
Chintagunta PK, Rao VR (1996) Pricing strategies in a dynamic duopoly: a differential game model. Manag Sci 42:1501–1514
Chintagunta PK, Vilcassim NJ (1992) An empirical investigation of advertising strategies in a dynamic duopoly. Manag Sci 38:1230–1244
Chintagunta PK, Vilcassim NJ (1994) Marketing investment decisions in a dynamic duopoly: a model and empirical analysis. Int J Res Mark 11:287–306
Chintagunta PK, Rao VR, Vilcassim NJ (1993) Equilibrium pricing and advertising strategies for nondurable experience products in a dynamic duopoly. Manag Decis Econ 14:221–234
Chutani A, Sethi SP (2012a) Optimal advertising and pricing in a dynamic durable goods supply chain. J Optim Theory Appl 154(2):615–643
Chutani A, Sethi SP (2012b) Cooperative advertising in a dynamic retail market oligopoly. Dyn Games Appl 2:347–375
De Giovanni P, Reddy PV, Zaccour G (2015) Incentive strategies for an optimal recovery program in a closed-loop supply chain. Eur J Oper Res 249(2):1–13
Deal K (1979) Optimizing advertising expenditure in a dynamic duopoly. Oper Res 27:682–692
Desai VS (1996) Interactions between members of a marketing-production channel under seasonal demand. Eur J Oper Res 90:115–141
Dockner EJ (1985) Optimal pricing in a dynamic duopoly game model. Zeitschrift für Oper Res 29:1–16
Dockner EJ, Gaunersdorfer A (1996) Strategic new product princing when demand obeys saturation effects. Eur J Oper Res 90:589–598
Dockner EJ, Jørgensen S (1984) Cooperative and non-cooperative differential game solutions to an investment and pricing problem. J Oper Res Soc 35:731–739
Dockner EJ, Jørgensen S (1988a) Optimal pricing strategies for new products in dynamic oligopolies. Mark Sci 7:315–334
Dockner EJ, Jørgensen S (1992) New product advertising in dynamic oligopolies. Zeitschrift für Oper Res 36:459–473
Dockner EJ, Gaunersdorfer A, Jørgensen S (1996) Government price subsidies to promote fast diffusion of a new consumer durable. In: Jørgensen S, Zaccour G (eds) Dynamic competitive analysis in marketing. Springer, Berlin, pp 101–110
Dockner EJ, Jørgensen S, Van Long N, Sorger G (2000) Differential games in economics and management science. Cambridge University Press, Cambridge
Dolan R, Jeuland AP, Muller E (1986) Models of new product diffusion: extension to competition against existing and potential firms over time. In: Mahajan V, Wind Y (eds) Innovation diffusion models of new product acceptance. Ballinger, Cambridge, pp 117–150
El Ouardighi F, Jørgensen S, Pasin F (2013) A dynamic game with monopolist manufacturer and price-competing duopolist retailers. OR Spectr 35:1059–1084
Eliashberg J, Chatterjee R (1985) Analytical models of competition with implications for marketing: issues, findings, and outlook. J Mark Res 22:237–261
Eliashberg J, Jeuland AP (1986) The impact of competitive entry in a developing market upon dynamic pricing strategies. Mark Sci 5:20–36

Eliashberg J, Steinberg R (1987) Marketing-production decisions in an industrial channel of distribution. Manag Sci 33:981–1000
Eliashberg J, Steinberg R (1991) Competitive strategies for two firms with asymmetric production cost structures. Manag Sci 37:1452–1473
Eliashberg J, Steinberg R (1993) Marketing-production joint decision-making. In: Eliashberg J, Lilien GL (eds) Handbooks of operations research and management science: marketing. North-Holland, Amsterdam, pp 827–880
Engwerda J (2005) Algorithms for computing nash equilibria in deterministic LQ games. Comput Manag Sci 4(2):113–140
Erickson GM (1985) A model of advertising competition. J Mark Res 22:297–304
Erickson GM (1991) Dynamic models of advertising competition. Kluwer, Boston
Erickson GM (1992) Empirical analysis of closed-loop duopoly advertising strategies. Manag Sci 38:1732–1749
Erickson GM (1993) Offensive and defensive marketing: closed-loop duopoly strategies. Mark Lett 4:285–295
Erickson GM (1995a) Differential game models of advertising competition. Eur J Oper Res 83:431–438
Erickson GM (1995b) Advertising strategies in a dynamic oligopoly. J Mark Res 32:233–237
Erickson GM (1996) An empirical comparison of oligopolistic advertising strategies. In: Jørgensen S, Zaccour G (eds) Dynamic competitive analysis in marketing. Springer, Berlin, pp 26–36
Falcone M (2006) Numerical methods for differential games based on partial differential equations. Int Game Theory Rev 8:231–272
Erickson GM (1997) Dynamic conjectural variations in a lanchester duopoly. Manag Sci 43:1603–1608
Feichtinger G, Dockner EJ (1985) Optimal pricing in a duopoly: a noncooperative differential games solution. J Optim Theory Appl 45:199–218
Feichtinger G, Hartl RF, Sethi SP (1994) Dynamic optimal control models in advertising: recent developments. Manag Sci 40:195–226
Fershtman C (1984) Goodwill and market shares in oligopoly. Economica 51:271–281
Fershtman C, Mahajan V, Muller E (1990) Market share pioneering advantage: a theoretical approach. Manag Sci 36:900–918
Friedman JW (1991) Game theory with applications to economics, 2nd edn. Oxford University Press, Oxford
Fruchter GE (2001) Advertising in a competitive product line. Int Game Theory Rev 3(4):301–314
Fruchter GE, Kalish S (1997) Closed-loop advertising strategies in a duopoly. Manag Sci 43: 54–63
Fruchter GE, Kalish S (1998) Dynamic promotional budgeting and media allocation. Eur J Oper Res 111:15–27
Fruchter GE, Erickson GM, Kalish S (2001) Feedback competitive advertising strategies with a general objective function. J Optim Theory Appl 109:601–613
Gaimon C (1989) Dynamic game results of the acquisition of new technology. Oper Res 37:410–425
Gaimon C (1998) The price-production problem: an operations and marketing interface. In: Aronson JE, Zionts S (eds) Operations research: methods, models, and applications. Quorum Books, Westpoint, pp 247–266
Gallego G, Hu M (2014) Dynamic pricing of perishable assets under competition. Manag Sci 60(5):1241–1259
Haurie A, Krawczyk JB, Zaccour G (2012) Games and dynamic games. Now/World Scientific, Singapore
He X, Prasad A, Sethi SP, Gutierrez GJ (2007) A survey of stackelberg differential models in supply and marketing channels. J Syst Sci Eng 16(4):385–413
Helmes K, Schlosser R (2015) Oligopoly pricing and advertising in isoelastic adoption models. Dyn Games Appl 5:334–360

Horsky D, Mate K (1988) Dynamic advertising strategies of competing durable goods producers. Mark Sci 7:356–367
Horsky D, Simon LS (1983) Advertising and the diffusion of new products. Mark Sci 2:1–18
Huang J, Leng M, Liang L (2012) Recent developments in dynamic advertising research. Eur J Oper Res 220:591–609
Janssens G, Zaccour G (2014) Strategic price subsidies for new technologies. Automatica 50:1998–2006
Jarrar R, Martín-Herrán G, Zaccour G (2004) Markov perfect equilibrium advertising strategies in a lanchester duopoly model: a technical note. Manag Sci 50(7):995–1000
Jørgensen S (1982a) A survey of some differential games in advertising. J Econ Dyn Control 4:341–369
Jørgensen S (1982b) A differential game solution to an investment and pricing problem. J Operat Res Soc 35:731–739
Jørgensen S (1986a) Optimal dynamic pricing in an oligopolistic market: a survey. In: Basar T (ed) Dynamic games and applications in economics. Springer, Berlin, pp 179–237
Jørgensen S (1986b) Optimal production, purchasing, and pricing: a differential game approach. Eur J Oper Res 24:64–76
Jørgensen S (2011) Intertemporal contracting in a supply chain. Dyn Games Appl 1:280–300
Jørgensen S, Gromova E (2016) Sustaining cooperation in a differential game of advertising goodwill. Eur J Oper Res 254:294–303
Jørgensen S, Kort PM (2002) Optimal pricing and inventory policies: centralized and decentralized decision making. Eur J Oper Res 138:578–600
Jørgensen S, Sigué SP (2015) Defensive, offensive, and generic advertising in a lanchester model with market growth. Dyn Games Appl 5(4):523–539
Jørgensen S, Sigue SP, Zaccour G (2000) Dynamic cooperative advertising in a channel. J Retail 76:71–92
Jørgensen S, Sigue SP, Zaccour G (2001) Stackelberg leadership in a marketing channel. Int Game Theory Rev 3(1):13–26
Jørgensen S, Taboubi S, Zaccour G (2003) Retail promotions with negative brand image effects: is cooperation possible. Eur J Oper Res 150:395–405
Jørgensen S, Taboubi S, Zaccour G (2006) Incentives for retailer promotion in a marketing channel. In: Haurie A, Muto S, Petrosjan LA, Raghavan TES (eds) Advances in dynamic games. Annals of ISDG, vol 8, Birkhäuser, Boston, Ma., USA
Jørgensen S, Martín-Herrán G, Zaccour G (2010) The Leitmann-Schmitendorf advertising differential game. Appl Math Comput 217:1110–1118
Jørgensen S, Zaccour G (1999a) Equilibrium pricing and advertising strategies in a marketing channel. J Optim Theory Appl 102:111–125
Jørgensen S, Zaccour G (1999b) Price subsidies and guaranteed buys of a new technology. Eur J Oper Res 114:338–345
Jørgensen S, Zaccour G (2003a) A differential game of retailer promotions. Automatica 39:1145–1155
Jørgensen S, Zaccour G (2003b) Channel coordination over time: incentive equilibria and credibility. J Econ Dyn Control 27:801–822
Jørgensen S, Zaccour G (2004) Differential games in marketing. Kluwer Academic Publishers, Boston
Jørgensen S, Zaccour G (2007) Developments in differential game theory and numerical methods: economic and management applications. Comput Manag Sci 4(2):159–181
Jørgensen S, Zaccour G (2014) A survey of game-theoretic models of cooperative advertising. Eur J Oper Res 237:1–14
Kalish S (1983) Monopolist pricing with dynamic demand and production costs. Mark Sci 2: 135–160
Kalish S (1988) Pricing new products from birth to decline: an expository review. In: Devinney TM (ed) Issues in pricing. Lexington Books, Lexington, pp 119–144

Kalish S, Lilien GL (1983) Optimal price subsidy for accelerating the diffusion of innovations. Mark Sci 2:407–420
Karray S, Martín-Herrán G (2009) A dynamic model for advertising and pricing competition between national and store brands. Eur J Oper Res 193:451–467
Kimball GE (1957) Some industrial applications of military operations research methods. Oper Res 5:201–204
Krishnamoorthy A, Prasad A, Sethi SP (2010) Optimal pricing and advertising in a durable-good duopoly. Eur J Oper Res 200(2):486–497
Kushner HJ, Dupuis PG (2001) Numerical methods for stochastic optimal control problems in continuous time. Springer, Berlin
Lanchester FW (1916) Aircraft in warfare: the dawn of the fourth arm. Constable, London
Leitmann G, Schmitendorf WE (1978) Profit maximization through advertising: nonzero-sum differential game approach. IEEE Trans Autom Control 23:645–650
Lieber Z, Barnea A (1977) Dynamic optimal pricing to deter entry under constrained supply. Oper Res 25:696–705
Mahajan V, Muller E, Bass FM (1990) New-product diffusion models: a review and directions for research. J Mark 54:1–26
Mahajan V, Muller E, Bass FM (1993) New-product diffusion models. In: Eliashberg J, Lilien GL (eds) Handbooks in operations research and management science: marketing. North-Holland, Amsterdam
Mahajan V, Muller E, Wind Y (eds) (2000) New-product diffusion models. Springer, New York
Martín-Herrán G, McQuitty S, Sigué SP (2012) Offensive versus defensive marketing: what is the optimal spending allocation? Int J Res Mark 29:210–219
Martín-Herrán G, Taboubi S (2005) Shelf-space allocation and advertising decisions in the marketing channel: a differential game approach. Int Game Theory Rev 7(3): 313–330
Martín-Herrán G, Taboubi S (2015) Price coordination in distribution channels: a dynamic perspective. Eur J Oper Res 240:401–414
Martín-Herrán G, Taboubi S, Zaccour G (2005) A time-consistent open-loop stackelberg equilibrium of shelf-space allocation. Automatica 41:971–982
Martín-Herrán G, Sigué SP, Zaccour G (2011) Strategic interactions in traditional franchise systems: are franchisors always better off? Eur J Oper Res 213:526–537
Mesak HI, Calloway JA (1995) A pulsing model of advertising competition: a game theoretic approach, part B – empirical application and findings. Eur J Oper Res 86:422–433
Mesak HI, Darrat AF (1993) A competitive advertising model: some theoretical and empirical results. J Oper Res Soc 44:491–502
Minowa Y, Choi SC (1996) Optimal pricing strategy and contingent products under a duopoly environment. In: Jørgensen S, Zaccour G (eds) Dynamic competitive analysis in marketing. Springer, Berlin, pp 111–124
Miranda MJ, Fackler PL (2002) Applied computational economics and finance. The MIT Press, Cambridge
Moorthy KS (1993) Competitive marketing strategies: game-theoretic models. In: Eliashberg J, Lilien GL (eds) Handbooks in operations research and management science: marketing. North-Holland, Amsterdam, pp 143–190
Mukhopadhyay SK, Kouvelis P (1997) A differential game theoretic model for duopolistic competition on design quality. Oper Res 45:886–893
Nair A, Narasimhan R (2006) Dynamics of competing with quality- and advertising-based Goodwill. Eur J Oper Res 175:462–476
Nerlove M, Arrow KJ (1962) Optimal advertising policy under dynamic conditions. Economica 39:129–142
Prasad A, Sethi SP (2004) Competitive advertising under uncertainty: a stochastic differential game approach. J Optim Theory Appl 123(1):163–185
Rao RC (1988) Strategic pricing of durables under competition. In: Devinney TM (ed) Issues in pricing. Lexington Books, Lexington, pp 197–215

Rao RC (1990) The impact of competition on strategic marketing decisions. In: Day G, Weitz B, Wensley R (eds) The interface of marketing and strategy. JAI Press, Greenwich, pp 101–152
Rao VR (2009) Handbook of pricing research in marketing. Edward Elgar, Cheltenham
Reynolds SS (1991) Dynamic oligopoly with capacity adjustment costs. Journal of Economic Dynamics & Control 15:491–514
Rishel R (1985) A partially observed advertising model. In: Feichtinger G (ed) Optimal control theory and economic analysis. North-Holland, Amsterdam, pp 253–262
Sadigh AN, Mozafari M, Karimi B (2012) Manufacturer-retailer supply chain coordination: a bi-level programming approach. Adv Eng Softw 45(1):144–152
Sethi SP (1983) Deterministic and stochastic optimization of a dynamic advertising model. Optim Control Methods Appl 4:179–184
Sigué SP (2002) Horizontal strategic interactions in franchising. In: Zaccour G (ed) Decision and control in management science: essays in honor of Alain Haurie. Kluwer, Boston
Sigué SP, Chintagunta P (2009) Advertising strategies in a franchise system. Eur J Oper Res 198:655–665
Simchi-Levi D, Wu SD, Shen Z-J (2004) Handbook of quantitative supply chain analysis: modeling in the e-business era. Springer, Berlin
Sorger G (1989) Competitive dynamic advertising: a modification of the case game. J Econ Dyn Control 13:55–80
Spengler JJ (1950) Vertical integration and antitrust policy. J Polit Econ 58:347–352
Sudhir K, Datta S (2009) Pricing in marketing channels. In: Rao VR (ed) Handbook of pricing research in marketing. Elgar E, Cheltenham, pp 169–215
Swartz TA, Iacobucci D (2000) Handbook of services marketing & management. Sage Publications, London
Tapiero CS (1975) Random walk models of advertising, their diffusion approximation and hypothesis testing. Ann Econ Soc Meas 4:293–309
Tapiero CS (1988) Applied stochastic models and control in management. North-Holland, Amsterdam
Teng JT, Thompson GL (1983) Oligopoly models for optimal advertising when production costs obey a learning curve. Manag Sci 29:1087–1101
Teng JT, Thompson GL (1985) Optimal strategies for general price-advertising models. In: Feichtinger G (ed) Optimal control theory and economic analysis, vol 2. Elsevier, Amsterdam, pp 183–195
Teng JT, Thompson GL (1998) Optimal strategies for general price-quality decision models of new products with learning production costs. In: Aronson JE, Zionts S (eds) Operations research: methods, models and applications. Greenwood Publishing Co., Westport
Thépot J (1983) Marketing and investment policies of duopolists in a growing industry. J Econ Dyn Control 5:387–404
Tirole J (1988) The theory of industrial organization. The MIT Press, Cambridge
Vega-Redondo F (2003) Economics and the theory of games. Cambridge University Press, Cambridge
Vidale ML, Wolfe HB (1957) An operations research study of sales response to advertising. Oper Res 5:370–381
Vives X (1999) Oligopoly pricing: old ideas and new tools. MIT Press, Cambridge
Wang Q, Wu Z (2001) A duopolistic model of dynamic competitive advertising. Eur J Oper Res 128:213–226
Xie J, Sirbu M (1995) Price competition and compatibility in the presence of positive demand externalities. Manag Sci 41(5):909–926
Zaccour G (1996) A differential game model for optimal price subsidy of new technologies. Game Theory Appl 2:103–114
Zaccour G (2008) On the coordination of dynamic marketing channels and two-part tariffs. Automatica 44:1233–1239
Zhang J, Gou Q, Liang L, Huang Z (2012) Supply chain coordination through cooperative advertising with reference price effect. Omega 41(2):345–353

Social Networks

21

Sadegh Bolouki, Angelia Nedić, and Tamer Başar

Contents

Abstract

In this chapter, some applications of game theory in social network analysis are presented. We first focus on the opinion dynamics of a social network. Viewing the individuals as players of a game with appropriately defined action (opinion) sets and utility functions, we investigate the best response dynamics

S. Bolouki (✉)
Coordinated Science Laboratory, University of Illinois, Urbana, IL, USA
e-mail: sadegh.bolouki@gmail.com

A. Nedić
School of ECEE, Arizona State University, Tempe, AZ, USA
e-mail: angelia.nedich@asu.edu

T. Başar
Coordinated Science Laboratory and Department of Electrical and Computer Engineering, University of Illinois at Urbana-Champaign, Urbana, IL, USA
e-mail: basar1@illinois.edu

T. Başar, G. Zaccour (eds.), *Handbook of Dynamic Game Theory*,
https://doi.org/10.1007/978-3-319-44374-4_32

and its variants for the game, which would in effect represent the evolution of the individuals' opinions within a social network. The action sets are defined according to the nature of the opinions, which may be continuous, as for the political beliefs of the individuals, or discrete, as for the type of technology adopted by the individuals to use in their daily lives. The utility functions, on the other hand, are to best capture the social behavior of the individuals such as conformity and stubbornness. For every formulation of the game, we characterize the formation of the opinions as time grows. In particular, we determine whether an agreement among all of the individuals is reached, a clustering of opinions occurs, or none of the said cases happens. We further investigate the Nash equilibria of the game and make clear if the game dynamics converges to one of the Nash equilibria. The rate of convergence to the equilibrium, if it is the case, is also obtained. We then turn our attention to decision-making processes (elections) in social networks, where a collective decision (social choice) must be made by multiple individuals (voters) with different preferences over the alternatives (candidates). We argue that the nonexistence of a perfectly fair social choice function that takes all voter preferences into account leads to the emergence of various strategic games in decision-making processes, most notably strategic voting, strategic candidacy, and coalition formation. While the strategic voting would be played among the voters, the other two games would be played among the candidates. We explicitly discuss the games of strategic candidacy and coalition formation.

Keywords

Game theory · Social networks · Opinion dynamics · Coordination games · Potential games · Social choice · Strategic voting · Strategic candidacy · Coalition formation

1 Introduction

Who should I vote for? What job should I choose? What smartphone should I buy next? These are only a few out of a wide range of questions whose answers are influenced significantly by different kinds of social networks surrounding us. Such undebatable influences on everyone's life have drawn attention to social networks and their analysis from a diverse group of researchers including, but not limited to, mathematicians, social scientists, social psychologists, physicists, and computer scientists.

The introduction of game theory as an analytical tool to model and study social behavior dates back to the 1950s. Jessie Bernard in her programmatic work (Bernard 1954) argues that game theory would in many ways contribute to sociology, particularly in analysis of social conflict and social organization. Since Bernard's article, to this date employment of game theory by sociologists and other researchers interested in social behavior has substantially grown. For a rather historical review of the relationship between game theory and sociology, the interested reader is referred to Swedberg (2001).

This chapter is devoted to applications of game theory to two important concepts in social networks with finite populations: *opinion dynamics* and *social choice*. As we will see in this chapter, these two concepts are fundamentally different from a game theoretic perspective, as one is associated with the evolution of the opinions over an extended period of time, hence resembling a multistage process, while the other one can be viewed a single-stage process. We further clarify these two concepts below:

Opinion Dynamics. Every human being has her/his own views of different topics: political views, religious views, etc. Engaged in conversations, people influence each other's opinions. As time goes by, they may reach an agreement about some topics while split into separate opinion-alike groups about some other topics. An *agreement*, as a type of social coordination, about a certain subject in a social network refers to the convergence of evolving opinions held by the individuals of the network about the subject. Another type of social coordination is *clustering*, that is the convergence of individual opinions, but to possibly different limits. A clustering thus corresponds to a partitioning of the entire group of individuals into several subgroups such that within each subgroup an agreement is reached. To investigate coordination in social networks, one has to first characterize how the individual opinions evolve over time, i.e., the so-called opinion dynamics of the network. The method by which the network dynamics is mathematically modeled varies according to exclusive circumstances of the network as well as the subject of the opinions.

Social Choice. A social choice is a collective decision made by the individuals in a social network through a decision-making process. Such a process, generally known as *election*, is essentially designed for a social network to reach a decision particularly in the absence of unanimity, that is, when a global agreement is not in place. A must-have in an election is *fairness* to all the candidates and the voters. However, the notion of fairness is as ill-defined as it is crucial. Very different interpretations of fairness have led to various *electoral systems* or *voting schemes* such as plurality, instant runoff, and Borda rules.

The objective of this chapter, as it pertains to opinion dynamics, is to analyze the network's opinion dynamics, particularly the asymptotic formation of the opinions, via game theoretic approaches. We shall view the individuals as the players of a game where the individuals' opinions comprise the player actions. This game would be played repeatedly among the individuals. A player's utility function must then be defined appropriately, in the sense that higher utilities are socially more desirable for her. The communication structure of the social network must also be taken into account that translates to a restricted information set for each player in the game. We will refer to such a game as a *coordination game*[1] since we are generally interested in situations in which some type of coordination, such as agreement or clustering, is achieved by the individual in the network. For various formulations of coordination

[1]This definition of coordination games is specific to this chapter and may be different from other definitions used in the literature.

games, we will establish whether the steady-state behavior of the game constitutes agreement or clustering. Furthermore, we will characterize the Nash equilibria of the game and determine whether the game dynamics converges to one of the Nash equilibria. In case convergence occurs, the convergence rate is also of particular interest. Our formulations of coordination games modeling the opinion dynamics of a network are split into two parts: one addressing continuous opinions, as carried out in Sect. 3, and the other one addressing discrete opinions, as carried out in Sect. 4.

Our discussion regarding the notion of social choice is covered in Sect. 5, where we will first address the fairness of a social choice function by introducing several fairness criteria. We will argue that no social choice function can meet all the fairness criteria, which consequently makes it susceptible to *strategic voting* and *strategic candidacy*. We will make explicit the game of strategic candidacy involved in an election. Finally, we will investigate the game of *coalition formation* by the candidates.

To fully comprehend the contents of this chapter, the reader is expected to be familiar with basic concepts of game theory, graph theory, and probability theory.

2 A Review of Opinion Dynamics

Let $\mathcal{P} = \{\mathcal{P}_1, \ldots, \mathcal{P}_n\}$ denote the group of individuals in a social network. For every individual $\mathcal{P}_i \in \mathcal{P}$, we model her opinion at a discrete time $t \geq 0$ as $x_i(t) \in \mathcal{A}_i \subset \mathbb{R}^d$, where $d \geq 1$ is a constant integer. The choice of $\mathcal{A}_i$ varies by the particular problem under consideration. In Sect. 3, we shall deal with the case in which $\mathcal{A}_i$s are convex subsets of $\mathbb{R}^d$, while in Sect. 4, it is assumed that $\mathcal{A}_i$s are finite sets. However, for the purposes of this section, we assume that $d = 1$, i.e., $\mathcal{A}_i = \mathbb{R}$, although the arguments can be carried over to higher dimensional spaces.

Dynamics of a general opinion network, to determine the evolution of individual opinions, is defined as follows.

Definition 1 (dynamics). Given a network of n individuals $\mathcal{P}_1, \ldots, \mathcal{P}_n$, a *dynamics* refers to a triplet (x, f, X_0) which satisfies:

$$\begin{cases} x(t+1) = f\big(t, x(t)\big), \ t \geq 0, \\ x(0) = X_0, \end{cases} \tag{21.1}$$

where:

(i) $x(t) \in \mathbb{R}^n$, $t \geq 0$, denotes the *opinion vector*;
(ii) $f(.,.) : \mathbb{N} \cup \{0\} \times \mathbb{R}^n \to \mathbb{R}^n$ is called an *update rule*; and
(iii) $X_0 \in \mathbb{R}^n$ is the *initial opinion vector*.

A *global agreement*, or simply an *agreement*, within the social network is formulated as follows.

Definition 2 (agreement). Given a dynamics (x, f, X_0), *agreement* is said to be achieved if for any $X_0 \in \mathbb{R}^n$, there exists a constant $c \in \mathbb{R}$ such that:

$$\lim_{t\to\infty} x_i(t) = c, \; \forall \mathcal{P}_i \in \mathcal{P}. \tag{21.2}$$

Furthermore, one defines *clustering* as follows.

Definition 3 (clustering). Given a dynamics (x, f, X_0), *clustering* is said to have occurred if for any $X_0 \in \mathbb{R}^n$, $\lim_{t\to\infty} x_i(t)$ exists for every individual $\mathcal{P}_i \in \mathcal{P}$.

Notice that, according to Definition 3, agreement can be considered as a special case of clustering when there is a single cluster. Coordination problems such as agreement and clustering arise in many research areas. In biology, such problems are seen in the emergent behavior of bird flocks, fish schools, etc. (Couzin et al. 2005; Cucker and Smale 2007; Flierl et al. 1999). Coordination models are designed to interpret, analyze, and predict flocking aggregation behavior. In robotics and control, coordination and cooperation of mobile agents have been studied (Jadbabaie et al. 2003), which are of great importance in sensor networks for environmental applications or space exploration. In sociology, the emergence of a common language in primitive societies (Cucker et al. 2004) indicates the phenomenon of reaching an agreement within a complex system. Coordination algorithms have also been extensively studied within the computer science community (Lynch 1996) as well as the management science community (see DeGroot 1974 and references therein). Applications in physics, biophysics, and neurobiology include synchronization of coupled oscillators, i.e., reaching a common frequency of coupled oscillators (Ermentrout 1992; Graver et al. 1984; Strogatz 2001).

An important type of opinion dynamics (21.1) is known as *distributed averaging dynamics* or, simply, *averaging dynamics*. In such dynamics, each individual has a neighboring set, which is a subset of individuals, and, at each time instant, updates her opinion to a convex combination of her neighbors' opinions. Thus, an averaging dynamics is associated with an update rule defined as:

$$f(t, x) \triangleq Ax, \tag{21.3}$$

for some fixed, row-stochastic matrix $A = [a_{ij}]$. A *row-stochastic* matrix refers to a matrix where every row of which sums up to 1. Thus, in view of (21.1), one now has:

$$\begin{cases} x(t+1) = Ax(t), \; t \geq 0, \\ x(0) = X_0. \end{cases} \tag{21.4}$$

We note that a translation of all elements of X_0 by an arbitrary constant $c \in \mathbb{R}$ simply leads to the same translation of all elements of $x(t)$ for each $t \geq 0$ since the matrix A is row-stochastic. Consequently, this translation preserves major asymptotic properties of $x(t)$, such as whether an agreement is achieved or a clustering occurs. The same argument is valid for any scaling of X_0. Therefore, without loss of generality, via a proper translation and scaling, we can assume that the elements of X_0 belong to the interval [0, 1]. Furthermore, once again, since A is row-stochastic, it is clear from (21.4) that $\min_{\mathcal{P}_i \in \mathcal{P}} x_i(t)$ is nondecreasing with respect to t. Similarly, $\max_{\mathcal{P}_i \in \mathcal{P}} x_i(t)$ is nonincreasing with respect to t. Therefore, given an averaging dynamics (x, f, X_0), the interval

$$\left[\min_{\mathcal{P}_i \in \mathcal{P}} x_i(t), \max_{\mathcal{P}_i \in \mathcal{P}} x_i(t) \right], \tag{21.5}$$

which represents the *range* of the individual opinions at time t, is non-expanding with respect to t and remains a subset of [0, 1] at all times. Thus, the following assumption is standard in the context of opinion dynamics.

Assumption 1. *For every individual $\mathcal{P}_i \in \mathcal{P}$: $x_i(t) \in [0, 1]$.*

Averaging dynamics, as a type of opinion dynamics, was first introduced by DeGroot (1974). He considered a group of individuals in a team or committee who seeks a common probability distribution for the unknown value of some parameter θ. Each individual is assumed to initially have her own subjective probability distribution for the unknown parameter. The simple DeGroot model (DeGroot 1974), which is perfectly consistent with dynamics (21.4), is the following: Assume that $F_i(t)$ denotes the belief of individual i at a discrete time instant t about the distribution of θ. At any time instant t, individual $\mathcal{P}_i$ updates her belief via the update equation $F_i(t+1) = \sum_{\mathcal{P}_j \in \mathcal{V}} a_{ij} F_j(t)$, where coefficients a_{ij} are nonnegative constants that satisfy $\sum_{\mathcal{P}_j \in \mathcal{V}} a_{ij} = 1, \forall \mathcal{P}_i \in \mathcal{P}$. Furthermore, the coefficients a_{ij} are assumed symmetric, i.e., $a_{ij} = a_{ji}$ for every pair $\mathcal{P}_i, \mathcal{P}_j \in \mathcal{P}$. Employing properties of Markov chains, DeGroot arrived at a sufficient condition for the convergence of individual distributions to a common distribution, which would in essence be the average of the initial distributions since the coefficients were assumed to be symmetric. In the ensuing years and decades, dynamics (21.4) has been pursued and generalized by a large number of researchers. Notable works include Chatterjee and Seneta (1977), for considering a time-varying version of the model, and Tsitsiklis (1984), for exploring the effect of asymmetric coefficients a_{ij}.

The literature on the subject of averaging dynamics has expanded in a number of directions. Here, we provide a brief overview of a few of those research directions:

– *Bounded confidence models.* Roughly speaking, in bounded confidence models, individuals are influenced only by those who lie in their confidence zone, i.e., those who hold beliefs "close" to theirs. The most well-studied bounded confidence models are the Hegselmann-Krause model (Hegselmann and Krause 2002) and the Deffuant-Weisbuch model (Deffuant et al. 2000; Weisbuch et al. 2002).

– *Gossip models.* In a gossip model, at each step, each individual communicates with only one other individual chosen randomly from a given set of neighbors or the entire network. Then, the states are updated via a gossip protocol which is generally viewed as an averaging algorithm (Karp et al. 2000; Kempe and Kleinberg 2002).
– *Models with imperfect information exchange.* In practical cases, one has to account for noise and disturbance in the network. Remarkable works addressing the presence of noise include (Huang and Manton 2009) for the case of noisy communication links and Kashyap et al. (2007), Nedić et al. (2009), and El Chamie et al. for quantized data exchange. A gossip model in the presence of stubborn individuals, who do not update their beliefs over time, with its analysis can be found in Acemoğlu et al. (2013).
– *Models with time delays.* Communication delays in networks are inevitable. Averaging problems in models with time delays have been widely studied in literature. For instance, see Lin and Jia (2009) for second-order models with time delays and Sun and Wang (2009) for time-varying delays.

3 Continuous Opinion Dynamics

Again let $\mathcal{P} = \{\mathcal{P}_1, \ldots, \mathcal{P}_n\}$ denote the set of individuals in a social network and $\mathcal{A}_i = \mathbb{R}$ be the opinion set of individual $\mathcal{P}_i \in \mathcal{P}$. Construct a game played repeatedly among individuals in which each individual $\mathcal{P}_i$'s opinion, i.e., $x_i \in \mathbb{R}$, is viewed as her action. In this game, player $\mathcal{P}_i \in \mathcal{P}$ is associated with a time-invariant utility function $u_i(x_i, x_{-i}) : \mathbb{R}^n \to \mathbb{R}$, where:

$$x_{-i} \triangleq \{x_j \mid \mathcal{P}_j \in \mathcal{P}, j \neq i\}. \tag{21.6}$$

If she plays her best response strategy, the opinion dynamics of the social network can be described by:

$$x_i(t+1) = \arg\max_{x_i} u_i(x_i, x_{-i}(t)), \ t \geq 0. \tag{21.7}$$

The game above will be referred to as $(\mathcal{P}, x, u)$ in the rest of the section.

The utility function u_i of an individual $\mathcal{P}_i \in \mathcal{P}$ should be formulated in such a way to best capture her social behavior. We are specifically interested in *coordination games*, i.e., those games whose utility functions capture *conformity*, which refers to a type of social influence involving a change in belief or behavior in order to fit in with a group.

In the rest of the section, we formulate utilities u_is in two different ways, and, for each resulting game, we investigate the formation of opinions as time grows. In case of agreement or clustering, we also discuss the rate of convergence of opinions to their final limits.

3.1 Coordination Game with Stubborn Individuals

In a coordination game $(\mathcal{P}, x, u)$, let the utility function u_i of an individual $\mathcal{P}_i \in \mathcal{P}$ be given as (Ghaderi and Srikant 2013):

$$u_i(x_i, x_{-i}) = -\frac{1}{2}\sum_{\mathcal{P}_j \in N_i}(x_i - x_j)^2 - \frac{1}{2}K_i(x_i - x_i(0))^2, \tag{21.8}$$

where (i) $N_i \subset \mathcal{P}$ is the subset of individuals who interact with $\mathcal{P}_i$, and $\mathcal{P}_i \notin N_i$, (ii) the summation term captures *conformity* as a social behavior, and (iii) $K_i \geq 0$ indicates the *stubbornness* of $\mathcal{P}_i$ regarding her initial opinion. We will see that in the absence of stubbornness, i.e., $K_i = 0$, $\forall i$, the best response dynamics of the game with utilities (21.8) becomes a special case of averaging dynamics (21.4), while in the presence of stubbornness, it takes a non-averaging form of (21.1).

Let an undirected, unweighted graph $\mathcal{G}(\mathcal{P}, \mathcal{E})$, referred to as the *social graph*, represent the interaction structure of the social network, i.e., for every pair of individuals $\mathcal{P}_i, \mathcal{P}_j \in \mathcal{P}$:

$$e_{ij} \in \mathcal{E} \Leftrightarrow \mathcal{P}_j \in N_i. \tag{21.9}$$

For the best response strategy (21.7) with u_i defined by (21.8), we obtain:

$$x_i(t+1) = \frac{1}{d_i + K_i}\sum_{\mathcal{P}_j \in N_i} x_j(t) + \frac{K_i}{d_i + K_i}x_i(0), \tag{21.10}$$

where $d_i = |N_i|$ is the degree of node $\mathcal{P}_i$ in the social graph $\mathcal{G}$. We point out that, with a slight abuse of terminology, we also allow *infinite* stubbornness, i.e., $K_i = \infty$, for which the best response dynamics (21.10) reduces to $x_i(t+1) = x_i(0)$.

If we define a matrix $A \in \mathbb{R}^{n\times n}$ as:

$$A_{ij} \triangleq \begin{cases} \frac{1}{d_i + K_i} & \text{if } e_{ij} \in \mathcal{E}, \\ 0 & \text{otherwise,} \end{cases} \tag{21.11}$$

and a diagonal matrix $B \in \mathbb{R}^{n\times n}$ with $B_{ii} = \frac{K_i}{d_i + K_i}$ for every $\mathcal{P}_i \in \mathcal{P}$, then the best response strategies (21.10) of all players can be combined into matrix form as follows:

$$x(t+1) = Ax(t) + BX_0. \tag{21.12}$$

In the following, we state results from Ghaderi and Srikant (2013) on the existence of a Nash equilibrium for the coordination game above, and convergence of its best response dynamics, by considering two cases: (i) there are no stubborn individuals, and (ii) there are stubborn individuals.

No stubborn individuals. Suppose that for each individual $\mathcal{P}_i \in \mathcal{P}$, $K_i = 0$. In this case, $B = \mathbf{0}$ and A becomes row-stochastic since the sum of the elements of the ith row of A is $d_i.\frac{1}{d_i} = 1$. Thus, dynamics (21.12) converts to a time-invariant distributed averaging algorithm $x(t+1) = Ax(t)$ which can be used to characterize the asymptotic behavior of individual opinions as follows. Let us impose the following condition on the social graph $\mathcal{G}$ associated with the network:

Assumption 2. *Social graph $\mathcal{G}$ is connected and non-bipartite.*

We note that the connectedness of $\mathcal{G}$ can be assumed without any loss of generality since otherwise one would proceed with decomposing dynamics (21.12) into independent subdynamics associated with each connected component of $\mathcal{G}$. Furthermore, $\mathcal{G}$ is assumed to be non-bipartite to ensure *consistent influence* of each individual on one another. In other words, if $\mathcal{G}$ is connected and non-bipartite, there exists $t_0 \geq 0$ such that for each pair of individuals $\mathcal{P}_i, \mathcal{P}_j \in \mathcal{P}$, $x_i(t+t')$ depends on $x_j(t)$ for any $t' > t_0$ and t. Assumption 2 can also be interpreted in terms of the row-stochastic matrix A defined by (21.11). Indeed, $\mathcal{G}$ is connected if and only if A is *irreducible*, while $\mathcal{G}$ is non-bipartite if and only if A is *primitive*. Finally, we note that theorems from graph theory can be used to verify whether a given graph is bipartite or not. In particular, it is well known that a graph is bipartite if and only if it contains no odd cycles.

Under Assumption 2, an agreement among all individuals is reached as time increases, that is, all individuals' opinions asymptotically converge to the same value. The agreement equilibrium can be characterized as in the following theorem.

Theorem 1. *Consider game $(\mathcal{P}, x, u)$ with individuals' utilities defined by (21.8) and social graph $\mathcal{G}$ satisfying Assumption 2. If there are no stubborn individuals, i.e., $K_i = 0$ for every $\mathcal{P}_i \in \mathcal{P}$, then the best response dynamics (21.10) asymptotically converges to the following agreement equilibrium:*

$$\lim_{t\to\infty} x_i(t) = \frac{1}{2|\mathcal{E}|} \sum_{j=1}^{n} d_j x_j(0), \ \forall \mathcal{P}_i \in \mathcal{P}. \tag{21.13}$$

According to Theorem 1, each individual's impact on the agreement equilibrium is proportional to the number of her neighbors. We next address the rate of convergence to the equilibrium. Toward that goal, let us first provide an interpretation of the agreement equilibrium. View matrix A as the transition probability matrix of an irreducible random walk over $\mathcal{G}$ with edge weights equal to one. Stationary distribution $\pi \in \mathbb{R}^n$ of such a random walk is the normalized left eigenvector of A defined as in (21.11), associated with left eigenvalue 1, i.e.,

$$\pi_j = \frac{d_j}{2|\mathcal{E}|}, \ \mathcal{P}_j \in \mathcal{P}. \tag{21.14}$$

Thus, (21.13) can also be written as:

$$\lim_{t\to\infty} x(t) = \mathbf{1}\pi^{\mathrm{T}} X_0, \tag{21.15}$$

where **1** is the vector of all ones. Recalling the goal to address the rate of convergence in (21.15), let $l^2(\pi)$ denote the vector space $\mathbb{R}^n$ endowed with the scalar product:

$$< z, y >_\pi \triangleq \sum_{i=1}^{n} z_i y_i \pi_i,\ z, y \in \mathbb{R}^n. \tag{21.16}$$

Consequently, let norm $\|.\|_\pi$ be defined as:

$$\|z\|_\pi \triangleq \left(\sum_{i=1}^{n} z_i^2 \pi_i\right)^{1/2}. \tag{21.17}$$

Furthermore, define the error term $e(t)$ as:

$$e(t) \triangleq x(t) - \lim_{t\to\infty} x(t),\ t \geq 0. \tag{21.18}$$

Then, the following theorem states that $\|e(t)\|$ converges to 0 at a geometric rate.

Theorem 2. *Consider game $(\mathcal{P}, x, u)$ with individuals' utilities defined as (21.8) and social graph $\mathcal{G}$ satisfying Assumption 2. If there are no stubborn individuals, i.e., $K_i = 0$ for every $\mathcal{P}_i \in \mathcal{P}$, under best response strategies we have:*

$$\|e(t)\|_\pi \leq \rho_2^t \|e(0)\|_\pi, \tag{21.19}$$

where ρ_2 is the second largest eigenvalue modulus of A, i.e., $\rho_2 \triangleq \max(|\lambda_2|, |\lambda_n|)$ and $1 = \lambda_1 > \lambda_2 \geq \cdots \geq \lambda_n$ comprise the eigenvalues of A defined in (21.11).

Define also convergence time $\tau(\nu)$ for a small positive scalar ν as:

$$\tau(\nu) \triangleq \inf\{t \geq 0 \,|\, \|e(t)\| \leq \nu\}. \tag{21.20}$$

Based on Theorem 2, one can now conclude that:

$$\left(\frac{1}{1-\rho_2} - 1\right) \log\left(\frac{\|e(0)\|_\pi}{\nu}\right) \leq \tau(\nu) \leq \frac{1}{1-\rho_2} \log\left(\frac{\|e(0)\|_\pi}{\nu}\right). \tag{21.21}$$

In particular, the convergence time is $\Theta\left(\frac{1}{1-\rho_2}\right)$ as the number n of individuals grows.

Let us now address the case in which social graph $\mathcal{G}$ is bipartite but still connected. In this case, under the best response strategies, none of the individuals' opinions converge, except for some special choices of X_0, as time grows. To see why, we recall that $x(t) = A^t X_0$ for every $t \geq 0$ and show that $\lim_{t\to\infty} A^t$ does not exist. Let the set $\mathcal{P}$ of nodes be partitioned into nonempty subsets $\mathcal{P}^1$ and $\mathcal{P}^2$ such that there is no link between the nodes in the same subset. Equivalently, from (21.11), we must have $A_{ij} = 0$ if $\mathcal{P}_i, \mathcal{P}_j \in \mathcal{P}^1$ or $\mathcal{P}_i, \mathcal{P}_j \in \mathcal{P}^2$. This implies that for every odd integer $t > 0$, $(A^t)_{ij} = 0$ if $\mathcal{P}_i$ and $\mathcal{P}_j$ both belong to the same subset. In a similar fashion, one obtains that for every even integer $t > 0$, $(A^t)_{ij} = 0$ if $\mathcal{P}_i$ and $\mathcal{P}_j$ are in different subsets. Thus, if A^t is to converge as t grows, the limit must be $\mathbf{0}$, which is not possible since A^t is always row-stochastic.

The issue of not converging to an equilibrium for the case of bipartite $\mathcal{G}$ can be overcome assuming that a *noisy* alternative to best response strategy is played by every individual:

$$\hat{x}_i(t+1) \triangleq (1-\epsilon)\left(\frac{1}{d_i}\sum_{\mathcal{P}_j \in N_i} \hat{x}_j(t)\right) + \epsilon \hat{x}_i(t), \tag{21.22}$$

where $\epsilon > 0$ is a constant indicating *self-confidence* and is common among all individuals and $\hat{x}(0) \triangleq X_0$. The modified dynamics in essence means that every individual also accounts for her current opinion while updating. Therefore, if one defines matrix $\hat{A}$ as:

$$\hat{A}_{ij} \triangleq \begin{cases} \frac{1-\epsilon}{d_i} & \text{if } e_{ij} \in \mathcal{E}, \\ \epsilon & \text{if } i = j, \\ 0 & \text{otherwise}, \end{cases} \tag{21.23}$$

then (21.22) can be written as $\hat{x}(t+1) = \hat{A}\hat{x}(t), \forall t \geq 0$. Therefore, the noisy best response dynamics (21.22) converges to an equilibrium for which:

$$\lim_{t\to\infty} \hat{x}(t) = \mathbf{1}\hat{\pi}^{\mathrm{T}}\hat{x}(0) = \mathbf{1}\hat{\pi}^{\mathrm{T}} X_0, \tag{21.24}$$

where $\hat{\pi}$ is the unique stationary distribution of a Markov chain with transition probability matrix $\hat{A}$. Hence, similar to the non-bipartite case, all opinions converge to the same value as t grows. Notice also that, due to reversibility of $\hat{A}$, $\hat{\pi}$ is independent of ϵ and is in fact equal to the stationary distribution of a Markov chain with transition probability A, i.e., $\hat{\pi}_i = \frac{d_i}{2|\mathcal{E}|}$.

Furthermore, the convergence time to the equilibrium for the noisy best response dynamics is $\Theta\left(\frac{1}{1-\rho_2(\hat{A})}\right)$, where $\rho_2(\hat{A})$ denotes the second largest eigenvalue modulus of $\hat{A}$, i.e., $\rho_2(\hat{A}) = \max(\lambda_2(\hat{A}), |\lambda_n(\hat{A})|)$. It is to be noted that since $\hat{A} = \epsilon I + (1-\epsilon)A$, eigenvalues of $\hat{A}$ are derived as $\lambda_i(\hat{A}) = \epsilon + (1-\epsilon)\lambda_i(A)$. Thus:

$$\lambda_2(\hat{A}) = \epsilon + (1-\epsilon)\lambda_2(A), \tag{21.25}$$

and

$$\lambda_n(\hat{A}) = \epsilon + (1-\epsilon)\lambda_n(A) = -1 + 2\epsilon. \tag{21.26}$$

In the last equation, we used the fact that $\lambda_n(A) = -1$ since $\mathcal{G}$ is bipartite. Relations (21.25) and (21.26) together imply that the convergence time of the noisy best response dynamics to the equilibrium is $\Theta\left(\frac{1}{1-\lambda_2(A)}\right)$.

Finally, we note that although ϵ was assumed to be uniform among all individuals, the convergence to an equilibrium associated with a global agreement would remain intact even if that were not the case. However, the agreement value, as well as the convergence time to it, would also depend on the self-confidence indices of individuals.

Stubborn Individuals. Suppose now that the set $\mathcal{S} \subset \mathcal{P}$ of stubborn individuals is nonempty. Recall that an individual $\mathcal{P}_i \in \mathcal{P}$ is called stubborn if $K_i \neq 0$. Thus, according to (21.11), matrix A is substochastic, that is all its elements are nonnegative and none of its row sums exceeds 1, while at least one of the row sums is strictly less than 1. Thus, for every eigenvalue λ_i of A, we have $\|\lambda_i\| < 1$. Therefore, $\lim_{t\to\infty} A^t = \mathbf{0}$. Noticing that for every $t \geq 1$,

$$x(t) = \left(\sum_{s=0}^{t} A^s\right) BX_0, \tag{21.27}$$

we have the following theorem.

Theorem 3. *Consider game* $(\mathcal{P}, x, u)$ *with individuals' utilities defined as in* (21.8) *and let social graph* $\mathcal{G}$ *be connected. If there is at least one stubborn individual, the best response dynamics* (21.10) *asymptotically converges to the following equilibrium:*

$$\lim_{t\to\infty} x(t) = (I-A)^{-1} BX_0. \tag{21.28}$$

One proves Theorem 3 by simply taking (21.27) into account and using relation $\sum_{s=0}^{\infty} A^s = (I-A)^{-1}$. Notice that $\sum_{s=0}^{\infty} A^s$ converges since each eigenvalue λ_i of A satisfies $\|\lambda_i\| < 1$. Based on (21.28), the following conclusions can be drawn:

(i) Each individual's opinion converges to a convex combination of the initial opinions, as in the case of no stubborn individuals. In other words, $(I-A)^{-1}B$ is row-stochastic. To show this, one first uses induction on t to show that $\left(\sum_{s=0}^{t} A^s\right) B$ is row-stochastic for every $t > 0$. Then, the row-stochasticity of $(I-A)^{-1}B$ follows from the fact that the row-stochasticity property is preserved under the limit.

(ii) Only the initial opinions of stubborn individuals influence the equilibrium. To see why, it is sufficient to notice that the ith column of B is $\mathbf{0}$ if $\mathcal{P}_i$ is not stubborn.

From (21.28) and the conclusions drawn above, we know that for every $\mathcal{P}_i \in \mathcal{P}$:

$$\lim_{t\to\infty} x_i(t) = \sum_{\mathcal{P}_j \in \mathcal{S}} \alpha_{ij} x_j(0), \tag{21.29}$$

where $\alpha_{ij} \geq 0$ for every $\mathcal{P}_j \in \mathcal{S}$ and $\sum_{\mathcal{P}_j \in \mathcal{S}} \alpha_{ij} = 1$. In what follows, an interpretation of coefficients α_{ij}s in (21.29) is presented via a random walk over a weighted graph $\hat{G}$ built based on the social graph $\mathcal{G}$.

Let $\mathcal{S}_1 \subset \mathcal{S}$ contain those stubborn individual whose stubbornness indices are finite, i.e.,

$$\mathcal{P}_i \in \mathcal{S}_1 \;\Leftrightarrow\; 0 < K_i < \infty. \tag{21.30}$$

Moreover, let $\mathcal{S}_2 \triangleq \mathcal{S} \backslash \mathcal{S}_1$ contain those stubborn individuals with infinite stubbornness. Without loss of generality, assume that:

$$\begin{aligned} \mathcal{S}_1 &= \{\mathcal{P}_1, \ldots, \mathcal{P}_{|\mathcal{S}_1|}\}, \\ \mathcal{S}_2 &= \{\mathcal{P}_{|\mathcal{S}_1|+1}, \ldots, \mathcal{P}_{|\mathcal{S}|}\} \end{aligned} \tag{21.31}$$

Consequently, $\mathcal{P} \backslash \mathcal{S} = \{\mathcal{P}_{|\mathcal{S}|+1}, \ldots, \mathcal{P}_n\}$. Construct now weighted graph $\hat{\mathcal{G}}(\hat{\mathcal{P}}, \hat{\mathcal{E}}, w)$ from the social graph $\mathcal{G}(\mathcal{P}, \mathcal{E})$ as follows: add $|\mathcal{S}_1|$ nodes to $\mathcal{G}$, namely, $\mathcal{P}_{n+1}, \ldots, \mathcal{P}_{n+|\mathcal{S}_1|}$, and connect $\mathcal{P}_i$ to $\mathcal{P}_{n+i}$ for every $\mathcal{P}_i \in \mathcal{S}_1$. Then, we have:

$$\hat{\mathcal{P}} = \mathcal{P} \cup \{\mathcal{P}_{n+i} \,|\, \mathcal{P}_i \in \mathcal{S}_1\}, \tag{21.32}$$

$$\hat{\mathcal{E}} = \mathcal{E} \cup \{e_{i,n+i} \,|\, \mathcal{P}_i \in \mathcal{S}_1\}. \tag{21.33}$$

Furthermore, keeping (21.33) in mind, define weight w_{ij} of an arbitrary edge $e_{ij} \in \hat{\mathcal{E}}$ as:

$$w_{ij} \triangleq \begin{cases} 1 & \text{if } e_{ij} \in \mathcal{E} \\ K_i & \text{if } j = n + i \text{ and } \mathcal{P}_i \in \mathcal{S}_1. \end{cases} \tag{21.34}$$

Consider a random walk Y over $\hat{\mathcal{G}}$ for which the probability p_{ij} of transition from $\mathcal{P}_i \in \hat{\mathcal{P}}$ to $\mathcal{P}_j \in \hat{\mathcal{P}}$ is defined as:

$$p_{ij} \triangleq \frac{w_{ij}}{\sum_{e_{ik} \in \hat{\mathcal{E}}} w_{ik}}. \tag{21.35}$$

We recall the goal to interpret α_{ij}, $\mathcal{P}_i \in \mathcal{P}$, $\mathcal{P}_j \in \mathcal{S} = \mathcal{S}_1 \cup \mathcal{S}_2$. Let the random walk start at an arbitrary but fixed $\mathcal{P}_i \in \mathcal{P}$. It is clear that the walk hits the following set in a finite time with probability 1:

$$\mathcal{S}_2 \cup (\hat{\mathcal{P}} \backslash \mathcal{P}) = \{\mathcal{P}_{|\mathcal{S}_1|+1}, \ldots, \mathcal{P}_{|\mathcal{S}|}, \mathcal{P}_{n+1}, \ldots, \mathcal{P}_{n+|\mathcal{S}_1|}\}. \tag{21.36}$$

Then, α_{ij}, $\mathcal{P}_j \in \mathcal{S}_2$, is the probability that $\mathcal{P}_j$ is the first node within $\mathcal{S}_2 \cup (\hat{\mathcal{P}} \backslash \mathcal{P})$ to be hit by the walk. Moreover, for $\mathcal{P}_j \in \mathcal{S}_1$, α_{ij} is the probability that $\mathcal{P}_{n+j}$ is the first to be hit within the same set.

To address the rate of convergence to the equilibrium in (21.28), let the error term $e(t)$ again be defined as in (21.18). The following theorem addresses the convergence rate that turns out to be geometric with a rate at least equal to the largest eigenvalue of A.

Theorem 4. *Consider game* $(\mathcal{P}, x, u)$ *with individuals' utilities given by* (21.8) *and let the social graph* $\mathcal{G}$ *be connected. If there is at least one stubborn individual, under the best response strategies* (21.10)*, we have:*

$$\|e(t)\|_2 \leq c\lambda_1^t \|e(0)\|, \tag{21.37}$$

where $c > 0$ *is constant and* λ_1 *is the largest eigenvalue of* A.

To recap, in this subsection, we have employed a game theoretic approach to model concurrently conformity and stubbornness. The opinions (players' actions) were assumed to be continuous in nature, as they might take any value in interval $[0, 1]$. The utility of each player, defined by (21.8), contains two terms: one that captures conformity and another one capturing stubbornness. It was shown that in the absence of stubborn individuals, a global agreement is asymptotically achieved at a geometric rate if the social graph of the network is connected, with an exception of bipartite social graphs. For the case where the social graph is bipartite, each individual's opinion may not converge although the number of accumulation points of opinions can be proved to be finite. In the presence of at least one stubborn individual, however, the convergence at a geometric rate of each individual's opinion, as time grows, is proved for all social graphs, but the occurrence of a global agreement is nearly impossible. In fact, even if the social graph is connected, the existence of at least two stubborn individuals with different initial opinions would eventually lead to a grouping of the individuals into multiple clusters.

3.2 Coordination Game with Bounded Confidence

In bounded confidence models, an individual communicates with only those who lie in her area of confidence. Such models were introduced independently in Hegselmann and Krause (2002), the so-called Hegselmann-Krause (HK) model, and in Deffuant et al. (2000) and Weisbuch et al. (2002), the so-called Deffuant-Weisbuch (DW) model. A thorough survey on both models and their extensions can be found in Lorenz and Urbig (2007).

In the well-received HK model (Hegselmann and Krause 2002), it is assumed that opinion x_i of an individual $\mathcal{P}_i \in \mathcal{P}$ evolves according to the following update equation:

$$x_i(t+1) = \frac{1}{n_i(t)} \sum_{\mathcal{P}_j \in \mathcal{N}_i(t)} x_j(t), \ t \geq 0, \tag{21.38}$$

where $n_i(t) = |\mathcal{N}_i(t)|$, i.e., the number of players in $\mathcal{N}_i(t)$, and $\mathcal{N}_i(t)$ denotes the neighbor set of individual $\mathcal{P}_i$ at time t, which is defined as:

$$\mathcal{N}_i(t) \triangleq \{\mathcal{P}_j \in \mathcal{P} \mid \|x_j(t) - x_i(t)\| \leq R\}, \tag{21.39}$$

where $R > 0$ is a fixed, common *confidence bound*. Roughly speaking, each individual modifies her opinion to the arithmetic mean of the current opinions of her neighbors, i.e., those opinions which are not too far from hers. Notice that, according to (21.39), each individual belongs to her neighbor set at all times.

Again, let $X_0 \in \mathbb{R}^n$ represent the vector of initial opinions. It is known that, under the dynamics (21.38), an individual limit x_i^* for each individual's opinion $x_i(t)$ exists and the convergence occurs in a finite time. Thus, depending on vector X_0 of initial opinions and confidence bound R, the individuals are split into one or several clusters, with those in the same cluster reaching an agreement, i.e., sharing a common limiting opinion.

The classical HK model lies within the category of *synchronous* models, since the individual opinions are updated simultaneously. An *asynchronous* version of the HK model would be obtained if only one, randomly selected, individual at a time was able to update her opinion while again using protocol (21.38) to do so. The asynchronous versions of the HK model demonstrate quite different behaviors when compared to their synchronous counterparts; most notably, their convergence times are infinite unless a global agreement had already been in place before the updates started. The steady-state behavior of asynchronous HK dynamics (21.38) is of interest and will be addressed in the following via a game theoretic approach (Etesami and Başar 2015).

In Etesami and Başar (2015), the asynchronous HK model was redesigned as a game $(\mathcal{P}, x, u)$ for which the opinion set $\mathcal{A}_i$ of each individual is $\mathbb{R}^d$, where $d \geq 1$ is an integer, and utility $u_i : (\mathbb{R}^d)^n \to \mathbb{R}$ of a player $\mathcal{P}_i \in \mathcal{P}$ is defined as:

$$u_i(x_i, x_{-i}) = (n-1)R^2 - \sum_{j=1}^{n} \min\{\|x_i - x_j\|^2, R^2\}, \tag{21.40}$$

where $R > 0$ is the fixed common confidence bound as introduced earlier. The game is played repeatedly in the following manner: at each discrete time $t \geq 0$, one player is selected at random, who then updates her action. It is assumed that player $\mathcal{P}_i$ is aware of the action of player $\mathcal{P}_j$ at time t if and only if $\|x_i(t) - x_j(t)\| \leq R$. Assuming that $\mathcal{P}_i \in \mathcal{P}$ is selected at time t to update her action, it turns out that strategy (21.38) serves as a *better*, although not necessarily the *best*, response strategy for her in the following sense: unless the game is at a Nash equilibrium, she strictly benefits from adopting strategy (21.38), meaning that her utility would increase if she updated her action via (21.38). Indeed, utility (21.40) is designed in such a way that strategy (21.38) becomes a better response strategy. Under the better response strategy (21.38), the game is now expected to converge to a Nash equilibrium. Before confirming that, let us first characterize the Nash equilibria of the game considered.

Proposition 1. *For the asynchronous game* $(\mathcal{P}, x, u)$*, where the action set for every player is* $\mathbb{R}^d$ *and utilities are defined as* (21.40)*, an action profile* $(x_1^*, \ldots, x_n^*)$ *is a Nash equilibrium if and only if for every pair* i, j*,* $1 \leq i, j \leq n$*, either* $x_i^* = x_j^*$ *or* $\|x_i^* - x_j^*\| > R$*.*

The game considered above is a *potential game* which is defined below:

Definition 4 (potential game). A game $(\mathcal{P}, x, u)$ with action sets $\mathcal{A}_i$, $1 \leq i \leq n$, and utility functions $u_i : \mathcal{A} \to \mathbb{R}$ constitute a *potential game* if there exists a function $\phi : \mathcal{A} \to \mathbb{R}$ such that:

$$u_i(x_i'', x_{-i}) - u_i(x_i', x_{-i}) = \phi(x_i'', x_{-i}) - \phi(x_i', x_{-i}), \tag{21.41}$$

for every player $\mathcal{P}_i \in \mathcal{P}$, all actions $x_i'', x_i' \in \mathcal{A}_i$, and all $x_{-i} \in \prod_{j \neq i} \mathcal{A}_j$. The function ϕ is then called a *potential function* associated with the game $(\mathcal{P}, x, u)$.

Lemma 1. *Asynchronous game* $(\mathcal{P}, x, u)$*, where action set is* $\mathbb{R}^d$ *and utilities are defined as* (21.40)*, is a potential game with* $\phi : (\mathbb{R}^d)^n \to \mathbb{R}$ *defined as:*

$$\phi(x) = \phi(x_1, \ldots, x_n) \triangleq \frac{1}{2} \sum_{i=1}^{n} u_i(x_i, x_{-i}). \tag{21.42}$$

In other words,

$$\phi(x_i', x_{-i}) - \phi(x_i'', x_{-i}) = u_i(x_i', x_{-i}) - u_i(x_i'', x_{-i}), \tag{21.43}$$

for every $\mathcal{P}_i \in \mathcal{P}$*,* $x_i' x_i'' \in \mathbb{R}^d$*, and* $x_{-i} \in (\mathbb{R}^d)^{n-1}$*.*

From the discussion above, we know that playing the better response strategy (21.38), potential function ϕ increases in value unless the game is at a Nash equilibrium. The following lemma addresses the amount of increase in ϕ at each time instant, which is crucial to prove the convergence of the game to a Nash equilibrium.

Lemma 2. *For the asynchronous game* $(\mathcal{P}, x, u)$*, where the action set for every player is* $\mathbb{R}^d$ *and utilities are defined as in* (21.40)*, under strategy* (21.38)*, if player* i *is selected to update at time* $t \geq 0$*, we have:*

$$\phi\big(x_i(t+1), x_{-i}(t+1)\big) - \phi\big(x_i(t), x_{-i}(t)\big) \geq 2n_i \, \|x_i(t+1) - x_i(t)\|^2. \qquad (21.44)$$

Consequently:

$$\begin{aligned} &\mathbb{E}\Big\{\phi\big(x(t+1)\big) - \phi\big(x(t)\big)\Big\} \\ &\geq \tfrac{2}{n} \textstyle\sum_{j=1}^{n} n_i \, \mathbb{E}\Big\{\|x_i(t+1) - x_i(t)\|^2\Big\} \\ &\geq \tfrac{2}{n} \textstyle\sum_{j=1}^{n} \, \mathbb{E}\Big\{\|x_i(t+1) - x_i(t)\|\Big\}^2. \end{aligned} \qquad (21.45)$$

Inequality (21.45) can comfortably be used to show that the game indeed converges to a Nash equilibrium. Since we are also interested in the rate of convergence to the equilibrium, we define a δ-Nash equilibrium of the considered game as follows:

Definition 5. Consider the asynchronous game $(\mathcal{P}, x, u)$, where the action set for each player is $\mathbb{R}^d$ and utilities are given by (21.40). Given $\delta > 0$, an action profile $(x_1^*, \ldots, x_n^*)$ is said to be a δ-Nash equilibrium if for a partitioning $(C_1, \ldots, C_m)$ of $\mathcal{P}$, we have:

(i) $\text{diam}\big(\text{conv}(C_k)\big) < \delta$, $\forall k$, $1 \leq k \leq m$, where $\text{conv}(.) : \mathbb{R}^d \to \mathbb{R}^d$ stands for the *convex hull* of a subset of $\mathbb{R}^d$ and $\text{diam}(.) : \mathbb{R}^d \to$ denotes the *diameter* of a subset of $\mathbb{R}^d$, i.e., the longest distance between two points in the subset;
(ii) $\text{dist}\big(\text{conv}(C_i), \text{conv}(C_j)\big) > \delta$, $\forall i \neq j$, $1 \leq i, j \leq m$, where $\text{dist}(.,.) : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ denotes the *distance* between two subsets of $\mathbb{R}^d$, which is the longest possible distance between two points in different subsets.

Theorem 5. *For the asynchronous game* $(\mathcal{P}, x, u)$*, where the action set for each player is* $\mathbb{R}^d$ *and utilities are defined as in* (21.40)*, under strategy* (21.38)*, we have:*

(i) *The game converges to a Nash equilibrium, where the set of all Nash equilibria of the game is characterized in Proposition 1;*
(ii) *The expected number of steps for the game to reach a* δ*-Nash equilibrium is upper bounded by* $2n^9(R/\delta)^2$.

To summarize, in this subsection, by employing game theory, we have concluded that for the given asynchronous HK dynamics (21.40), $\lim_{t\to\infty} x(t)$ exists, which means that clustering occurs. The steady-state opinions of those in the same cluster are equal by definition. Moreover, according to Proposition 1, for any two individuals in different clusters, their steady-state opinions differ by at least R, where R is the confidence bound. Furthermore, if we define T_δ as the smallest time such that $\|x(t) - \lim_{t' \to \infty} x(t')\| < \delta$, for all $t \geq T_\delta$, we have $\mathbb{E}[T_\delta] \leq 2n^9(R/\delta)^2$.

4 Discrete Opinion Dynamics

4.1 Coordination Via a Potential Game Model

In this subsection, we discuss a potential game approach from Marden et al. (2009) inspired by the pioneering work (DeGroot 1974) on opinion dynamics. In DeGroot (1974), DeGroot considers a committee of individuals, each holding an initial opinion about the probability distribution of an unknown parameter, seeking a consensus. In the model, an interaction graph, which is fixed, weighted, and undirected, is assigned to the committee, based on which the committee members' opinions evolve. More specifically, at each time instant, each member updates her opinion to a convex combination of her neighbors' opinions, where the weights in the combination coincide with the weights in the graph. It is well known that all the committee members eventually agree on a common distribution if and only if the interaction graph is connected.

In the game theoretic framework introduced in Marden et al. (2009), a game $(\mathcal{P}, x, u)$ is considered in which each player $\mathcal{P}_i \in \mathcal{P}$ takes an action from a predefined, finite action set $\mathcal{A}_i \subset \mathbb{R}^d$, where d is a positive integer. Furthermore, each player $\mathcal{P}_i \in \mathcal{P}$ is assigned a utility function $u_i : \mathcal{A} \to \mathbb{R}$, with $\mathcal{A} = \prod_{i=1}^n \mathcal{A}_i$, given by:

$$u_i(x_i, x_{-i}) = - \sum_{\mathcal{P}_j \in \mathcal{N}_i} \|x_i - x_j\|, \tag{21.46}$$

where $\mathcal{N}_i$ denotes the neighbor set of player $\mathcal{P}_i$ in the social graph $\mathcal{G}$, which is a fixed, unweighted, and undirected.

Assume that each individual at every time step adopts her best response strategy, i.e., takes an action in such a way to maximize her utility. Notice that the dynamics derived by playing the above game differs from the traditional averaging dynamics in two ways: (i) the action sets are finite in the game, whereas in averaging dynamics, the opinions can virtually take any real value, (ii) in the game defined, there is a lack of self-confidence, that is, an individual's current opinion does not directly influence her opinion at the next time instant, while in averaging dynamics, the opposite is generally true.

Recalling Definition 4 of potential games, it can be shown that the game defined above is a potential game with potential function $\phi : \mathcal{A} \to \mathbb{R}$ defined as:

$$\phi(x_1, \ldots, x_n) \triangleq -\frac{1}{2} \sum_{i=1}^{n} \sum_{\mathcal{P}_j \in \mathcal{N}_i} \|x_i - x_j\|. \tag{21.47}$$

Let us also assume that the action sets $\mathcal{A}_i$s have nonempty intersection, meaning that an agreement is possible. It is then clear that any agreement maximizes the potential function. Furthermore, any joint action profiles maximizing the potential function have to constitute an agreement if $\mathcal{G}$ is connected. Two strategies for the players resulting in an agreement are discussed next.

Spatial Adaptive Play (SAP). Assume that the game is played repeatedly over time and each player adopts a mixed strategy called *Spatial Adaptive Play* (SAP) (Young 2001), which is reminiscent of a Gibbs-based exponential learning algorithm, described as follows: each player $\mathcal{P}_i \in \mathcal{P}$ at time t takes an action $x_i \in \mathcal{A}_i$ with probability $p_i^{x_i}(t)$ defined as:

$$p_i^{x_i}(t) \triangleq \frac{\exp\{\beta u_i(x_i, x_{-i}(t))\}}{\sum_{\bar{x}_i \in \mathcal{A}_i} \exp\{\beta u_i(\bar{x}_i, \bar{x}_{-i}(t))\}}, \tag{21.48}$$

where $\beta \geq 0$ is the *exploration parameter*. The larger β is, the more likely it is for a player $\mathcal{P}_i$ to take an action from her best response set with respect to her utility function $u_i(x_i, x_{-i})$. The following theorem addresses the stationary distribution of the joint action profiles under the SAP.

Theorem 6 (Young 2001). *In a potential game where players adopt the SAP* (21.48) *as their strategies, stationary distribution* $\mu : \mathcal{A} \to [0, 1]$ *of the joint action profiles is given by:*

$$\mu(x) = \frac{\exp\{\beta\phi(x)\}}{\sum_{\bar{x} \in \mathcal{A}} \exp\{\beta\phi(\bar{x})\}}. \tag{21.49}$$

In other words, as $t \to \infty$, $x(t)$ approaches x with probability $\mu(x)$ for any $x \in \mathcal{A}$. Consequently, if β is sufficiently large, $\mu(x)$ is arbitrarily small for those joint action profiles which do not maximize the potential function. Thus, we have the following corollary.

Corollary 1. *If* $\mathcal{G}$ *is connected, an agreement is asymptotically reached with an arbitrarily high probability, provided that* β *is sufficiently large.*

Restricted Spatial Adaptive Play (RSAP). From a social network perspective, it could be reasonably argued that a player's action at any time is correlated with her previous action(s) which is an issue not accounted for in the setting discussed above. An attempt to resolve this issue leads to a more realistic setting where a player $\mathcal{P}_i \in \mathcal{P}$ takes an action x_i with a non-zero probability at time t only if

$x_i \in R_i(x_i(t-1))$, where $R_i : \mathcal{A}_i \to 2^{\mathcal{A}_i}$ and $2^{\mathcal{A}_i}$ denotes the set of all subsets of $\mathcal{A}_i$. Thus, $R_i(x_i(t-1))$ is a *restricted* action set within $\mathcal{A}_i$. The following reasonable assumptions are made on R_i, for every player $\mathcal{P}_i \in \mathcal{P}$:

1. For each $x_i \in \mathcal{A}_i$: $x_i \in R_i(x_i)$;
2. For each pair of actions $x_i^1, x_i^2 \in \mathcal{A}_i$: $x_i^2 \in R_i(x_i^1) \Leftrightarrow x_i^1 \in R_i(x_i^2)$;
3. For each pair of actions $x_i^0, x_i^k \in \mathcal{A}_i$, there exists a sequence of actions $x_i^0, x_i^1, \ldots, x_i^k$ satisfying $x_i^{k'} \in R_i(x_i^{k'-1})$ for all $k' = 1, \ldots, k$.

A modified version of SAP, the so-called Restricted Spatial Adaptive Play (RSAP), can be described as follows: at each time t, a player, say $\mathcal{P}_i \in \mathcal{P}$, is selected at random to update her action. She first selects a *trial* action from within $R_i(x_i(t-1))$, excluding her previous action, with probability $1/m_i$, where $m_i \triangleq \max_{x_i \in \mathcal{A}_i} |R_i(x_i)|$. Her trial action would be her previous action with probability $1 - \frac{|R_i(x_i(t-1))|-1}{m_i}$. She then updates her action to the selected trial action, say $\hat{x}_i$, with the following probability:

$$\frac{\exp\{\beta u_i(\hat{x}_i, x_{-i}(t-1))\}}{\exp\{\beta u_i(\hat{x}_i, x_{-i}(t-1))\} + \exp\{\beta u_i(x(t-1))\}}, \tag{21.50}$$

where $\beta \geq 0$ is the exploration parameter, and repeats her previous action $x_i(t-1)$ with probability one minus the probability (21.50), i.e.,

$$\frac{\exp\{\beta u_i(x(t-1))\}}{\exp\{\beta u_i(\hat{x}_i, x_{-i}(t-1))\} + \exp\{\beta u_i(x(t-1))\}}. \tag{21.51}$$

The following theorem addresses the stationary distribution of the joint action profiles under the RSAP, which happens to be equal to that under SAP.

Theorem 7. *In a potential game where players adopt the RSAP* (21.50) *and* (21.51) *as their strategies, stationary distribution* $\mu : \mathcal{A} \to [0, 1]$ *of the joint action profiles is given by* (21.49).

Hence, similar to the discussion on the SAP, given $\mathcal{G}$ is connected and β is sufficiently large, an agreement is achieved with an arbitrarily high probability.

Thus, to summarize, we have employed a game theoretic approach in this subsection to model conformity in social networks and to investigate when a global agreement would be achieved. The opinions (players' actions) were assumed to be discrete, i.e., they could take only a finite number of values in $\mathbb{R}^d$. Two strategies, namely, SAP and RSAP, were introduced, and it was made clear that provided that the social graph of the network is connected, a global agreement is asymptotically achieved with high probability if it is feasible in the sense that the action sets have a nonempty intersection.

4.2 Coordination Game: Innovation Spread

Coordination games can also be used to model the spread of innovations throughout a social network. One wonders, for instance, how the use of smartphones grew to become so widespread as it is today. Addressing this problem is of great importance, for smartphone companies in particular, since it helps predict the spread. To investigate the spread of an innovation, we view the individuals as players of a coordination game. Each individual has two possible actions: whether to adopt the innovation or not. The players' utilities are then defined in such a way that the adoption of a common strategy among the individuals results in higher utilities. Once the game and its specifications are defined formally, we shall examine whether the players' strategies converge to an equilibrium in time. Moreover, in order to predict finite time behavior of the social network, the rate of convergence to the equilibrium, if it exists, is of interest.

In this subsection, we cover two outstanding work Ellison (1993) and Montanari and Saberi (2009) whose game theoretic frameworks are inspired by pioneering work (Kandori et al. 1993). More recent game models on the subject of innovation spread and competitive diffusion can be found in Goyal et al. (2014) and Etesami and Başar (2016).

In Ellison (1993), a game $(\mathcal{P}, x, u)$, where $x_i \in \{0, 1\}$ for every $\mathcal{P}_i \in \mathcal{P}$, is considered. Let the utilities be defined as:

$$u_i(x_i, x_{-i}) = \sum_{\substack{\mathcal{P}_j \in \mathcal{P} \\ \mathcal{P}_j \neq \mathcal{P}_i}} w_{ij}\, g(x_i, x_j), \tag{21.52}$$

where $W = [w_{ij}] \in \mathbb{R}^{n\times n}$ is a symmetric weight matrix and payoffs g are determined according to the underlying 2×2 coordination game in Table 21.1.

Before stating the player strategies, let us further assume that $a > d$ and $b > c$ so that $(0, 0)$ and $(1,1)$ are the Nash equilibria of the underlying 2×2 game. Moreover, let $(a - d) > (b - c)$ so that $(0, 0)$ is the *risk dominant* equilibrium (Harsanyi et al. 1988). Therefore, when the strategies have the same security level, i.e., $c = d$, the equilibrium $(0, 0)$ is also the Pareto optimum.

One can view matrix W as the adjacency matrix of the weighted social graph of the network, which is henceforth assumed to be undirected. Higher weights indicate larger influences. Incorporating the weight matrix would also enable us to address *random matching* games within the same framework. In a random matching game, players $\mathcal{P}_i, \mathcal{P}_j \in \mathcal{P}$ are matched at a given time instant with probability w_{ij}.

Table 21.1 Underlying 2×2 coordination game

	0	1
0	a, a	c, d
1	d, c	b, b

Assume now that game $(\mathcal{P}, x, u)$ described above is played synchronously by all individuals over time. We pursue the following two strategies:

1. *Best response strategy*: every player $\mathcal{P}_i \in \mathcal{P}$ at time $t+1$ plays

$$\arg\max_{x_i \in \{0,1\}} u_i(x_i, x_{-i}(t)), \tag{21.53}$$

   for any $t \geq 0$.
2. *Noisy best response strategy*: every player $\mathcal{P}_i$ at time $t+1$ plays her best response strategy (21.53) with probability $1-2\epsilon$, while with probability 2ϵ she plays 0 or 1 randomly, where $0 < \epsilon < \frac{1}{2}$ is fixed. Consequently, at each time instant, she plays her best response strategy with probability $1-\epsilon$ and the other one with probability ϵ.

We shall focus particularly on the following cases representing two specific types of the social graph:

1. *Uniform matching rule*: Let the social graph of the network be a *complete* graph, i.e., w_{ij} be non-zero for every pair $\mathcal{P}_i, \mathcal{P}_j \in \mathcal{P}$. More precisely, assume that

$$w_{ij} = \frac{1}{n-1}, \ \forall \mathcal{P}_i, \mathcal{P}_j \in \mathcal{P}. \tag{21.54}$$

2. *Local (2k-neighbor) matching rule*: Let the social graph of the network be a $2k$*-regular* graph, with $1 \leq k \leq [(n-1)/2]$, in which

$$w_{ij} = \begin{cases} \frac{1}{2k} & \text{if } i - j = \pm 1, \pm 2, \ldots, \pm k \pmod{n}, \\ 0 & \text{otherwise.} \end{cases} \tag{21.55}$$

   This graph corresponds to the $(2k, n)$-Harary graph (Harary 1962). In particular, for the case $k = 1$, i.e., the so-called 2-neighbor matching rule, the above graph turns to the *ring* graph.

In the following, we discuss the limiting behavior of the game under best response strategies and noisy best response strategies, in particular for uniform and local matching rules. Our analysis mainly relies on the evolution of the number of players who take action 0. Thus, we shall use the following notation in the reminder of this subsection. Given an action profile $x \in \{0,1\}^n$, we denote by x^{zero} the number of zero elements of x. Subsequently, $x^{\text{zero}}(t)$ denotes the number of players who play 0 at time t. Obviously, $x^{\text{zero}}(t)$ would be a random variable if $x(t)$ were a random variable, as under noisy best response strategies. We also denote by $q_i(t)$, $0 \leq q_i(t) \leq 1$, the fraction of a player $\mathcal{P}_i$'s neighbors playing 0 at a given time t.

Best response strategies. Assuming that $x(0) = X_0 \in \{0,1\}^n$ is given, action profile $x(t)$ is deterministic for each $t \geq 0$ if the best response strategies are played. Let us first consider the uniform matching rule. In view of best response strategy (21.53), for $\mathcal{P}_i$ to play 0 at time $t+1$, the following must hold:

$$u_i(0, x_{-i}(t)) \geq u_i(1, x_{-i}(t)), \tag{21.56}$$

which is equivalent to:

$$q_i(t) \geq \frac{b-c}{(a-d)+(b-c)} \triangleq q^*. \tag{21.57}$$

If condition (21.57) does not hold, $\mathcal{P}_i$ plays 1 at time $t+1$. Notice that we assumed that a player plays 0 when she is indifferent between 0 and 1. Notice also that $q^* < \frac{1}{2}$ since we assumed $(a-d) > (b-c)$, i.e., $(0,0)$ is the risk dominant equilibrium. Recalling that X_0^{zero} is the number of players who play 0 initially, we have:

$$q_i(0) = \begin{cases} \frac{X_0^{\text{zero}}-1}{n-1} & \text{if } x_i(0) = 0, \\ \frac{X_0^{\text{zero}}}{n-1} & \text{if } x_i(0) = 1. \end{cases} \tag{21.58}$$

From (21.58) and condition (21.57) for playing 0, one concludes that at time $t=1$: (i) every player plays 0 if $\frac{X_0^{\text{zero}}-1}{n-1} \geq q^*$, (ii) every player switches her action if $\frac{X_0^{\text{zero}}}{n-1} \geq q^* > \frac{X_0^{\text{zero}}-1}{n-1}$, and (iii) every player plays 1 if $q^* > \frac{X_0^{\text{zero}}}{n-1}$. In cases (i) and (iii), no player will ever change her action, and the strategies have thus converged to an equilibrium where all actions are the same. In case (ii), however, the strategies are yet to converge. Notice that in case (ii), $x^{\text{zero}}(1) = n - X_0^{\text{zero}}$. Therefore, by a similar argument, the strategies converge at time $t=2$ unless $\frac{n-X_0^{\text{zero}}}{n-1} \geq q^* > \frac{n-X_0^{\text{zero}}-1}{n-1}$. Recalling that $q^* < \frac{1}{2}$ and taking the following two inequalities into account,

$$\begin{aligned} \frac{X_0^{\text{zero}}}{n-1} &\geq q^* > \frac{X_0^{\text{zero}}-1}{n-1}, \\ \frac{n-X_0^{\text{zero}}}{n-1} &\geq q^* > \frac{n-X_0^{\text{zero}}-1}{n-1}, \end{aligned} \tag{21.59}$$

we conclude that:

$$\begin{aligned} X_0^{\text{zero}} &= \frac{n}{2}, \\ \frac{1}{2} - \frac{1}{2(n-1)} &< q^* < \frac{1}{2}. \end{aligned} \tag{21.60}$$

Hence, the strategies converge to an equilibrium at $t=1$ unless the two conditions of (21.60) hold, for which the strategies will never converge. More precisely, under (21.60), every player keeps on switching her action at each time instant $t > 0$.

We now consider the local matching rule (21.55) with $k=1$, i.e., the case where the social graph is a ring graph. We investigate the asymptotic behavior of the actions by analyzing the evolution of $x^{\text{zero}}(t)$ as t grows. Assume first that n is odd. It is easy to show that:

$$\begin{cases} x^{\text{zero}}(t+1) = x^{\text{zero}}(t) & \text{if } x^{\text{zero}}(t) = 0 \text{ or } n, \\ x^{\text{zero}}(t+1) > x^{\text{zero}}(t) & \text{otherwise.} \end{cases} \tag{21.61}$$

From (21.61), we conclude that $x^{\text{zero}}(t)$ converges after at most n time steps to an equilibrium where all actions are the same. A crucial difference which an even n makes is that the social graph becomes bipartite, leading to the possibility of non-convergence of the actions. For instance, at time $t = 0$, let all the players in one part of the social graph play 0 at $t = 0$, while the players in the other part play 1. Then, according to the dynamics of the game, every player will switch her action at each time instant. The limiting behavior of the actions can again be investigated via the evolution of $x^{\text{zero}}(t)$. One can see that at an arbitrary $t \geq 0$, $x^{\text{zero}}(t+1) > x^{\text{zero}}(t)$ unless the set of players with action 0 at time t form an empty set, either part of the bipartite social graph or the entire set of players. Therefore, after at most n time steps, the action profile either converges to an equilibrium, in which all actions are the same, or enters a limit cycle, in which the players in the same part of the bipartite social graph take the same action while switching their actions at each time instant.

For a local matching rule with $k > 1$, the analysis is slightly different, though it can once again be shown that the game reaches an equilibrium or enters a limit cycle in a finite time.

Noisy best response strategies. The inclusion of the noise in player strategies turns out to help the convergence of the game dynamics to an equilibrium. Let the weight matrix W, i.e., the matching rule, be given and ϵ, $0 < \epsilon < 1$ be arbitrary but fixed. Under the noisy strategies, the evolution of $x(t)$, which is now a random variable over $\{0,1\}^n$, can be viewed as a Markov chain. Let $s_1, \ldots, s_{2^n}$ be all pure strategy profiles. Then, there is a fixed transition matrix $P = [p_{ij}]$ of size $2^n \times 2^n$ such that for every $t \geq 0$ and every pair of pure strategy profiles $s_i, s_j \in \{0,1\}^n$:

$$\mathbb{P}\big[x(t+1) = s_j \mid x(t) = s_i\big] = p_{ij}. \tag{21.62}$$

We note that the Markov chain $\{x(t)\}$ is irreducible since $p_{ij} > 0, 1 \leq i, \, and \, j \leq 2^n$. In fact, $p_{ij} \geq \min\big(\epsilon^n, (1-\epsilon)^n\big)$. Thus, the distribution of $x(t)$ over $\{0,1\}^n$ converges as t grows. Let row vector π of size 2^n denote the corresponding stationary distribution, i.e., for every $1 \leq i \leq 2^n$:

$$\lim_{t\to\infty} \mathbb{P}\big[x(t) = s_i\big] = \pi_i. \tag{21.63}$$

Obviously, the stationary distribution π of $x(t)$ over $\{0,1\}^n$ depends on the choice of ϵ and the matching rule. Let $\pi^u(\epsilon)$ and $\pi^{2k}(\epsilon)$ denote the stationary distributions of $x(t)$ over $\{0,1\}^n$ under uniform and $2k$-neighbor matching rules, respectively. The objective is to characterize $\pi^u(\epsilon)$ and $\pi^{2k}(\epsilon)$, and the rate of convergence to them, when $\epsilon \to 0$. Furthermore, we would like to obtain the probability by which an agreement is achieved asymptotically. Thus, for a stationary distribution π, let $\pi_{\mathbf{0}}$ and $\pi_{\mathbf{1}}$ denote the probability of an asymptotic agreement on 0 and 1, respectively. The following theorem characterizes $\pi_{\mathbf{0}}$ and $\pi_{\mathbf{1}}$ under both uniform and $2k$-matching rules.

Theorem 8. *Let $\pi^u(\epsilon)$ and $\pi^{2k}(\epsilon)$ be the stationary distributions of $x(t)$ over $\{0,1\}^n$ under uniform and $2k$-neighbor matching rules, respectively. Then, for n sufficiently large:*

(i) $\lim_{\epsilon\to 0} \pi_{\mathbf{0}}^u(\epsilon) = 1,$

$$\lim_{\epsilon\to 0} \pi_{\mathbf{0}}^{2k}(\epsilon) = 1;$$

(ii) $\pi_{\mathbf{1}}^u(\epsilon) = O(\epsilon^{n-2\lceil q^*(n-1)\rceil+1}),$

$$\pi_{\mathbf{1}}^{2k}(\epsilon) = \begin{cases} O(\epsilon^{n-2}) & \text{if } n \text{ is even,} \\ O(\epsilon^{n-1}) & \text{if } n \text{ is odd.} \end{cases}$$

We point out that n in Theorem 8 above must be sufficiently large so that $q^* \le \frac{1}{2} - \frac{1}{2(n-1)}$ holds. As we intuitively explain in the following, this assumption is crucial when the matching rule is uniform and n is even: Notice in the settings of the game, action 0 has been favored over action 1 in two ways:

- 0 is played when a player is indifferent between 0 and 1;
- $q^* < \frac{1}{2}$, i.e., equilibrium $(0,0)$ is the risk dominant equilibrium of the underlying 2×2 coordination game in Table 21.1.

The former becomes irrelevant under the uniform matching rule and even n, as no tie situation can possibly exist. Thus, the condition $q^* \le \frac{1}{2} - \frac{1}{2(n-1)}$ in Theorem 8 is there to ensure that the latter is relevant, i.e., the risk dominancy of $(0,0)$ counts. To be more precise, recall that in the game under the uniform matching rule, the best response strategy of player $\mathcal{P}_i$ at time $t+1$ is 0 if condition (21.57) holds. For risk dominancy to count, i.e., for $q^* < \frac{1}{2}$ to have any impact on the dynamics of the game, there must exist situations where condition (21.57) is not equivalent to $q_i(t) > \frac{1}{2}$. There is no such situation unless $q^* \le \frac{1}{2} - \frac{1}{2(n-1)}$.

Hence intuitively, action 0 has been favored over action 1 in the settings of the game. This, according to Theorem 8 (i), turns out to have significant impact on the stationary distribution of $x(t)$ over $\{0,1\}^n$. In fact, Theorem 8 (i) says that for sufficiently small $\epsilon > 0$, every player's strategy is 0 in steady state with an arbitrarily high probability under either uniform or $2k$-neighbor matching rules. Consequently, as $\epsilon \to 0$, the probability of all players' strategies being 1 in steady state vanishes. Theorem 8 (ii) clarifies how fast that vanishing happens.

It is noteworthy that the uniform and $2k$-neighbor matching rules result in identical stationary outcomes. However, the same cannot be said once we move beyond 2×2 games as the underlying coordination game. For instance, consider the

Table 21.2 Examples of the underlying 3×3 coordination games.

	0	1	2
0	6, 6	0, 5	0, 0
1	5, 0	7, 7	5, 5
2	0, 0	5, 5	8, 8

	0	1	2
0	6, 6	0, 5	0, 1
1	5, 0	7, 7	5, 5
2	1, 0	5, 5	8, 8

3×3 games of Table 21.2. If the one on the left is the underlying coordination game in the settings of our game, under the uniform matching rule, for sufficiently small $\epsilon > 0$, every player's strategy is 1 in steady state with an arbitrarily high probability. We also assumed that a player favors 1 over 2 when she is indifferent between them. For the 3×3 game on the right in Table 21.2 as the underlying coordination game, every player's strategy is 2 in steady state with an arbitrarily high probability as $\epsilon \to 0$. Let us now consider the $2k$-neighbor matching rule for $k = 1$, i.e., the case in which every player has two neighbors. Possible actions of the neighbors of a player are $\{0, 0\}$, $\{0, 1\}$, $\{0, 2\}$, $\{1, 1\}$, $\{1, 2\}$, and $\{2, 2\}$. It turns out that for each of these possibilities, the player's best response strategy is the same for both games of Table 21.2. Therefore, unlike the uniform matching rule, under the 2-neighbor matching rule, whichever of the 3×3 games is the underlying coordination game, the stationary distribution of $x(t)$ remains the same. Thus, the stationary outcomes of the two matching rules are not always the same as $\epsilon \to 0$.

In the reminder of this subsection, we focus on the rate of convergence of the distribution of $x(t)$ over $\{0, 1\}^n$ to its stationary distribution π. Let row vector ρ be the initial distribution. Therefore:

$$\lim_{t \to t_0} \rho P(\epsilon)^t = \pi, \tag{21.64}$$

or equivalently:

$$\lim_{t \to t_0} \|\rho P(\epsilon)^t - \pi\| = 0, \tag{21.65}$$

where for convenience $\|.\|$ is assumed to be the *max* norm. From the theory of Markov chains, we know that the rate of convergence in (21.65) is geometric, i.e., there exists some $r < 1$ such that $\|\rho P(\epsilon)^t - \pi\| = O(r^t)$. Therefore, regarding the rate of convergence, we are interested in the value of r. In the following theorem, the parameter r is characterized for small ϵ under the uniform and 2-neighbor matching rules. It is generally expected that $r \to 1$ as $\epsilon \to 0$.

Theorem 9. *Let $P^u(\epsilon)$ and $P^2(\epsilon)$ be transition matrices for the uniform and 2-neighbor matching rules and $\pi^u(\epsilon)$ and $\pi^2(\epsilon)$ be the corresponding associated stationary distributions of $x(t)$ over $\{0, 1\}^n$. Let Δ be the set of all probability distributions over $\{0, 1\}^n$. Assume that $q^* \le \frac{1}{2} - \frac{1}{2(n-1)}$. Defining:*

$$\begin{aligned} r^u(\epsilon) &\triangleq \sup_{\rho\in\Delta} \limsup_{t\to\infty} \|\rho P^u(\epsilon)^t - \pi^u(\epsilon)\|^{1/t}, \\ r^2(\epsilon) &\triangleq \sup_{\rho\in\Delta} \limsup_{t\to\infty} \|\rho P^2(\epsilon)^t - \pi^2(\epsilon)\|^{1/t}, \end{aligned} \tag{21.66}$$

we have

$$\begin{aligned} 1 - r^u(\epsilon) &= O(\epsilon^{\lceil q^*(n-1)\rceil}), \\ 1 - r^2(\epsilon) &= O(\epsilon), \end{aligned} \tag{21.67}$$

as $\epsilon \to 0$.

Theorem 9 implies that, under the uniform matching rule, the convergence to the stationary distribution becomes drastically slow as it is exponential in n, the number of players. The 2-neighbor matching rule, however, shows a faster rate of convergence for small ϵ.

The results obtained above for the convergence rate were also confirmed in Montanari and Saberi (2009). However, the framework considered in Montanari and Saberi (2009) is more general than that of Ellison (1993). Their respective frameworks are both based on the 2×2 coordination game of Table 21.1 and incorporate identical player utilities and hence identical best response strategies. What is different between the two frameworks is in how the noise enters the strategies of the players. In Ellison (1993), as discussed above, for any player, the probability of not playing her best response strategy, namely ϵ, is independent of the potential loss value she would bear. In the more realistic framework of Montanari and Saberi (2009), however, a player is less likely to not play her best response strategy if the potential value of her loss in that is higher. More precisely, let $g_i : \{0, 1\}^{|N_i|} \to \mathbb{R}$ denote the potential loss value, scaled by the constant factor $\frac{2}{(a-d)+(b-c)}$, of player $\mathcal{P}_i$ for not playing her best response strategy given the actions of her neighbors, i.e., those who form the set N_i. Then, she is assumed to play her best response strategy with probability:

$$\frac{e^{\beta g_i}}{e^{\beta g_i} + e^{-\beta g_i}}, \tag{21.68}$$

where $\beta > 0$ is a constant noise parameter common among all players. Consequently, she plays the opposite strategy to her best response strategy with probability:

$$\frac{e^{-\beta g_i}}{e^{\beta g_i} + e^{-\beta g_i}}. \tag{21.69}$$

Notice that $\beta = \infty$ corresponds to noise-free or best response dynamics.

The modified dynamics turns out to be associated with a reversible Markov chain, yielding a stationary distribution. The convergence time to the stationary distribution is studied for the following types of networks:

- *random graphs* with either a fixed degree sequence with minimum degree 3 or in preferential-attachment model with minimum degree 2;
- d*-dimensional networks*, $d \geq 1$, which are those whose nodes can be positioned in $\mathbb{R}^d$ in such a way that (i) for a fixed $K > 0$, any two nodes within a distance K are connected, and (ii) for any $v > 0$, any cube of volume v contains no more than $2v$ nodes;
- *small-world networks*, whose nodes are those of a d-dimensional grid of size $n^{1/d}$ with (i) each node connected to its nearest neighbors and (ii) each node i connected to k other nodes $j(1), \ldots, j(k)$ drawn independently with distribution $C(n)|i - j|^{-r}$.

Given β and a graph $\mathcal{G}$ representing the matching rule, let T_+ denote the *hitting time* to the state where all players play 0 and define *typical hitting time* τ_+ as:

$$\tau_+ = \sup_{X_0 \in \{0,1\}^n} \inf\{t \geq 0 \,|\, \mathbb{P}(T^+ \geq t) \leq e^{-1}\}. \tag{21.70}$$

The following theorem addresses τ_+ as an indicator of the convergence time to the state where all players play 0.

Theorem 10. *Let* $h \triangleq \frac{(a-d)-(b-c)}{(a-d)+(b-c)}$. *As* $\beta \to \infty$, $\tau_+(\mathcal{G}) = \exp\{2\beta\Gamma(\mathcal{G}) + o(\beta)\}$ *where:*

(i) $\Gamma(\mathcal{G}) = \Omega(n)$ *if* h *is sufficiently small and* $\mathcal{G}$ *is a random* k*-regular graph with* $k \geq 3$, *a random graph with a fixed degree sequence with minimum degree 3 or a preferential-attachment graph with minimum degree 2;*
(ii) $\Gamma(\mathcal{G}) = O(1)$ *for all* $h > 0$ *and* d*-dimensional graphs* $\mathcal{G}$ *with bounded range;*
(iii) $\Gamma(\mathcal{G}) = \Omega(\log n/ \log\log n)$ *if* h *is sufficiently small and* $\mathcal{G}$ *is a small-world network with* $r \geq d$;
(iv) $\Gamma(\mathcal{G}) = \Omega(n)$ *if* h *is sufficiently small and* $\mathcal{G}$ *is a small-world network with* $r < d$.

A general conclusion from Theorem 10 is that for those well-connected graphs, i.e., random regular graphs or power-law graphs, the convergence time is high, while for graphs in which a link is drawn between two nodes if they are geographically close to each other, the convergence time is expected to be lower.

To recap, in this subsection we have employed game theory to model a social behavior, namely, conformity, which leads to the spread of a certain innovation throughout the network. Each individual is assumed to adopt the innovation if the ratio of her neighbors doing so reaches a certain threshold exclusive to that individual. If all individuals follow this conformity rule, given that the social graph is connected, their actions converge to the same value in a finite time, unless the social graph is bipartite, which results in a periodic behavior. Similar to the case of Sect. 3.1, however, the introduction of noise, referring to the non-zero probability of best response strategies not being played, helps the convergence and eliminates the possibility of periodicity. Consequently, given that the social graph is connected, the innovation becomes widespread with a high probability when the thresholds are less than $\frac{1}{2}$. A similar conclusion was drawn for a more realistic, modified model in which the noise depends on the current action profile in the sense that an individual is more likely not to adopt her best response strategy if her threshold is closer to the ratio of her neighbors adopting the innovation.

5 Social Choice

Whether it is a jury of twelve deliberating on a verdict or a democratic state of millions choosing its head, it has to convert votes into a decision, a *social choice*. The election process that results in a social choice involves three major elements: *voting scheme*, *candidates*, and *voters*. Let $\mathcal{V} = \{\mathcal{V}_1, \ldots, \mathcal{V}_n\}$ and $\mathcal{C} = \{\mathcal{C}_1, \ldots, \mathcal{C}_m\}$ denote, respectively, the sets of voters and candidates. We shall assume throughout this section that the preference of each voter $\mathcal{V}_k \in \mathcal{V}$ is represented as a binary relation $\mathbf{P}_k$ over the set of candidates, which is both

(i) *total*, meaning that for each pair of candidates $\mathcal{C}_i, \mathcal{C}_j \in \mathcal{C}$, either $\mathcal{C}_i \, \mathbf{P}_k \, \mathcal{C}_j$ or $\mathcal{C}_j \, \mathbf{P}_k \, \mathcal{C}_i$; and
(ii) *transitive*, meaning that for every three candidates $\mathcal{C}_i, \mathcal{C}_j, \mathcal{C}_l \in \mathcal{C}$,

$$\mathcal{C}_i \, \mathbf{P}_k \, \mathcal{C}_j \text{ and } \mathcal{C}_j \, \mathbf{P}_k \, \mathcal{C}_l \;\Rightarrow\; \mathcal{C}_i \, \mathbf{P}_k \, \mathcal{C}_l. \tag{21.71}$$

Let $\mathcal{P}$ denote the set of all profiles of such preferences and $\mathbf{P}$ denote a generic preference profile in $\mathcal{P}$. Note that $\mathbf{P} = \mathcal{R}^n$, where $\mathcal{R}$ denotes the set of all possible rankings of the candidates. A voting scheme is then characterized by a *social choice function* $f : \mathcal{P} \to \mathcal{C}$. Thus, a voting scheme in essence returns a winner candidate taking into account the voter preferences. Some of the well-known voting schemes are described below where their outcomes, when the voter preferences are according to Table 21.3, are also given for illustration.

1. *Plurality*: Each voter votes for its most preferred candidate. Then, the candidate with the most votes wins. Candidate $\mathcal{C}_3$ is the clear winner of the preference profile of Table 21.3.
2. *Instant runoff*: Each voter states its complete ranking of candidates. Then, the winner candidate is the most preferred one by the majority of the voters, if there exists one. Otherwise, the candidate who is the most preferred one by the least number of voters is eliminated, and the voter rankings are updated accordingly. The process continues until a winner candidate emerges. For the preference profile of Table 21.3, no winner candidate emerges instantly as no candidate is the most preferred one by the majority. Thus, candidate $\mathcal{C}_1$ is eliminated as she is the most preferred one by the least number of voters and the voter rankings are

Table 21.3 An example of a preference profile

24 voters:	$\mathcal{C}_1 > \mathcal{C}_2 > \mathcal{C}_3$
35 voters:	$\mathcal{C}_2 > \mathcal{C}_3 > \mathcal{C}_1$
41 voters:	$\mathcal{C}_3 > \mathcal{C}_1 > \mathcal{C}_2$

updated. Candidate $\mathcal{C}_2$ wins the election as she becomes the most preferred by the majority in the updated rankings.

3. *Borda rule*: Each voter rates its least preferred candidate 0, its second least preferred candidate 1, and so on. Then, for each candidate, the rates are accumulated, and the candidate with the highest total score wins the election. It can easily be verified that under this rule candidate $\mathcal{C}_3$ wins for the preference profile of Table 21.3.
4. *Successive elimination*: Each voter states its ranking of the candidates. The winner candidate is then determined in $m - 1$ stages based on a *predefined* list of candidates. In stage 1, the first two candidates from the list are compared. The one that is preferred, based on the voter rankings, by the majority is then compared with the third candidate from the list and so on. The candidate preferred by the majority in the last stage is the winner candidate of the election. It turns out that the outcome of an election based on the successive elimination rule may depend on the predefined order of the candidates list. For instance, in the example of Table 21.3, the winner candidate is the one who appears last in the predefined list, whether it is candidate $\mathcal{C}_1$, $\mathcal{C}_2$, or $\mathcal{C}_3$.

5.1 Fair Choice

All constitutional arrangements forming a voting scheme are made to ensure a *fair* outcome of the election. Different interpretations of the notion of "fairness" in an election are the sole reason for the existence of various voting schemes or, equivalently, various social choice functions. We exploit different fairness criteria in this subsection.

Condorcet Criterion. Given the voter rankings of candidates, it is possible that a candidate, referred to as the *Condorcet winner*, exists, who would win a two-candidate election against each of the other candidates. It is clear that a Condorcet winner is unique if it exists. However, not every preference profile admits a Condorcet winner, a phenomenon known as the *Condorcet paradox*. In other words, collective preferences may be non-transitive even though each voter preference is indeed transitive. For instance, in the example of Table 21.3, in two-candidate elections, $\mathcal{C}_1$ would win against $\mathcal{C}_2$, $\mathcal{C}_2$ would win against $\mathcal{C}_3$, and $\mathcal{C}_3$ would win against $\mathcal{C}_1$.

A social choice function is said to meet the Condorcet criterion if it always returns the Condorcet winner in case it exists. Counterintuitively, most of the social choice functions that are used in real-world elections do not meet such a criterion.

Table 21.4 An example of a preference profile admitting a Condorcet winner ($\mathcal{C}_1$)

3 voters:	$\mathcal{C}_1 > \mathcal{C}_2 > \mathcal{C}_3 > \mathcal{C}_4$
48 voters:	$\mathcal{C}_2 > \mathcal{C}_1 > \mathcal{C}_3 > \mathcal{C}_4$
49 voters:	$\mathcal{C}_3 > \mathcal{C}_4 > \mathcal{C}_1 > \mathcal{C}_2$

Among the voting schemes described in this section, only successive elimination is associated with a social choice function meeting the Condorcet criteria. As an illustrative example, for the preference profile of Table 21.4, a Condorcet winner, namely, $\mathcal{C}_1$, exists, but she is not elected under either plurality ($\mathcal{C}_3$), or instant run-off ($\mathcal{C}_2$), or Borda ($\mathcal{C}_3$) rules.

Contradictory Criteria (Muller and Satterthwaite 1977). Muller and Satterthwaite argue that an ideally fair social choice function $f : \mathcal{P} \to \mathcal{C}$ should satisfy the following three criteria:

(i) *Weak Pareto Efficiency (WPE)*: For every preference profile $\mathbf{P} \in \mathcal{P}$ and two candidates $\mathcal{C}_i, \mathcal{C}_j \in \mathcal{C}$:

$$\mathcal{C}_i\, \mathbf{P}_k\, \mathcal{C}_j,\ \forall \mathcal{V}_k \in \mathcal{V} \ \Rightarrow\ f(\mathbf{P}) \neq \mathcal{C}_j. \tag{21.72}$$

In other words, if all the voters prefer a candidate $\mathcal{C}_i$ over another candidate $\mathcal{C}_j$, then $\mathcal{C}_j$ must not be elected.

(ii) *Monotonicity*: Let $\mathbf{P} \in \mathcal{P}$ be an arbitrary preference profile and $\mathcal{C}_i = f(\mathbf{P})$. If a preference profile $\mathbf{P}' \in \mathcal{P}$ is such that for every $\mathcal{C}_j \in \mathcal{C}\backslash\{\mathcal{C}_i\}$ and $\mathcal{V}_k \in \mathcal{V}$:

$$\mathcal{C}_i\, \mathbf{P}_k\, \mathcal{C}_j \ \Rightarrow\ \mathcal{C}_i \mathbf{P}'_k \mathcal{C}_j, \tag{21.73}$$

then $\mathcal{C}_i = f(\mathbf{P}')$. In other words, the winner candidate for some preference profile would still win if she enjoyed no less support from any voter.

(iii) *Non-dictatorship*: There does not exist a voter $\mathcal{V}_i \in \mathcal{V}$ such that, for every $\mathbf{P} \in \mathcal{P}$, $f(\mathbf{P})$ is the most preferred candidate by $\mathcal{V}_i$. In other words, the winner candidate must not coincide with the most preferred candidate by a single voter for all preference profiles.

Although all the three criteria above seem reasonable, they are not all met by the social choice function associated with any of the well-known voting schemes discussed in this section. In particular, none of them meets monotonicity. For instance, if the preference profile of Table 21.3 is modified to that of Table 21.5, candidate $\mathcal{C}_3$ enjoys no less support from any voter, but she no longer wins the election under plurality or Borda rules. In fact, both plurality and Borda rules now point to candidate $\mathcal{C}_2$ as the winner. One can easily find scenarios in which the social choice functions associated with instant runoff and successive elimination rules also violate the monotonicity condition.

Table 21.5 Modified preference profile of Table 21.3

24 voters:	$\mathcal{C}_2 > \mathcal{C}_1 > \mathcal{C}_3$
35 voters:	$\mathcal{C}_2 > \mathcal{C}_3 > \mathcal{C}_1$
41 voters:	$\mathcal{C}_3 > \mathcal{C}_1 > \mathcal{C}_2$

Table 21.6 Successive elimination by the list $(\mathcal{C}_2, \mathcal{C}_1, \mathcal{C}_4, \mathcal{C}_3)$ returns $\mathcal{C}_3$ as the winner although every voter prefers $\mathcal{C}_2$ over her

3 voters:	$\mathcal{C}_1 > \mathcal{C}_2 > \mathcal{C}_3 > \mathcal{C}_4$
48 voters:	$\mathcal{C}_4 > \mathcal{C}_1 > \mathcal{C}_2 > \mathcal{C}_3$
49 voters:	$\mathcal{C}_2 > \mathcal{C}_3 > \mathcal{C}_4 > \mathcal{C}_1$

Table 21.7 Various voting schemes and fairness criteria

Voting schemes/criteria	WPE	Monotonicity	Non-dictatorship	Condorcet criterion
Plurality	✓		✓	
Instant runoff	✓		✓	
Borda rule	✓		✓	
Successive elimination			✓	✓

Besides monotonicity, WPE criterion is not met by the successive elimination rule either. For instance, for the preference profile of Table 21.6, if successive elimination is done according to the predefined list $(\mathcal{C}_2, \mathcal{C}_1, \mathcal{C}_4, \mathcal{C}_3)$, candidate $\mathcal{C}_3$ emerges as the winner although every voter prefers $\mathcal{C}_2$ over her. Remember, however, that the successive elimination rule, unlike the other three rules of interest, does meet the Condorcet criterion.

It is demonstrated in Table 21.7 as to what voting schemes meet what criteria including the Condorcet criterion.

It now may not come as a surprise that all the three fairness criteria, i.e., WPE, monotonicity, and non-dictatorship, cannot be met by any social choice function.

Theorem 11 (Muller and Satterthwaite 1977). *In an election with more than two candidates and unrestricted preference profiles, any social choice function which satisfies WPE and monotonicity is dictatorial.*

A crucial assumption in Theorem 11 is that the domain of voter preferences must be unrestricted. Indeed, by imposing restrictions on the voter preferences, it is possible to satisfy all the three seemingly contradicting fairness criteria. A well-known type of restricted preference domain is *single-peaked preferences*, discussed below:

Assume that each candidate is associated with a real number in the interval [0, 1]. To create a set of single-peaked preferences, assume also that the *ideal choice* of each voter is a real number in [0, 1]. Then, each voter ranks the candidates according to the distances between their numbers and its ideal choice. More precisely, for every voter $\mathcal{V}_k$ and two candidates $\mathcal{C}_i$ and $\mathcal{C}_j$, we have $\mathcal{C}_i \, \mathbf{P}_k \, \mathcal{C}_j$ if the ideal choice of $\mathcal{V}_k$ is closer to the number associated with $\mathcal{C}_i$ than $\mathcal{C}_j$. It has been proved that for any profile of single-peaked voter preferences, a Condorcet winner exists and is elected under the plurality rule. Furthermore, such a voting scheme meets all the three contradicting fairness criteria. It is also worth mentioning that the winner candidate turns out to be the most preferred candidate by the voter, whose ideal choice is the median of the set of voter ideal choices.

In real-world elections, restrictions on the preference profiles cannot be imposed nor should it be assumed. Therefore, according to Theorem 11, at least one of the three contradicting fairness criteria must be compromised. A glance at Table 21.7 makes it clear that the favorite option to compromise is the monotonicity criterion. We shall see that such a compromise is costly as it makes the voting schemes susceptible to strategic manipulations by both voters and candidates.

Social Welfare. In some elections, not a winner candidate alone but a complete ranking of the candidates is desired. Recall that $\mathcal{R}$ denotes the set of all possible rankings of the candidates. A *social welfare function* is a function $g : \mathcal{P} \to \mathcal{R}$, which captures a complete ranking of the candidates taking the voter preferences into account. Note that most of the voting schemes characterized as social choice functions can also be viewed as social welfare functions. Some examples include plurality, instant runoff, and Borda rules.

The crucial concept of fairness appears to be just as ill-defined for social welfare functions as it is for social choice functions. Arrow (1951, 2nd ed., 1963) argues that for a social welfare function $g : \mathcal{P} \to \mathcal{R}$ to ideally, fairly determine the outcome of an election, it must satisfy the following three conditions:

(i) *Pareto Efficiency (PE)*: For every preference profile $\mathbf{P} \in \mathcal{P}$ and two candidates $\mathcal{C}_i, \mathcal{C}_j \in \mathcal{C}$:

$$\mathcal{C}_i \, \mathbf{P}_k \, \mathcal{C}_j, \ \forall \mathcal{V}_k \in \mathcal{V} \ \Rightarrow \ \mathcal{C}_i \, g(\mathbf{P}) \, \mathcal{C}_j. \tag{21.74}$$

In other words, if every voter ranks a candidate $\mathcal{C}_i$ higher than another candidate $\mathcal{C}_j$, so does the social welfare function.

(ii) *Independence of Irrelevant Alternatives (IIA)*: Let two preference profiles $\mathbf{P}, \mathbf{P}' \in \mathcal{P}$ and two candidates $\mathcal{C}_i, \mathcal{C}_j \in \mathcal{C}$ be arbitrary. If for every $\mathcal{V}_k \in \mathcal{V}$,

$$\mathcal{C}_i \, \mathbf{P}_k \, \mathcal{C}_j \ \Leftrightarrow \ \mathcal{C}_i \, \mathbf{P}'_k \, \mathcal{C}_j, \tag{21.75}$$

then,

$$\mathcal{C}_i \, g(\mathbf{P}) \, \mathcal{C}_j \ \Leftrightarrow \ \mathcal{C}_i \, g(\mathbf{P}') \, \mathcal{C}_j. \tag{21.76}$$

In other words, the ranking of every pair of candidates $\mathcal{C}_i$ and $\mathcal{C}_j$ by the social welfare function only depends on the voters' rankings of those two candidates with respect to each other.

(iii) *Non-dictatorship*: There does not exist a voter $\mathcal{V}_i \in \mathcal{V}$ such that, for every $\mathbf{P} \in \mathcal{P}$, $g(\mathbf{P}) = \mathcal{P}_i$. In other words, the social welfare function does not mirror any single voter's preference.

One can verify that each of plurality, instant runoff, and Borda rules are non-dictatorial and meet the PE criterion. However, neither meets the IIA criterion. In fact, Arrow (1951, 2nd ed., 1963) has proved that no fair social welfare function exists, in the sense of meeting the three criteria above, unless the voter preference domain is restricted or there are at most two candidates:

Theorem 12 (Arrow 1951, 2nd ed., 1963). *In an election with more than two candidates, if the preference domain of the voters is unrestricted, any social welfare function which satisfies the PE and IIA criteria is dictatorial.*

5.2 Strategic Voting

Recall the voting preferences of Table 21.3. Assume that the winner candidate is to be elected under the plurality rule. If all voters vote for their most preferred candidate, then candidate $\mathcal{C}_3$ wins the election. Consider now the following two statements:

(i) If the 24 voters of the first row of Table 21.3 (or at least 7 among them) voted for their second most preferred candidate, that is, $\mathcal{C}_2$, then $\mathcal{C}_2$ would be elected;
(ii) All those 24 voters prefer $\mathcal{C}_2$ over $\mathcal{C}_3$.

Thus, if those 24 voters perceived that their most preferred candidates, that is, $\mathcal{C}_1$, would have little or no chance to be elected, they might vote for their second most preferred candidates, $\mathcal{C}_2$, hoping to get her elected instead of $\mathcal{C}_3$, who is their least preferred candidate. This is an example of what is known as *strategic voting* or *sophisticated voting*.

As it is evident from the example above, strategic voting occurs in elections with more than two candidates when some voters support a candidate more strongly than they sincerely should, to prevent a highly undesirable outcome. In numerous real elections of the past, the outcomes are widely believed to have been influenced by the voters' strategic voting. High-profile examples with successful strategic voting campaigns include the 1997 UK general election and the 2004 Canadian federal election.

In a groundbreaking work (Myerson and Weber 1993), the authors proposed a model within a game theoretic framework for strategic voting behavior. The proposed game model is applicable to various voting schemes such as plurality and Borda rules. In this game, the set of players is comprised of n voters, $\{\mathcal{V}_1, \ldots, \mathcal{V}_n\}$. Each voter $\mathcal{V}_k$ has a payoff vector u_k of size m whose ith element represents the payoff which $\mathcal{V}_k$ would get if candidate $\mathcal{C}_i$ is elected. We assume that the set T of all permissible payoff vectors is finite. This payoff vector also indicates the preference of $\mathcal{V}_k$ over the set of all candidates.

Furthermore, $\mathcal{V}_k$ casts a vector v_k of size m as its vote, whose ith element is its vote for candidate $\mathcal{C}_i \in \mathcal{C} = \{\mathcal{C}_1, \ldots \mathcal{C}_m\}$. This general vector form for the votes has been chosen to cover different voting schemes. For instance, in the plurality voting scheme, v_k is either a vector of all zeros, when the voter does not cast a vote, or a vector with all but one element equal to 0 and one element (corresponding to the voter's most preferred candidate) equal to 1. Alternatively, in the Borda rule, v_k is either a vector of all zeros, if the voter does not cast a vote, or a vector whose elements are a permutation of $0, 1, \ldots, m-1$, where 0 is assigned to the element corresponding to the least preferable candidate by $\mathcal{V}_k$ and $m-1$ is assigned to the element corresponding to the most preferable candidate by $\mathcal{V}_k$. We further assume that set V of all permissible vote vectors is finite.

Each voter $\mathcal{V}_k$ also has a pivot probability $(p_k)_{ij}$ for every pair $\mathcal{C}_i, \mathcal{C}_j \in \mathcal{C}$. More specifically, if $\mathcal{V}_k$ casts a vote v_k, it perceives that it might change the outcome of the election from $\mathcal{C}_j$ to $\mathcal{C}_i$ with probability:

$$(p_k)_{ij} \,.\, \max\{(v_k)_i - (v_k)_j, 0\}.$$

In the following, we shall drop the indices k keeping in mind that all parameters are specific to the voter $\mathcal{V}_k$. For instance, we write v_i, u_i, and p_{ij} instead of $(v_k)_i$, $(u_k)_i$, and $(p_k)_{ij}$. The expected utility of $\mathcal{V}_k$, to be maximized, is defined as:

$$U(p, v, u) \triangleq \sum_{\{\mathcal{C}_i, \mathcal{C}_j\} \subset \mathcal{C}} p_{ij}(v_i - v_j)(u_i - u_j).$$

Defining

$$R_i \triangleq \sum_{\mathcal{C}_j \neq \mathcal{C}_i} p_{ij}(u_i - u_j),$$

as the *prospective rating* of $\mathcal{C}_i$ by $\mathcal{V}_k$, one can write:

$$U(p, v, u) = \sum_{\mathcal{C}_i \in \mathcal{C}} v_i R_i. \tag{21.77}$$

The expected utility (21.77) is to be maximized by $\mathcal{V}_k$ over V, i.e., the set of all permissible vote vectors. Then, according to Myerson and Weber (1993), to maximize its utility,

– under the plurality voting scheme, each voter casts a vote for the candidate with the highest prospective rating; and
– under the Borda rule, each voter ranks the candidates in decreasing order of their prospective ratings.

Optimal votes under other voting schemes, such as approval and range, were also derived in Myerson and Weber (1993). Moreover, the authors define a certain type of voting equilibrium given all the voters cast their optimal vote vectors and share common pivotal probabilities. Such equilibria may not necessarily apply to practical cases and were later refined in De Sinopoli (2000).

Another interesting topic, which has been investigated from a game theoretical viewpoint, but we do not further discuss herein, is *rational voting*. This topic mainly deals with the voter turnout and decisions of the voters to vote or abstain from voting. A few notable game models and their analyses can be found in Ledyard et al. (1981), Palfrey and Rosenthal (1985), Feddersen and Pesendorfer (1996), Dhillon and Peralta (2002), and James and Lahti (2006).

5.3 Strategic Candidacy

Strategic candidacy has long been present in political elections when the voter preferences can be estimated via opinion polls or the outcome of the first round in two-round elections. Recall once again the voter preferences of Table 21.3 and assume the plurality rule as the voting scheme, which means that candidate $\mathcal{C}_3$ is elected. It should be clear that if candidate $\mathcal{C}_1$ is dropped out of the election, candidate $\mathcal{C}_2$ would win. This arguably undesirable phenomenon, in which a *losing* candidate may have the opportunity to manipulate the outcome of the election by dropping out, is referred to as susceptibility to *strategic candidacy*.

If the preference profile of the voters is known, strategic candidacy is a single-stage game in which:

(i) The players are represented by the candidates;
(ii) Each candidate's action is either to enter or exit the election;
(iii) Given an action profile of the candidates and hence given a winner candidate, each candidate's utility indicates how she prefers the winner candidate. More precisely, the candidates, like the voters, are assumed to have preferences over themselves, and they benefit more if a candidate more preferred by them is elected.

In this subsection, reporting from Dutta et al. (2001), we argue that, under the weakest of fairness conditions, namely, non-dictatorship and *unanimity*, for all candidates to enter the election, cannot be a Nash equilibrium for all voter preference profiles, which highlights the importance of strategic candidacy in real-world elections. In connection with the previous subsection, i.e., strategic voting, the interested reader is referred to Brill and Conitzer (2015), where both sets of voters and candidates are assumed to act strategically.

According to Dutta et al. (2001), unanimity requires that if all the voters have the same most preferred candidate, that candidate must be elected. View set $\mathcal{C} = \{\mathcal{C}_1, \ldots, \mathcal{C}_m\}$ as the set of *potential* candidates and assume that an *unknown*, nonempty subset of $\mathcal{C}$ is going to enter the election. Assume that the sets of voters and candidates do not overlap, i.e., $\mathcal{V} \cap \mathcal{C} = \emptyset$. Let once again the preference of each voter $\mathcal{V}_k \in \mathcal{V}$s be represented as a total, transitive binary relation $\mathbf{P}_k$ over the set of potential candidates $\mathcal{C}$. Moreover, for any $A \subset \mathcal{C}$, let $\mathbf{P}_k\,|_A$ denote the preference, induced by $\mathbf{P}_k$, of $\mathcal{V}_k$ over A. Finally, let $\mathbf{P}|_A$ denote the profile of voter preferences, induced by $\mathbf{P}$, over A.

Recall that definition of a social choice function requires a *fixed* set $\mathcal{C}$ of candidates entering the election. Thus, the notion of social choice function has to be generalized if one is to investigate strategic candidacy for which the set of candidates is an unknown subset of $\mathcal{C}$. The related generalized notion shall be referred to as a *voting procedure*.

Definition 6 (voting procedure). A *voting procedure* is a function $V : 2^{\mathcal{C}}\backslash\{\emptyset\} \times \mathcal{P} \to \mathcal{C}$ such that for every subset of candidates $A \in 2^{\mathcal{C}}\backslash\{\emptyset\}$ and preference profile $\mathbf{P} \in \mathcal{P}$:

(i) $V(A, \mathbf{P}) \in A$; and
(ii) $V(A, \mathbf{P}) = V(A, \mathbf{P}')$ for any preference profile $\mathbf{P}' \in \mathcal{P}$ such that $\mathbf{P}|_A = \mathbf{P}'|_A$.

Item (i) above asserts that the outcome of the election must be determined from the subset of candidates who entered the election. Item (ii) on the other hand, asserts that the outcome of the election is determined only based on the voter preferences over the candidates entering the election. We shall see that a voting procedure would inevitably be susceptible to strategic candidacy if it meets the weakest of fairness criteria, i.e., non-dictatorship and unanimity.

Definition 7 (unanimity). A voting procedure $V : 2^{\mathcal{C}}\backslash\{\emptyset\} \times \mathcal{P} \to \mathcal{C}$ is said to satisfy *unanimity* if $V(A, \mathbf{P}) = a$ for every nonempty subset of candidates $A \subset \mathcal{C}$, preference profile $\mathbf{P} \in \mathcal{P}$, and $a \in A$ such that a is the most preferred candidate by each voter $\mathcal{V}_i \in \mathcal{V}$.

Formal definition of susceptibility to strategic candidacy can be derived as the opposite of *candidate stability* defined below:

Definition 8 (candidate stability). A voting procedure $V : 2^{\mathcal{C}}\backslash\{\emptyset\} \times \mathcal{P} \to \mathcal{C}$ is said to be *candidate stable* if for every candidate $\mathcal{C}_i \in \mathcal{C}$ and preference profile $\mathbf{P} \in \mathcal{P}$such that $\mathcal{C}_i \neq V(\mathcal{C}, \mathbf{P})$, we have $V(\mathcal{C}, \mathbf{P}) = V(\mathcal{C}\backslash\{\mathcal{C}_i\}, \mathbf{P})$.

Candidate stability implies that no losing candidate affects the outcome by exiting the election. In other words, if a voting procedure is candidate stable, all candidates entering the election make up a Nash equilibrium for all voter preference profiles in the candidates game described at the beginning of the subsection. The following theorem says that this situation cannot happen for any non-dictatorial, unanimous voting procedure.

Theorem 13. *If a voting procedure* $V : 2^{\mathcal{C}} \backslash \{\emptyset\} \times \mathcal{P} \rightarrow \mathcal{C}$ *is unanimous and candidate stable, it is dictatorial.*

Assume now that each candidate, like voters, has preferences over the entire set of candidates, with herself being the most preferable. Then, *weak candidate stability* as an alternative to candidate stability may be defined as follows:

Definition 9 (weak candidate stability). A voting procedure $V : 2^{\mathcal{C}} \backslash \{\emptyset\} \times \mathcal{P} \rightarrow \mathcal{C}$ is said to be *weakly candidate stable* if for every preference profile $\mathbf{P} \in \mathcal{P}$, every candidate $\mathcal{C}_i \in \mathcal{C}$ prefers $V(\mathcal{C}, \mathbf{P})$ over $V(\mathcal{C} \backslash \{\mathcal{C}_i\}, \mathbf{P})$.

If a voting procedure is weakly candidate stable, no candidate gains by exiting the election. Thus again, all candidates entering the election form a Nash equilibrium for all voter preference profiles. Similar to Theorem 13, one could show that any voting procedure which is weakly candidate stable and unanimous is dictatorial.

5.4 Theory of Coalitions

In this subsection, we consider elections whose winners are *coalitions* of candidates. Consider an election where each voter votes for its most preferred candidate. Having collected the votes, the candidates are allowed to form nonoverlapping coalitions. The coalition with the majority of the votes then wins the election.

For instance, recall the voting preferences of Table 21.3. The following coalition formations are possible, where sets represent coalitions:

(1) $\{\mathcal{C}_1\}\{\mathcal{C}_2\}\{\mathcal{C}_3\}$;
(2) $\{\mathcal{C}_1, \mathcal{C}_2\}\{\mathcal{C}_3\}$;
(3) $\{\mathcal{C}_1, \mathcal{C}_3\}\{\mathcal{C}_2\}$;
(4) $\{\mathcal{C}_1\}\{\mathcal{C}_2, \mathcal{C}_3\}$;
(5) $\{\mathcal{C}_1, \mathcal{C}_2, \mathcal{C}_3\}$.

In case (1), there is no winning coalition as none has the majority of the votes. In each of the cases (2)–(5), the coalition with more than one member candidate wins the election.

Coalition formation can be viewed as a *cooperative* game in which:

(i) The players are represented by the candidates;
(ii) An action profile is characterized as a partition of the set of candidates into coalitions;
(iii) Given an action profile of the candidates, and therefore given a winning coalition, a candidate's utility is 0, if she is not part of the winning coalition, and non-zero otherwise. The total utility of winning coalition members is fixed.

Note that we did not mention specifically what the utility of a winning coalition member should be, although their total sum must be fixed. Without loss of generality, let us assume that the fixed total utility is 1. How to distribute 1 among the members is determined during the process of forming the coalitions. For instance, if a candidate $\mathcal{C}_i$ has more votes than candidate $\mathcal{C}_j$, she may ask for a greater portion of the fixed total utility to be part of the same coalition with $\mathcal{C}_j$.

Minimal size coalitions (Gamson 1961). Assume that the fixed total utility 1 is distributed among the candidates in the winning coalition in proportion to the number of their votes, while the rest of the candidates receive nothing. Thus, the utility of a candidate in the winning coalition is calculated by her number of votes divided by the total number of votes of candidates in the winning coalition. It can be shown that this coalition game among the candidates has at least one *strong Nash equilibrium*, of which a definition is given below:

Definition 10 (strong Nash equilibrium). Given a game $(\mathcal{V}, x, u)$, an action profile $(x_1^*, \ldots, x_n^*)$ is said to be a *strong Nash equilibrium* if for every $\mathcal{S} \subset \mathcal{V}$ and any action profile $(x_1, \ldots, x_n)$ for which

$$x_j = x_j^*, \ \forall \mathcal{V}_i \in \mathcal{V} \backslash \mathcal{S}, \tag{21.78}$$

there exists $\mathcal{V}_j \in \mathcal{S}$ such that

$$u_j(x) \le u_j(x^*). \tag{21.79}$$

In other words, an action profile is a strong Nash equilibrium if there is not a group of players who can all benefit by cooperatively changing their actions.

To characterize such equilibria, let us define *minimal size coalitions*.

Definition 11 (minimal size coalitions). Given the number of each candidate's votes, a *minimal size coalition* is a winning coalition with a minimal number of total votes.

More specifically, among all coalitions whose members altogether have the majority of the votes, the one with the least number of total votes is the minimal size coalition.

Obviously, given any distribution of the votes, there exists at least one minimal size coalition. We show in the following that any action profile involving a minimal size coalition is a strong Nash equilibrium. Consider such an action profile $\mathbf{A}$, and let $\mathbf{A}'$ be an arbitrary action profile. Let $\mathcal{C}_i$ be an arbitrary candidate in the winning (minimal) coalition associated with action profile $\mathbf{A}$. Since the size of the winning coalition for $\mathbf{A}$ is minimal, it is not greater than the size of the winning coalition for $\mathbf{A}'$. Thus, candidate $\mathcal{C}_i$'s utility for $\mathbf{A}$ is not less than her utility for $\mathbf{A}'$, whether

Table 21.8 Strong Nash equilibria of a coalition game associated with Caplow's Theory

Votes distribution	Strong Nash equilibria
$c_1 = c_2 = c_3$	$\{\mathcal{C}_1, \mathcal{C}_2\}\{\mathcal{C}_3\}, \{\mathcal{C}_1, \mathcal{C}_3\}\{\mathcal{C}_2\}, \{\mathcal{C}_1\}\{\mathcal{C}_2, \mathcal{C}_3\}$
$c_1 < c_2 = c_3$	$\{\mathcal{C}_1, \mathcal{C}_2\}\{\mathcal{C}_3\}, \{\mathcal{C}_1, \mathcal{C}_3\}\{\mathcal{C}_2\}$
$c_1 > c_2 = c_3,\ c_1 < c_2 + c_3$	$\{\mathcal{C}_1\}\{\mathcal{C}_2, \mathcal{C}_3\}$
$c_1 > c_2 = c_3,\ c_1 = c_2 + c_3$	$\{\mathcal{C}_1, \mathcal{C}_2\}\{\mathcal{C}_3\}, \{\mathcal{C}_1, \mathcal{C}_3\}\{\mathcal{C}_2\}, \{\mathcal{C}_1, \mathcal{C}_2, \mathcal{C}_3\}$
$c_1 > c_2 = c_3,\ c_1 > c_2 + c_3$	$\{\mathcal{C}_1\}\{\mathcal{C}_2\}\{\mathcal{C}_3\}, \{\mathcal{C}_1\}\{\mathcal{C}_2, \mathcal{C}_3\}, \{\mathcal{C}_1, \mathcal{C}_2, \mathcal{C}_3\}$
$c_1 > c_2 > c_3,\ c_1 < c_2 + c_3$	$\{\mathcal{C}_1, \mathcal{C}_3\}\{\mathcal{C}_2\}, \{\mathcal{C}_1\}\{\mathcal{C}_2, \mathcal{C}_3\}$
$c_1 > c_2 > c_3,\ c_1 = c_2 + c_3$	$\{\mathcal{C}_1, \mathcal{C}_2\}\{\mathcal{C}_3\}, \{\mathcal{C}_1, \mathcal{C}_3\}\{\mathcal{C}_2\}, \{\mathcal{C}_1, \mathcal{C}_2, \mathcal{C}_3\}$
$c_1 > c_2 > c_3,\ c_1 > c_2 + c_3$	$\{\mathcal{C}_1\}\{\mathcal{C}_2\}\{\mathcal{C}_3\}, \{\mathcal{C}_1\}\{\mathcal{C}_2, \mathcal{C}_3\}$

$\mathcal{C}_i$ is in the winning coalition for $\mathbf{A}'$ or not. Therefore, given action profile $\mathbf{A}$, $\mathcal{C}_i$'s utility has reached its maximum over all the action profiles. Thus, if a group of candidates are to strictly increase their utilities by changing their actions, it cannot involve any of the winning coalition members. Consequently, the winning coalition remains intact and wins again, meaning that those who updated their actions have lost again and still have utility 0.

As an example, for the preference profile of Table 21.3, where candidates $\mathcal{C}_1, \mathcal{C}_2$, and $\mathcal{C}_3$ have 24, 35, and 41 votes, respectively, coalition $\{\mathcal{C}_1, \mathcal{C}_2\}$ with a total of 59 votes is a minimal size coalition. Thus, the action profile $\{\mathcal{C}_1, \mathcal{C}_2\}\{\mathcal{C}_3\}$ is a strong Nash equilibrium. Notice that the utilities of $\mathcal{C}_1$ and $\mathcal{C}_2$ are calculated as $\frac{24}{59}$ and $\frac{35}{59}$, respectively. Neither $\mathcal{C}_1$ nor $\mathcal{C}_2$ would be willing to form a coalition with $\mathcal{C}_3$ since their utilities would then strictly decrease.

Caplow's Theory and its generalization (Caplow 1968; Shenoy 1978). Focusing on the case of three candidates, Caplow does not believe that a minimal size winning coalition should necessarily have advantage over other winning coalitions. In particular, out of the five coalition formations for the example of Table 21.3, Caplow argues that action profiles $\{\mathcal{C}_1, \mathcal{C}_2\}\{\mathcal{C}_3\}$ and $\{\mathcal{C}_1, \mathcal{C}_3\}\{\mathcal{C}_2\}$ are just as likely to occur. A mathematical formulation of Caplow's Theory, reported from Shenoy (1978), is given below.

Let an action profile $\mathbf{A}$ involving a winning coalition be given and assume that $\mathcal{C}_i$ is a candidate in the winning coalition. A candidate $\mathcal{C}_j$ is said to be *controlled* by $\mathcal{C}_i$ if she has fewer votes than $\mathcal{C}_i$ or she is not part of the winning coalition. The power index of $\mathcal{C}_i$ is then defined as the number of candidates controlled by her. Finally, $\mathcal{C}_i$'s utility is calculated as her power index divided by the total cumulative power index of the candidates in the winning coalition. Once again, strong Nash equilibria of the game are of interest. For the case of three candidates $\mathcal{C}_1, \mathcal{C}_2$, and $\mathcal{C}_3$ whose number of votes are c_1, c_2, and c_3, respectively, Table 21.8 demonstrates the game's strong Nash equilibria according to how c_1, c_2, and c_3 are related.

In this subsection, we have dealt with the formation of coalitions in an election as a single-stage game among the candidates. Each candidate was assumed to receive a utility if she is a member of a coalition after the game is played. Her utility is not necessarily uniform over the collection of possible coalitions. The point of interest now becomes to find the strong Nash equilibria of the game given its associated utilities. Two approaches toward defining the utilities were introduced, and the strong Nash equilibria for each case were obtained.

6 Conclusion

In this chapter, we have presented some applications of game theory in social networks. In particular, two fundamental concepts in social networks, namely, opinion dynamics and social choice, were modeled as games among the individuals in a social network. While opinion dynamics is associated with multistage games in general, as it resembles a time-varying behavior of a social network, the social choice is modeled as a single-stage game.

Depending on the topic, the set containing all possible opinions may be of continuous or discrete nature. For instance, political opinions can be viewed as continuous opinions since political views can range from very extremist to moderate. When adopting an innovation, however, one has a finite number of choices from what would be available in the market, which corresponds to a discrete opinion. Hence, in the games modeling of opinion dynamics, the set of actions may be finite or infinite, requiring very different techniques to be used in their analysis.

In modeling opinion dynamics via games, one has to define the action set and utility function of each player in such a way that they reasonably capture the social behavior. Since there is no single best way to do that, there does not exist an inclusive game model to address opinion dynamics, whether continuous or discrete. Therefore, we have discussed a number of very different approaches employing game theory to model opinion dynamics. For each approach, various playing strategies, such as best response strategy or a noisy version of it, were investigated; it was determined whether the game dynamics converges, or equivalently whether an agreement is reached or a clustering occurs, as time goes by; it was made clear whether the game dynamics converges to a Nash equilibrium in case it converges at all; and the convergence rate to the equilibrium was made explicit.

A fundamentally different problem from opinion dynamics is the social choice process, which corresponds to the voting process. Social choice is generally viewed as a single-stage game, which is what we adopted in this chapter, although in some practical cases, two-stage elections are also possible. We argued that any voting scheme is fundamentally flawed in the sense that it cannot meet all of certain fairness criteria. This makes a voting scheme prone to manipulation, leading to the emergence of gaming in elections. Two well-known manipulation games regarding elections are strategic voting and strategic candidacy, with the latter discussed in the previous section.

Another type of game emerging in elections is to address the formation of candidate coalitions. It was assumed herein that the coalitions are to be formed after the votes are counted and the number of votes of each candidate is known. A coalition whose members altogether have the majority of the votes is a winning coalition. Thus, for the formation of coalitions, one investigates a single-stage game among the candidates to determine how a winning coalition emerges. Since the game is a cooperative one, strong Nash equilibria of the game are of great importance. We have introduced two well-known games of coalition formation and obtained their strong Nash equilibria.

References

Acemoğlu D, Como G, Fagnani F, Ozdaglar A (2013) Opinion fluctuations and disagreement in social networks. Math Oper Res 38(1):1–27

Arrow KJ (1951, 2nd ed., 1963) Social choice and individual values. Yale University Press, New Haven/London

Bernard J (1954) The theory of games of strategy as a modern sociology of conflict. Am J Sociol 59(5):411–424

Brill M, Conitzer V (2015) Strategic voting and strategic candidacy. AAAI 15:819–826

Caplow T (1968) Two against one: coalitions in triads. Prentice-Hall, Englewood Cliffs

Chatterjee S, Seneta E (1977) Towards consensus: some convergence theorems on repeated averaging. J Appl Probab 14(1):89–97

Couzin ID, Krause J, Franks NR, Levin SA (2005) Effective leadership and decision-making in animal groups on the move. Nature 433(7025):513–516

Cucker F, Smale S (2007) Emergent behavior in flocks. IEEE Trans Autom Control 52(5):852–862

Cucker F, Smale S, Zhou D-X (2004) Modeling language evolution. Found Comput Math 4(3):315–343

De Sinopoli F (2000) Sophisticated voting and equilibrium refinements under plurality rule. Soc Choice Welf 17(4):655–672

Deffuant G, Neau D, Amblard F, Weisbuch G (2000) Mixing beliefs among interacting agents. Adv Complex Syst 3(01n04):87–98

DeGroot MH (1974) Reaching a consensus. J Am Stat Assoc 69(345):118–121

Dhillon A, Peralta S (2002) Economic theories of voter turnout. Econ J 112(480):332–352

Dutta B, Jackson MO, Le Breton M (2001) Strategic candidacy and voting procedures. Econometrica 69(4):1013–1037

El Chamie M, Liu J, Başar T (2016) Design and analysis of distributed averaging with quantized communication. IEEE Trans Autom Control 61(12):3870–3884

Ellison G (1993) Learning, local interaction, and coordination. Econometrica J Econ Soc 61:1047–1071

Ermentrout GB (1992) Stable periodic solutions to discrete and continuum arrays of weakly coupled nonlinear oscillators. SIAM J Appl Math 52(6):1665–1687

Etesami SR, Başar T (2015) Game-theoretic analysis of the Hegselmann-Krause model for opinion dynamics in finite dimensions. IEEE Trans Autom Control 60(7):1886–1897

Etesami SR, Başar T (2016) Complexity of equilibrium in competitive diffusion games on social networks. Automatica 68:100–110

Feddersen TJ, Pesendorfer W (1996) The swing voter's curse. Am Econ Rev 86:408–424

Flierl G, Grünbaum D, Levins S, Olson D (1999) From individuals to aggregations: the interplay between behavior and physics. J Theor Biol 196(4):397–454

Gamson WA (1961) A theory of coalition formation. Am Sociol Rev 26:373–382

Ghaderi J, Srikant R (2013) Opinion dynamics in social networks: a local interaction game with stubborn agents. In: American control conference (ACC). IEEE, pp 1982–1987

Goyal S, Heidari H, Kearns M (2014) Competitive contagion in networks. Games Econ Behav. Online 5 Oct 2014

Graver J, Servatius H, Servatius B (1984) Chemical oscillators, waves, and turbulance. Springer, Berlin
Harary F (1962) The maximum connectivity of a graph. Proc Natl Acad Sci 48(7):1142–1146
Harsanyi JC, Selten R et al (1988) A general theory of equilibrium selection in games, vol 1. MIT Press, Cambridge
Hegselmann R, Krause U (2002) Opinion dynamics and bounded confidence models, analysis, and simulation. J Artif Soc Soc Simul 5(3):1–33
Huang M, Manton JH (2009) Coordination and consensus of networked agents with noisy measurements: stochastic algorithms and asymptotic behavior. SIAM J Control Optim 48(1):134–161
Jadbabaie A, Lin J, Morse AS (2003) Coordination of groups of mobile autonomous agents using nearest neighbor rules. IEEE Trans Autom Control 48(6):988–1001
James S, Lahti T (2006) A theory of participation in elections. Am Econ Rev 96(4):1271–1282
Kandori M, Mailath GJ, Rob R (1993) Learning, mutation, and long run equilibria in games. Econometrica J Econ Soc 61:29–56
Karp R, Schindelhauer C, Shenker S, Vocking B (2000) Randomized rumor spreading. In: Proceedings of 41st annual symposium on foundations of computer science, pp 565–574
Kashyap A, Başar T, Srikant R (2007) Quantized consensus. Automatica 43(7):1192–1203
Kempe D, Kleinberg J (2002) Protocols and impossibility results for gossip-based communication mechanisms. In: Proceedings of 43rd annual IEEE symposium on foundations of computer science, pp 471–480
Ledyard JO et al (1981) The paradox of voting and candidate competition: a general equilibrium analysis. Technical report
Lin P, Jia Y (2009) Consensus of second-order discrete-time multi-agent systems with nonuniform time-delays and dynamically changing topologies. Automatica 45(9):2154–2158
Lorenz J, Urbig D (2007) About the power to enforce and prevent consensus by manipulating communication rules. Adv Complex Syst 10(02):251–269
Lynch NA (1996) Distributed algorithms. Morgan Kaufmann, San Francisco
Marden JR, Arslan G, Shamma JS (2009) Cooperative control and potential games. IEEE Trans Syst Man Cybern B Cybern 39(6):1393–1407
Montanari A, Saberi A (2009) Convergence to equilibrium in local interaction games. In: 50th annual IEEE symposium on foundations of computer science, FOCS'09. IEEE, pp 303–312
Muller E, Satterthwaite MA (1977) The equivalence of strong positive association and strategy-proofness. J Econ Theory 14(2):412–418
Myerson RB, Weber RJ (1993) A theory of voting equilibria. Am Polit Sci Rev 87(1):102–114
Nedić A, Olshevsky A, Ozdaglar A, Tsitsiklis JN (2009) On distributed averaging algorithms and quantization effects. IEEE Trans Autom Control 54(11):2506–2517
Palfrey TR, Rosenthal H (1985) Voter participation and strategic uncertainty. Am Polit Sci Rev 79(1):62–78
Shenoy PP (1978) On coalition formation in simple games: a mathematical analysis of Caplow's and Gamson's theories. J Math Psychol 18(2):177–194
Strogatz SH (2001) Exploring complex networks. Nature 410(6825):268–276
Sun YG, Wang L (2009) Consensus of multi-agent systems in directed networks with nonuniform time-varying delays. IEEE Trans Autom Control 54(7):1607–1613
Swedberg R (2001) Sociology and game theory: contemporary and historical perspectives. Theory Soc 30(3):301–335
Tsitsiklis JN (1984) Problems in decentralized decision making and computation. Technical report, DTIC Document
Weisbuch G, Deffuant G, Amblard F, Nadal J-P (2002) Meet, discuss, and segregate! Complexity 7(3):55–63
Young HP (2001) Individual strategy and social structure: an evolutionary theory of institutions. Princeton University Press, Princeton

Pursuit-Evasion Games

22

Valerii Patsko, Sergey Kumkov, and Varvara Turova

Contents

V. Patsko (✉) · S. Kumkov
N. N. Krasovskii Institute of Mathematics and Mechanics, Ural Branch of Russian Academy of Sciences, Ekaterinburg, Russia
e-mail: patsko@imm.uran.russkumk@gmail.com

V. Turova
Department of Mathematics, Technical University of Munich, Garching, Germany
e-mail: turova@ma.tum.de

T. Başar, G. Zaccour (eds.), *Handbook of Dynamic Game Theory*,
https://doi.org/10.1007/978-3-319-44374-4_30

Abstract

Applied problems whose investigation involves methods of pursuit-evasion differential games are described. The main focus of this chapter is on time-optimal problems close to R. Isaacs' "homicidal chauffeur" game and to linear differential games of fixed terminal time with J. Shinar's space interception problem as the major example. These problems are taken because after a change of variables they can be reduced to models with two state variables. This allows us to provide adequate graphical representations of the level sets of the value functions being obtained numerically and emphasize important peculiarities of these sets. Also, other conflict control problems and control problems with uncertainties being extensively investigated nowadays are briefly outlined.

Keywords

Differential game · Homicidal chauffeur · Space interception · Semipermeable curves · Barriers · Singular surfaces · Maximal stable bridge

1 Introduction

1. Pioneering works on differential games were accomplished by R. Isaacs and published, starting from 1951, in his reports for the RAND Corporation. Already in the first of them (1951), R. Isaacs used the term "differential game" and formulated the "homicidal chauffeur" game, which later became one of the most famous problems. Note that in the beginning of the 1950s, modern mathematical optimal control theory relevant to accounting for "geometric constraints" on control had only just started to develop. The theorem on necessary conditions for optimal open-loop control, which received the name "Pontryagin maximum principle" was published in 1956–1957. A little earlier, the works of D. W. Bushaw and A. A. Feldbaum on construction of optimal feedback control in linear control problems in the plane had appeared.

In the homicidal chauffeur game a "car" pursues a "pedestrian." The car (a point in the plane) has constant velocity magnitude, but the direction of the velocity vector cannot change instantaneously because the angle velocity is bounded. In other words, there is a restriction on the turn radius of the car. The pedestrian (another point in the plane) is a non-inertia object whose velocity magnitude is bounded, but the direction can change instantaneously. Such non-inertia objects were called by R. Isaacs as objects with "simple motion." The pursuer minimizes the time of capture in a given neighborhood of the evader, while the pedestrian hinders this.

Of course, when considering this game, R. Isaacs kept in mind an applied problem, in which a torpedo pursues an evading small ship (Breitner 2005). It was the genius of R. Isaacs to give "catchy names" to applied problems in his reports and his book "Differential Games" (Isaacs 1965) and to leave only principal features in mathematical description, putting away a plenty of details that always exist when investigating applied problems. Studying differential games (in R. Isaacs' understanding) is first to model the main important features of complex antagonist problems and then to solve them using analytical or numerical methods.

2. Isaacs' method is based on the consideration of a first-order partial differential equation for the value function, which is analogous to the well-known in mechanics Hamilton-Jacobi equation. R. Isaacs derived the corresponding equation by letting feedback controls as admissible classes of controls in zero-sum games and formulated the "transition principle," or the "principle of guaranteed nondeterioration of result in the process of motion." Nowadays, this is a well-known principle of backward constructions (dynamic programming), when the value function is recomputed by going back from the terminal conditions. The operator of such recalculation implements the operations of minimum over controls of one player and maximum over controls of another one. Thereby, R. Isaacs clearly realized that the value function is, as a rule, non-smooth or even discontinuous. The latter is typical for time-optimal games, in which the payoff is the capture time.

R. Isaacs introduced the notion of "singular" surfaces in the game space and performed a classification of possible types of surfaces. He considered the construction of singular surfaces as a basis for solving differential games. Correctly constructed singular surfaces form a skeleton of solution by generating a peculiar separation of solution into cells. In the interior of each cell a single smooth optimal trajectory goes through every point. On singular surfaces, kinks of optimal trajectories, violation of uniqueness, etc. occur.

After the publication of R. Isaacs' book, theoretical investigation of singular surfaces, their analysis for particular applied problems were performed by J. Breakwell and his postgraduate students A. Merz, P. Bernhard, J. Lewin, and G.-J. Olsder. P. Bernhard in the paper (1977) and A. A. Melikyan in the book (1998) obtained differential equations for typical singular surfaces. Consideration of differential games with the use of singular surfaces was done in the book by J. Lewin (1994). However, one should be clearly aware of the fact that solving differential games by construction of singular surfaces requires enormous effort even for problems in the plane. In the latter case, we construct not singular surfaces but singular lines. Most likely, in high-dimensional differential games, detecting and classification of singular surfaces are to be appreciated as very useful research that should, however, be performed after constructing level sets of the value function.

3. The theory of differential games has been extensively developed in the Soviet Union in the 1960s–1980s. There existed four centers where differential games were intensively investigated: The mathematical school of L. S. Pontryagin in Moscow, the school of N. N. Krasovskii in Sverdlovsk (now Ekaterinburg), the school of B. N. Pschenichnyi in Kiev, and the school headed by L. A. Petrosyan in Leningrad (now St.-Petersburg).

In his works on the pursuit problem, L. S. Pontryagin assumed that the first player (the pursuer) discriminates the second player (the evader). The discrimination is reduced to the requirement of informing on a small current time interval about the second player's control. In the problem from the side of the second player, on the contrary, the first player is discriminated.

A similar concept was followed by B. N. Pschenichnyi and partly by L. A. Petrosyan.

From the very beginning of his investigations, N. N. Krasovskii followed positional formalization, in which the control is constructed using only current position of the game but is being applied using a discrete control scheme. The latter means that the control chosen at some time instant of a given time grid remains constant until the next time instant of the grid. When solving differential games in positional formalization, the final result is the generation of (optimal) feedback strategy that guarantees the best outcome to the respective player, provided the step width of the discrete control scheme goes to zero.

4. In two papers Pontryagin (1967a,b) devoted to differential games with linear dynamics, L. S. Pontryagin showed how one can account for the advantage of the pursuer over the evader using the notion of geometric difference (Minkowski difference) and how, based on the feedback procedure, the solvability set in the problem of approaching a given target set by a conflict-controlled system can be constructed. Among other papers by L. S. Pontryagin, let us note the works (Pontryagin and Mischenko 1971; Pontryagin 1971) devoted to evasion problem on an infinite time interval. For objects with linear dynamics, very "fine" condition of dynamic advantage of the second player over the first player has been formulated. Once this condition is fulfilled, the evader performs an evading maneuver in dangerous situations of approach. Then he is waiting for the next dangerous situation and so on.

In the book (1970), N. N. Krasovskii proposed effective methods for solving linear differential games based on the notion of reachable sets, i.e., on solving the problem in the class of open-loop controls. Though these methods give an optimal result only in the case where some "regularity conditions" are fulfilled, from the practical point of view, they can be also applied in cases where regularity conditions are not satisfied, because very often the difference between the optimal result and the result obtained is unessential. Moreover, these methods are very clear and can be easily understood by engineers.

Somewhat later, N. N. Krasovskii and A. I. Subbotin introduced (1974, 1988) for a wide class of differential games with nonlinear dynamics the notions of stable bridge and maximal stable bridge. The latter is the maximal set in the space (time $\times$ state vector), from which the first player can solve the problem of approaching a given target set under the assumption of discrimination of his opponent (the second player). Thus, being absolutely unacceptable from the point of view of the engineering practice, the idealized assumption on the discrimination was included into the theoretical construction. It was shown that if the stable bridge (or the maximal stable bridge) is somehow constructed, then an extremal to the stable bridge positional strategy of the first player holds trajectories of the control system in a sufficiently small neighborhood of the stable bridge, provided that the discrete control scheme with sufficiently small time step is used. The concept of stable bridges allowed N. N. Krasovskii and A. I. Subbotin to prove the existence of the value function for different classes of differential games. Efficient numerical methods for the construction of maximal stable bridges were developed in Ekaterinburg (Grigor'eva et al. 2005; Subbotin and Patsko 1984; Taras'yev et al. 1988; Ushakov 1998).

The theory of positional control is developed for differential games with nonlinear dynamics and separable controls of the players. Thereby, usually local Lipschitz condition and sublinear growth in state variable, measurability in time, and continuity in controls are required for the function in the right-hand side. It is extremely important that the results achieved are generalized (Krasovskii and Subbotin 1974, 1988; Subbotin and Chentsov 1981) to the case of systems with inseparable controls, including the case where Isaacs' condition (equality of minmax and maxmin-Hamiltonians) is not fulfilled. The above conditions are assumed to be satisfied also in numerical constructions. In numerical procedures, matching of time and spatial discretization step widths is required additionally.

First works accomplished by L. A. Petrosyan are related to the "lifeline" game that was introduced in the book by R. Isaacs. In this game, the evader strives to reach a given terminal set, whereas the pursuer tries to catch the evader as soon as possible. In the papers Petrosjan (1965) and Petrosyan and Dutkevich (1972), this game is completely solved in the case of simple pursuit (i.e., when the objects have dynamics with simple motion). In addition, it is revealed that for the considered class of dynamics of the pursuer and evader, in the case of point capture, the optimal strategy of the pursuer is the well-known in the engineering practice parallel approach strategy. Among the works by L. A. Petrosyan from the late 1960s, let us mention the paper Petrosyan (1970) where the problem with the evader information time lag is considered. Here, it is proved that the optimal strategy of the evader is mixed. On this topic, see also Petrosyan (1977, 1993). From the middle of 1970s, the school of L. A. Petrosyan started to pay more attention, along with zero-sum differential games, to noncooperative and cooperative dynamic games with many players, which find use in applied economic theory.

5. In the theory of differential games, problems with complete information are distinguished from those ones with incomplete information. The problems with complete information assume precise knowledge of the current position of the game by all participants. This is not the case in the problems with incomplete information. For example, the pursuer forms his control based not on precise information on the state of the evader but on information obtained from inexact measurements only. Moreover, in practice, one should account for information and processing delays. This creates difficulties even on the problem statement stage. Problems with incomplete information accomplished at the beginning of 1970s are presented in the books Chernous'ko and Melikyan (1978), Krasovskii (1970), Krasovskii and Subbotin (1974), Petrosyan (1977), and Kurzhanski (1977). In the years ahead, numerous attempts were made by the scientific school of

N. N. Krasovskii to develop a theory of problems with incomplete information in such a way that it would contain similar elements (maximal stable bridges, extremal positional strategies) as in the theory of problems with complete information. These approaches are reflected in the works Kurzhanski (2004), Osipov (2006), and Kryazhimskiy and Osipov (2010).

As problems with incomplete information, also settings can be considered, in which one of the conflicting players knows the precise phase state of the game at some instants only. There exist statements that assume a bounded number of observation times and that the respective player chooses a future observation instant based on the information about the phase state being available at the previous observation. Using the description of the informal problem from Neveu et al. (1995), let us imagine a helicopter that detected a submarine and tries to approach it (as projected to the horizontal plane) close enough to deliver a weapon. The submarine maneuvers in order to escape to a secure zone. The helicopter is only equipped with a dipping active sonar. Thus, to get information concerning the submarine position, the helicopter has to choose instant for dip position to detect and localize the submarine. After this instant, it chooses the next dip instant and so on. No information is available for the helicopter between the two dips. Therefore, from the point of view of the first player (helicopter), a problem of the choice of observation instants and of the construction of an open-loop control between two subsequent observation points that ensures appropriate result at the end of pursuit occurs. Considering the problem from the point of view of the second player (submarine) also yields in a statement with bounded number of observations of the game phase states but for the submarine.

Similar problems were considered by A. A. Melikyan (1973, 1975) in frames of differential game theory in the early 1970s. He found model examples, in which the correct (optimal) choice of the observation instants provides the equality of the best guaranteed result of the observing player and the optimal result that this player could guarantee under continuous observation of the current position of the game. The results obtained are included in the book Chernous'ko and Melikyan (1978). It occurred that questions on optimal choice of observation instants are closely connected to important theoretical questions related to the coincidence of the value of differential game (under condition of continuous observation of the game state) with the iterations of programming max-min function. Such programmed iterations were proposed by A. G. Chentsov (1976, 1978a). Corresponding results are included in the book Subbotin and Chentsov (1981).

At the beginning of the 1990s, topics related to bounded number of observation instants were further developed in the works by P. Bernhard, O. Pourtallier and their colleagues (Bernhard and Pourtallier 1994; Neveu et al. 1995; Olsder and Pourtallier 1995). Having in mind some applied problems, they investigated statements, in which the observing player needs some specified time to determine the phase state, and this player is immovable during the measurement (helicopter during acting a dipping active sonar). For games with simple motion and games with linear dynamics, construction of sets of initial states from which the reach of a given target set can be guaranteed under a specified number of observation intervals, as well as problems of minimization of the whole observation time, and some others, were considered.

6. In the 1970s–1990s, in different countries, investigations of pursuit-evasion games with objects separated into two groups were performed. For example, several objects collected in one group should capture all evaders combined in another group in finite time. Of course, from the point of view of existence of the value function and optimal feedback strategies, problems with many objects, as a rule, are included into the general theory of differential games. However, what would be effectively verifiable conditions of successful capture? How can one construct optimal strategies of the players? The papers Pschenichnyi et al. (1981), Petrov (1988), and Grigorenko (1989) and the books Chikrii (1997), Grigorenko (1990), and Blagodatskih and Petrov (2009) are devoted to the investigation of such questions for different classes of linear differential games. A great stimulating role in the creation of methods of group pursuit game theory had the work (1976) by B. N. Pschenichnyi, which considered the problem of the successful capture of one evader by a group of several pursuers in the case where all objects are identical and their dynamics are that of simple motion. In the paper Mishchenko et al. (1977), the local evasion maneuver from the work (1971) by L. S. Pontryagin and E. F. Mishchenko is extended to the situation of many pursuers. In the work (1976), for problems with simple motion, F. L. Chernous'ko suggested his method of preventing the capture of the evader by a group of pursuers. Under assumption of advantage in velocity of the evader over each pursuer, the method provides a certain distance evasion from all pursuers with keeping the motion inside a prescribed neighborhood of a given basic trajectory. Particular problems with evident applied character were considered in Hagedorn and Breakwell (1976) and Levchenkov and Pashkov (1990). In Petrosyan (1966) and Petrosyan and Shiryaev (1980), for differential games with several pursuers and several evaders, the notion of Nash equilibrium is used. Close results are presented in the book Petrosyan (1993). A survey of publications on pursuit-evasion games with many players is given in the paper Kumkov et al. (2017).

7. In the 1980s and early 1990s, the attention of many researchers was drawn to problems of aircraft control in the presence of wind disturbances. A tremendous role in the development of this topic played the publications by A. Miele and his collaborators (Miele et al. 1986, 1987, 1988), in which aircraft take off, landing, and abort landing problems were formulated for nonlinear system of vertical channel. These papers were followed by the works of other authors Leitmann and Pandey (1991), Bulirsch et al. (1991a), Bulirsch et al. (1991b), Botkin et al. (1984), and Patsko et al. (1994), in which various methods of optimal control theory and differential games were applied to similar problems.

Of course, the works mentioned were of research nature. The design of autopilots for different stages of aircraft motion is traditionally based on methods of the automatic regulation theory being intensively developed in the 1930s–1950s and relied heavily on achievements of stability theory for linear systems. The outcome of these algorithms when applied in mathematical modeling in the presence of wind disturbances is essentially worse compared to algorithms based on comprehensive mathematical optimal control theory (including differential game theory) that directly accounts for geometric bounds on deviations of steering mechanisms. However, such novel algorithms require as input data almost all state variables, many of which are difficult or even impossible to measure. The interest to study and to apply pursuit-evasion games to aircraft problems could be increased nowadays based on new realistic settings and comprehensive efficient numerical methods of differential game theory.

Many applied-oriented works accomplished in the 1980s are presented in the collection (Yavin et al. 1987) edited by Y. Yavin, M. Pachter, and E. Y. Rodin.

8. The presentation in this chapter is as follows. We describe several mathematical problems that should be regarded as model problems assimilating principal aspects of very important practical problems.

In Sect. 2, we consider the time-optimal homicidal chauffeur game and its modifications. For each problem, we give the statement and the corresponding references to journal publications. Then our results on numerical construction of the level sets of the value function and, for one of the modifications, results of modeling optimal strategies are presented.

In Sect. 3, a space interception problem with linear dynamics is considered. Here again, the main attention is paid to the computation of the level sets of the value function (maximal stable bridges with a given value of miss). It is stressed that the level sets in the space time $\times$ state vector can have narrow throats with complex geometry. Investigation of such seemingly pure mathematical peculiarities is important for understanding the structure of the solvability sets of the interception problem. Significant attention is paid to adaptive control of the first (minimizing) player, being developed by us for the case where, according to the statement of the problem, no geometric constraint on the control of the second player is specified.

Thus, our objective is to give a vivid presentation of two canonical classes of differential games and applied problems, which can be solved using numerical methods developed for these classes. The presentation is accompanied by a large number of figures to demonstrate the structure and peculiarities of the value function.

In the last subsections of Sects. 2 and 3, we mention comprehensive complex applied problems related to time-optimal games and to games with linear dynamics and stress that their investigation requires development of new efficient numerical methods.

The investigation of applied model problems considered in this chapter was initiated and to a large extent explored by the outstanding mathematicians: R. Isaacs, J. Breakwell, A. Merz, and J. Shinar. We place their photographs here (Figs. 22.1 and 22.2).

The Introduction and Sect. 2 of this chapter are written by V. S. Patsko and V. L. Turova, Sect. 3 and Conclusion are prepared by S. S. Kumkov and V. S. Patsko.

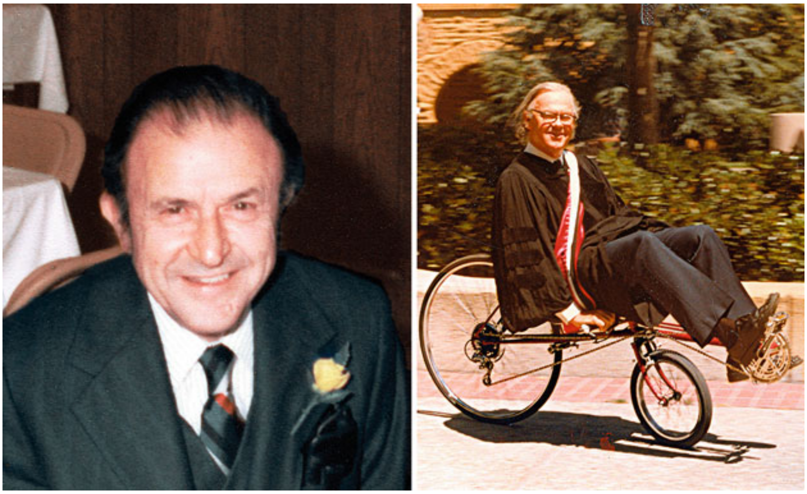


Fig. 22.1 Rufus Isaacs (1979) and John Valentine Breakwell ($\approx$ 1986)

Fig. 22.2 Antony Merz (2008) and Josef Shinar (2007)

2 Time-Optimal Problems: Homicidal Chauffeur Game and Its Modifications

Among pursuit-evasion games, the most popular ones are time-optimal problems, where one player wishes to minimize and another wishes to maximize the terminal time of the game. In turn, the most famous problem among time-optimal problems is the homicidal chauffeur game. It formed the basis of the book by R. Isaacs. After the publication of this book, a huge number of applied studies were performed on the homicidal chauffeur game and its modifications.

The significance of the problem is the following. On one side, many practical situations fall under this mathematical description, e.g., the abovementioned conflict situation between a controlled torpedo and an evading small motor boat or an aircraft pursuing a helicopter in a horizontal plane and so on. On the other hand, the problem is formulated mathematically in such a way that after passing to reduced coordinates we deal with two state variables. This was important in the middle of the last century as numerical investigation of the problem was not yet possible. Thus, intuition could help here. Also presently, when the application of numerical methods is not uncommon, systematic numerical analysis for various values of parameters with the aim to reveal regular and singular parts of solutions can really be performed only in the case where the problem is reduced to the one with two state variables. A third thing to mention is that the problem is very interesting as a test example when developing various numerical methods of differential game theory.

2.1 Dynamics of Conflicting Objects

Two moving objects, a "pedestrian" and a "car," present in the game.

The pedestrian is a non-inertia point object with coordinates x_e, y_e in the plane, which can change the direction of the motion instantaneously. The magnitude of the velocity v is bounded from above by a given number. Using differential equations, this can be expressed in the form

$$\begin{aligned} \dot{x}_e &= v_1, \\ \dot{y}_e &= v_2, \quad v = (v_1, v_2)', \ |v| \le \rho. \end{aligned} \tag{22.1}$$

Such an object was called by R. Isaacs as "object with simple motion." Here and below, the prime means transposition.

Dynamics of the car:

$$\begin{aligned} \dot{x}_p &= w \sin\theta, \\ \dot{y}_p &= w \cos\theta, \\ \dot{\theta} &= wu/R, \ \ |u| \le 1. \end{aligned} \tag{22.2}$$

Here x_p, y_p are the coordinates of the point object, θ is the angle specifying the direction of the velocity vector (measured clockwise from the $+y_p$-axis), $w = \text{const}$

is the given velocity magnitude, u is the scalar control (bounded in absolute value by 1), and R is the minimum turn radius.

By normalizing time and geometric coordinates, one can achieve $w = 1$, $R = 1$. In the new dimensionless variables, the dynamics of the objects is written as follows:

$$\begin{aligned} P: \dot{x}_p &= \sin\theta, & E: \dot{x}_e &= v_1, \\ \dot{y}_p &= \cos\theta, & \dot{y}_e &= v_2, \\ \dot{\theta} &= u, \ \ |u| \le 1; & v &= (v_1, v_2)', \ \ |v| \le \nu. \end{aligned} \tag{22.3}$$

The constraint on v in (22.3) has changed compared to (22.1) because of the joint normalization for (22.1) and (22.2). In the following, the notation $Q = \{v : |v| \le \nu\}$ will be often used.

It was R. Isaacs who introduced the name "car" for the object (22.2). After the normalization and assuming that the velocity $w = 1$, the path length run by such an object is $wT = T$, where T is the elapsed time. Therefore, minimization of time for the object (22.2) is equivalent to minimization of the path length.

A. A. Markov addressed in his paper (Markov 1889) four optimization problems for railway track laying. In the first two of them, he assumed that the movement along the railway track is performed with constant velocity, the curvature radius of the railway track is bounded, and the path length is used as an optimum criterion. This means that he studied practically time-optimal problems for the object with dynamics (22.2) but using other terms.

L. Dubins published the paper (Dubins 1957) on the line of shortest length, which connects two points in the plane. The lines whose curvature is bounded from below by the same number were admitted for comparison; herewith each line should have the same given outgoing direction at the initial point and the same given incoming direction at the terminal point. Obviously, movement along such lines is also described by system (22.2). It came about in works on theoretical robotics that objects with dynamics (22.2) are often called "Dubins' car."

The next in complexity model is the car model from the paper by J. Reeds and L. Shepp (1990):

$$\begin{aligned} \dot{x}_p &= w\sin\theta \\ \dot{y}_p &= w\cos\theta \\ \dot{\theta} &= u, \qquad |u| \le 1, \ |w| \le 1. \end{aligned} \tag{22.4}$$

The control u determines the angular velocity of motion. The control w is responsible for the instantaneous change of the linear velocity magnitude. In particular, the car can instantaneously change the direction of motion to the opposite one. A non-inertia change of the linear velocity magnitude is a mathematical idealization. But, citing (Reeds and Shepp 1990, p. 373), "for slowly moving vehicles, such as carts, this seems like a reasonable compromise to achieve tractability."

It is natural to consider problems where the range for changing the control w is $[a, 1]$. Here, $a \in [-1, 1]$ is the parameter of the problem. If $a = 1$, Dubins' car is obtained. For $a = -1$, one arrives at Reeds-Shepp's car.

2.2 Dynamics in Reduced Coordinates

Place the origin of the reduced coordinates x, y to the position of player P. Let $h(t)$ be a unit vector in the direction of motion of player P at time t. The orthogonal to $h(t)$ unit vector is denoted by $k(t)$ (see Fig. 22.3). We have

$$h(t) = \begin{pmatrix} \sin\theta(t) \\ \cos\theta(t) \end{pmatrix}, \quad k(t) = \begin{pmatrix} \cos\theta(t) \\ -\sin\theta(t) \end{pmatrix}.$$

Differentiating the relations

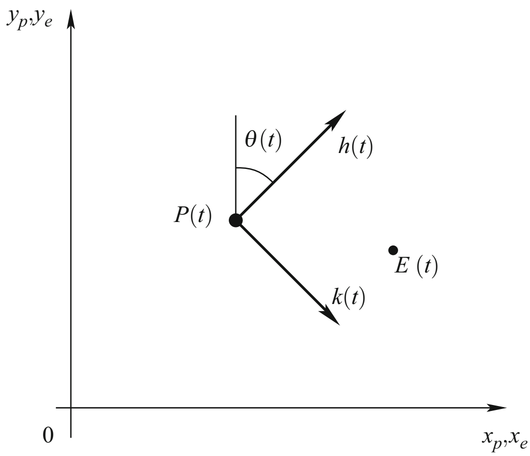
$$\begin{aligned} x(t) &= \cos\theta(t)(x_e(t) - x_p(t)) - \sin\theta(t)(y_e(t) - y_p(t)), \\ y(t) &= \sin\theta(t)(x_e(t) - x_p(t)) + \cos\theta(t)(y_e(t) - y_p(t)), \end{aligned}$$

we turn from system (22.3) to system

$$\begin{aligned} &\dot{x} = -yu + v_x, \\ &\dot{y} = xu - 1 + v_y, \\ &|u| \le 1, \ v = (v_x, v_y)', \ |v| \le \nu. \end{aligned} \tag{22.5}$$

Here $v_x = v_1\cos\theta - v_2\sin\theta$, $v_y = v_1\sin\theta + v_2\cos\theta$. Note that the form of the circular constraint on the control of player E remains the same in the reduced

Fig. 22.3 Movable reference system



coordinates. However, this might be not the case if the geometric constraint on the control of player E in (22.3) would be of another kind.

In the case where player P is described by (22.4), let the axis y of the relative coordinate system be directed toward the forward motion of the car to obtain

$$\begin{aligned} \dot{x} &= -yu + v_x, \\ \dot{y} &= xu - w + v_y, \\ |u| &\le 1, \ w \in [a, 1], \ v = (v_x, v_y)', \ |v| \le \nu. \end{aligned} \tag{22.6}$$

2.3 Isaacs' Method for Games of Kind and Games of Degree. Iterative Viability Methods

The problems considered in Sects. 2.5 and 2.6 were originally solved using Isaacs' method. The problem from Sect. 2.7 was initially investigated using an iterative method based on the concept of viability trajectories. Below, we give a schematic description of these methods.

1. In classical mathematics, smooth solutions to first-order partial differential equations are searched using Cauchy characteristics (see, e.g., Courant 1962; Evans 1998; Melikyan 1998).

Consider a partial differential equation

$$F\Big(x, J, \frac{dJ}{dx}\Big) = 0, \quad x \in R^n. \tag{22.7}$$

Here, F is a scalar function, J is the unknown function $x \to J(x)$, and $\dfrac{dJ}{dx}$ is its derivative. Depending on the context, the derivative of the scalar function of a vector argument will be considered either as a row matrix or a column matrix. Assume that $F \in C^2$. Also, let the function J satisfy a given boundary condition. Typically, values $J(x)$ are defined on a smooth manifold of dimension $n - 1$. With some additional regularity condition, the theorem on local parametrization of the graph of the function $J \in C^2$ holds. The parametrization is performed using the system of ordinary differential equations:

$$\dot{x} = F_\psi(x, s, \psi), \ \ \dot{s} = \langle \psi, \ F_\psi(x, s, \psi) \rangle, \ \ \dot{\psi} = -F_x(x, s, \psi) - \psi F_s(x, s, \psi), \tag{22.8}$$

where the angular brackets denote the scalar product of two vectors.

The functions $t \to x(t)$ and $t \to s(t)$, being the solution (together with the function $t \to \psi(t)$) to system (22.8), define curves carpeting the graph of the function J. Initial values $x(t_*)$, $s(t_*)$, $\psi(t_*)$ correspond to the boundary condition on the function J and depend on some parameter ξ of dimension $n - 1$. The scalar variable t taking values close to t_* is also a parameter. The functions $t \to x(t)$, $t \to s(t)$, and $t \to \psi(t)$ are called characteristics of equation (22.7).

Necessity. Given a solution J to equation (22.7) $\Longrightarrow$ system (22.8) defines a parametrization of the graph of the function J. Here, $s(t) = J(x(t))$, $\psi(t) = \frac{dJ}{dx}(x(t))$.

Sufficiency. Given functions $t \to x(t)$, $t \to s(t)$, and $t \to \psi(t)$, satisfying system (22.8) and relation $F(x(t), s(t), \psi(t)) = 0$. The initial values $x(t_*)$, $s(t_*)$, and $\psi(t_*)$ of these functions, being dependent on a parameter ξ, have the sense of a point on some sub-manifold of the initial manifold, the value $J(x(t_*))$ of the unknown function, and the value $\frac{dJ}{dx}(x(t_*))$ of its derivative, respectively. The regularity condition ensures that only one x-characteristics goes through each point $x = x(t)$ of this characteristics in some neighborhood of the distinguished sub-manifold. Then, the function J, being a solution to (22.7), is constructed in such neighborhood. The curves $t \to x(t)$ and $t \to s(t) = J(x(t))$ follow the graph of this function. Here, $\psi(t) = \frac{dJ}{dx}(x(t))$.

The proof of the sufficiency is constructive and gives a receipt for constructing the function J.

If the function F in (22.7) does not depend on J, then the system (22.8) is simplified as follows:

$$\dot{x} = F_\psi(x, \psi), \;\; \dot{s} = \langle \psi, \; F_\psi(x, \psi) \rangle, \;\; \dot{\psi} = -F_x(x, \psi). \tag{22.9}$$

The first and third equations in (22.9) are separated from the second equation and can be integrated independently. Once the functions $t \to x(t)$ and $t \to \psi(t)$ are found, the function $t \to s(t)$ can be then determined.

2. R. Isaacs applied the Cauchy characteristic method for solving differential games.

For clarity, let us consider time-optimal differential game in the plane with the dynamics:

$$\dot{x} = f_1(x, u) + f_2(x, v), \;\; x \in R^2, \; u \in P, \; v \in Q, \tag{22.10}$$

and a closed target set M. The first player has the control u at his disposal and minimizes the transfer time of system (22.10) to the set M. The second player being responsible for the control v has the opposite interest. The controls u and v are bounded by geometric constraints. The differential game is considered in the class of feedback controls.

In the theory of differential games, existence of the value function $x \to V(x)$ is established. Here, we do not go into details of a particular formalization. In typical examples, the value function is not differentiable or even continuous for $x \in R^2 \backslash M$. However, there exist regions (cells) in which $V \in C^2$. In each of such regions, the derivative of the function $t \to V(x(t))$ along an arbitrary trajectory $t \to x(t)$ can be computed:

$$\frac{dV}{dx}(x(t)) \cdot \dot{x}(t) = \frac{dV}{dx}(x(t)) \cdot (f_1(x, u(t)) + f_2(x, v(t))).$$

Accounting for the nondeterioration principle along optimal trajectories, we obtain

$$\min_{u \in P} \frac{dV}{dx}(x) f_1(x, u) + \max_{v \in Q} \frac{dV}{dx}(x) f_2(x, v) = -1. \tag{22.11}$$

Relation (22.11) is just R. Isaacs' transition rule (principle) written in terms of derivatives. Introducing the notation for the Hamiltonian

$$H(x, \psi, u, v) = \psi'(f_1(x, u) + f_2(x, v)),$$

we rewrite relation (22.11) as follows:

$$1 + \min_{u \in P} \max_{v \in Q} H(x, \frac{dV}{dx}(x), u, v) = 0. \tag{22.12}$$

Thus, the value function V satisfies the partial differential equation (22.12). Computing the extremal elements u^* and v^* in (22.12), assume additionally their smoothness in x. We obtain

$$1 + H(x, \frac{dV}{dx}(x), u^*(x), v^*(x)) = 0. \tag{22.13}$$

The equations of characteristics for (22.13) have the form

$$\dot{x} = H_\psi(x, \psi, u^*(x), v^*(x)), \quad \dot{\psi} = -H_x(x, \psi, u^*(x), v^*(x)). \tag{22.14}$$

Having under some boundary conditions a solution to (22.14), we obtain the functions $u^*(t) = u^*(x(t), \psi(t))$ and $v^*(t) = v^*(x(t), \psi(t))$. With that, relation (22.13) is fulfilled:

$$1 + H(x, \psi(t), u^*(t), v^*(t)) = 0. \tag{22.15}$$

Thereby,

$$\begin{aligned} &\min_{u \in P} \psi'(t) \cdot f_1(x(t), u) = \psi'(t) \cdot f_1(x(t), u^*(t)), \\ &\max_{v \in Q} \psi'(t) \cdot f_2(x(t), v) = \psi'(t) \cdot f_2(x(t), v^*(t)). \end{aligned} \tag{22.16}$$

Relationships (22.16) together with equations

$$\dot{x} = H_\psi(x, \psi, u^*(t), v^*(t)), \quad \dot{\psi} = -H_x(x, \psi, u^*(t), v^*(t)) \tag{22.17}$$

are similar to Pontryagin's maximum principle for optimal control problems. Here, they were derived in the form of necessary conditions based on the Cauchy characteristic method and on the assumption about smoothness of the function V and extremal elements u^* and v^* in (22.12).

We can also consider the question of constructing the unknown function V. Suppose that values of the function and its derivative are given on some curve. Consider the trajectories of system (22.10) that satisfy the minimax principle (22.16), (22.17), and, additionally, together with the function $t \to \psi(t)$, relation (22.15).

Assume that only a single trajectory satisfying the first equation of system (22.17) goes through each point x of some region that includes a curve on which values of the function V and its derivative are given. Then for every point x on the distinguished curve, we reconstruct the state $\xi(x)$ of the corresponding x-trajectory and the time $t(x)$ of passing through the point x. For every point x, let $u(x) = u^*(t)$, where $u^*(t)$ is defined by $\xi(x)$ and $t(x)$. Similarly, we introduce $v(x) = v^*(t)$. Assume that the functions $u(x)$ and $v(x)$ are continuously differentiable. Then the system

$$\begin{aligned} \dot{x} &= f_1(x, u(x)) + f_2(x, v(x)) = H_\psi(x, \psi, u(x), v(x)), \\ \dot{\psi} &= -H_x(x, \psi, u(x), v(x)) \end{aligned}$$

can be put in correspondence with system (22.16), (22.17). With that,

$$1 + H(x(t), \psi(x(t)), u(x(t)), v(x(t))) = 0.$$

Further, the unknown function V is reconstructed in the region considered using the same technique as in the proof of sufficiency of the Cauchy characteristic method.

However, we cannot say that this is the value function of the differential game. This is merely a function that satisfies the partial differential equation, the given values, the values of derivatives, and boundary condition on the distinguished curve. To obtain the value function, a solution in the whole space should be found. If the reconstruction region spans the whole space, the problem is solved. However, this is a very rare situation. Typically, finding the value function (and associated with it optimal strategies) with Isaacs' method requires subsequent (backward in time) covering the whole game space with cells filled out with regular x-characteristics.

In more detail, the construction of the value function with the method of characteristics is described, apart from R. Isaacs' book, in the paper (Berkovitz 1994) and in the books (Başar and Olsder 1995; Lewin 1994).

3. In R. Isaacs' approach to differential games with complete information, two main ideas can be emphasized.

The first idea is related to believing that the solution region of the differential game typically is divided into cells in which interior the value function is smooth and can be found using the Cauchy characteristic method for a properly derived first-order partial differential equation. But how can one set boundary conditions for each cell? The only possible way is primary analysis of terminal conditions of the differential game. For example, for time-optimal differential game in the plane with the target set M, the parts of the boundary of M which can be penetrated by optimal trajectories should be determined. We accept that the value function on such parts is equal to zero and try to compute the derivative of the value function in x on these parts. Then Cauchy's characteristics are emitted *backward* in time for each of the parts. Analysis of x-characteristics enables to determine where such backward construction should be stopped, and examination of next cells should be started.

The second idea by R. Isaacs concerns singular lines (in general case, singular surfaces). Singular lines are the curves on which optimal trajectories lose usual regularity. For example, the dispersal line is a curve each point of which is approached by two backward x-characteristics with equal optimal result values. The universal line is a curve which, to the contrary, is leaved by two backward optimal trajectories. Hence, the direct time-optimal trajectories approach the universal line. Thereby, there exists an optimal motion that goes along the universal line, but the corresponding trajectory (being the universal line itself) is not a regular Cauchy's x-characteristic.

R. Isaacs discovered a new type of singular lines (surfaces), which he called equivocal. Optimal trajectories approach such lines in direct time; then each optimal trajectory splits into two branches: the first one comes with a kink to the other side of the singular line, and the second one goes along the singular line. The value function is continuous but not differentiable on equivocal lines. R. Isaacs mentioned that equivocal lines are inherent to differential games. In contrast to other singular lines, curves with such properties cannot exist in problems where only one player optimizing the dynamic system behavior is present.

One more types of singular lines called barriers are curves where the value function is discontinuous.

The book Isaacs (1965) contains many remarkable pictures explaining the sense of singular lines and singular surfaces. R. Isaacs described how singular lines can arise in the backward constructions.

Detection and construction of singular lines is a key to consideration of the next cell in Isaacs' method. Thus, Isaacs' approach is a backward construction of cells based on the analysis of arising singular lines (surfaces). The cells are filled with optimal trajectories. By constructing cells backward in time, we hope to cover with them the whole game space. The obtaining value function is, as a rule, not differentiable or even continuous.

4. The theory of differential games essentially influenced the development of the theory of partial differential equations. In the beginning of the 1980s, new notions of generalized solutions of the first-order partial differential equations have been introduced. The generalized solution suggested by M. G. Crandall and P.-J. Lions was called viscosity solution (Crandall et al. 1984; Crandall and Lions 1983; Lions 1982), whereas the concept proposed by A. I. Subbotin was specified as minimax solution (Subbotin 1980, 1984). The equivalence of these two notions was established, and new concepts of generalized solutions were developed to cover formulations of typical differential games with non-smooth or even discontinuous value function (Subbotin 1995). Many facts from the theory of singular lines and surfaces revealed in the theory of differential games earlier were reformulated for

generalized solutions of partial differential equations. On this way, A. A. Melikyan developed (Melikyan 1998) a theory of singular surfaces for the first-order partial differential equations. Numerical methods for solving Hamilton-Jacobi equations associated with differential games and based on the concept of viscosity solutions are being developed (see, e.g., Botkin et al. 2011; Chen et al. 2015; Falcone 2006; Grigor'eva et al. 2000).

5. Considering differential games, R. Isaacs distinguished between games of kind and games of degree. Let us explain this using differential game with dynamics (22.10) and a closed target set M. The terminal time of the control process is not fixed.

In the game of kind, we are interested in finding the set $\mathcal{A}$ of all initial states x_0 from which the first player guarantees approaching the set M within a finite time, using a feedback control $u(x)$ implemented in a discrete control scheme. For initial states in the set $R^2 \backslash \mathcal{A}$, such a guarantee is absent. Surely, in frames of accurate formalization, one should correctly specify what "approaching M" means. Namely, whether "approaching M" implies precise transition to M or transition to an arbitrarily small neighborhood of M. We will not do this here. In any case, there are only two possible (guaranteed) outcomes in the game of kind: yes (approaching is possible) or no (approaching is not possible).

In the game of degree, the first player minimizes the time of approach M. Here, compared to the game of kind, one should determine a minimum guaranteed time of transition to M for each initial state $x_0 \in \mathcal{A}$. It is established that this time coincides with the best guaranteed time of the second player (who maximizes the time of transition to M); therefore, one can speak about the value function $V(x_0)$. For initial states $x_0 \notin \mathcal{A}$, nothing new compared to the game of kind arises.

A common feature for the games of kind and degree is the construction of the boundary of the set $\mathcal{A}$. In the game of degree, we have $V(x_0) < \infty$ for $x_0 \in \mathcal{A}$, and $V(x_0) = \infty$ for $x_0 \in R^2 \backslash \mathcal{A}$. Therefore, the curves constituting the boundary of the set $\mathcal{A}$ are barrier lines.

R. Isaacs formulated the main property of smooth curves that comprise the boundary of $\mathcal{A}$. Namely, he considered the following relation:

$$\min_{u \in P} \max_{v \in Q} \ell'(x)(f_1(x,u) + f_2(x,v)) = 0. \tag{22.18}$$

Here $\ell(x)$ is the normal to the smooth curve at the point x. Let the side of the curve to which the normal is directed be referred as negative and the opposite side be indicated as positive. R. Isaacs called the smooth curves satisfying the relation (22.18) the semipermeable curves. Families of semipermeable curves are defined only by dynamics of the game (including constraints on the controls) and do not depend on the objectives of the players. Having the dynamics of the game, we can perform an analysis of families of such curves in the plane in advance to use them later for constructing barriers. Each semipermeable curve is often bounded. Considering one of the two directions of moving along the curve, one can specify its start point and end point. For each point x in the plane, one of the following possibilities is realized: there is no semipermeable curve, there is only one semipermeable curve, or there are several semipermeable curves going through x.

Relation (22.18) can be considered as a partial differential equation with respect to unknown scalar function $x \to J(x)$:

$$\min_{u \in P} \max_{v \in Q} \frac{dJ}{dx}(x)(f_1(x,u) + f_2(x,v)) = 0. \tag{22.19}$$

The family of smooth semipermeable curves is a family of x-characteristics of equation (22.19). On each x-characteristic, the value $J(x)$ is constant. The whole family can be found by specifying some (not arbitrary) curve in the plane and values of derivative $\frac{dJ}{dx}(x)$ in points of this curve, so that (22.19) holds.

It is useful to distinguish families of semipermeable curves of the first and second type. For families of the first type, when constructing semipermeable curves backward in time, the vector of moving direction along the curve is related to the vector ℓ by a clockwise rotation through the angle $\pi/2$. For families of semipermeable curves of the second type, the corresponding vectors are related by a counterclockwise rotation through the angle $\pi/2$.

Remark. The greater the number of families of semipermeable curves for a given dynamics is, the more complex the differential game with a particular payoff is.

6. Smooth semipermeable curves are the basis for solving games of kind in Isaacs' method. Suppose that the target set M is convex and has a smooth boundary. Visiting the boundary of M, find those parts of it through which the first player guarantees the transition of system (22.10) to the interior of the set M for any counteraction of the second player. R. Isaacs called such pieces "the usable part" (UP). Let for simplicity UP consists of a single arc. Taking an arbitrary internal point x of this arc and denoting by $\ell(x)$ the vector of outward normal to the set M at this point, write down the inequality:

$$\min_{u \in P} \ell'(x) f_1(x,u) + \max_{v \in Q} \ell'(x) f_2(x,v) < 0, \tag{22.20}$$

which provides a guaranteed approach of the interior of the set M by the trajectories of system (22.10) not only from the point x but also from the points outside M, being close to M. For two boundary points x_*, x^* of this arc (BUP), we obtain

$$\min_{u \in P} \ell'(x) f_1(x,u) + \max_{v \in Q} \ell'(x) f_2(x,v) = 0. \tag{22.21}$$

In the coarse case, for any other points $x \in \partial M$ outside the arc $[x_*, x^*]$, the inequality >0 holds for the left part of (22.21). Then the arc $[x_*, x^*]$ is a "gate," through which entering M is only possible. The next step is to construct smooth barrier lines, being "boards of the way" leading to the arc $[x_*, x^*]$, from the points x_*, x^*. Herewith, with respect to each of the two boards, the first player is able to prevent the transition of the system from the positive side of the board (faced to the way) to its negative side, no matter how the second player acts. Conversely, the second player is able to prevent the transition from the negative side to the positive side. Thus, one of the boards is smooth semipermeable curve of the first type and the other one is of the second type.

The most simple situation in the coarse case is realized, when the semipermeable curves under consideration intersect without tangency. In this case, it can be often proved that the set $\mathcal{A}$ of successful termination of the game of quality with respect to the first player is the union of the set M and the part of the plane bounded by the curve UP and by the pieces of two semipermeable curves between their start points x_*, x^* and the intersection point.

In the paper Patsko (1975), the game of kind for differential games in the plane with arbitrary linear dynamics, a point target set M, scalar control of the first player with a bounded absolute value, and an arbitrary convex polygonal constraint on the control of the second player is completely solved. An algorithm for constructing the set $\mathcal{A}$ for games of kind in the plane in the case of complex roots of the characteristic polynomial of linear system and arbitrary polygonal constraints on the controls of the first and second players is described in Turova (1984).

7. Isaacs' method for games of kind and games of degree uses constructions that provide a precise answer without any iterations. The error obtained is determined by only inaccuracy in the implementation of prescribed operations.

In the middle of the 1970s, A. G. Chentsov proposed the method of programmed iterations for various classes of differential games (Chentsov 1976, 1978a,b). In this method, the solution to the problem in the form of some set is obtained as a result of iterative descent to this set from above. The method is based on the concepts used in the scientific school of N. N. Krasovskii. As it was mentioned in the Introduction, the central notion in this school is the notion of maximal stable bridge. For example, for time-optimal games with stationary dynamics and a closed target set M, the maximal stable bridge terminating at the time T on the set M is the collection of all positions (t_*, x_*), from which the first player by discriminating the second one can bring the state vector x to the set M within the time not exceeding $T - t_*$. The corresponding "tube" W (maximal stable bridge) in the space t, x $(t \le T)$ contains the cylinder $\{(t, x) : t \le T, \ x \in M\}$. It is known that any t-section $W(t)$ is the level set (Lebesgue set) of the value function $x \to V(x)$ of the time-optimal game, i.e., $W(t) = \{x : V(x) \le T - t\}$.

For the construction of the set W on some interval $[\bar{t}, T]$, $\bar{t} < T$, A. G. Chentsov proposed the following iterative procedure. The iterations start from the set $W^{(0)} = [\bar{t}, T] \times R^n$, where R^n is the phase space of the dynamic system. Then a closed subset $W^{(1)} \subset W^{(0)}$ is distinguished by the following property: for any position $(t_*, x_*) \in W^{(1)}$, any constant control $v \in Q$ of the second player on the interval $[t_*, T]$, the first player can choose his open-loop control $u(\cdot)$, so that the trajectory

$(t, x(t))$ generated by the controls $u(\cdot)$ and v comes to the cylindric set $\{(t, x) : t_* \le t \le T, x \in M\}$. Then the set $W^{(2)} \subset W^{(1)}$ is introduced, so that for any point $(t_*, x_*) \in W^{(2)}$ and any constant control $v \in Q$, the first player can choose his open-loop control such that the trajectory of the system generated by the controls $u(\cdot)$ and v satisfies the constraint $(t, x(t)) \in W^{(1)}$. Thus, the set $W^{(1)}$ plays the role of a closed state constraint. Further, the set $W^{(3)} \subset W^{(2)}$ is constructed, where the set $W^{(2)}$ being the state constraint, and so on. It is proved that $W^{(i)} \to W$ on $[\bar{t}, T]$ as $i \to \infty$.

A. G. Chentsov interpreted the method of programmed iterations as a method explaining the structure of the differential game and helping to establish particular theoretical facts. The method was formulated and proved for very wide variety of problems, but A. G. Chentsov did not attempt to develop efficient numerical procedures based on this approach.

P. Cardaliaguet, M. Quincampoix, and P. Saint-Pierre proposed a conceptually similar iterative method (Cardaliaguet et al. 1995, 1999), which, however, was directed to the numerical implementation. The method uses ideas of viability theory developed by J.-P. Aubin (1991). The most clear application of the method is its employment in the game of kind for the construction of maximal set $\mathcal{A}^*$, from every point of which the second player (by discriminating the first player) guarantees the evasion of the dynamic system from approaching the set M for infinite time. The iterations are computed in the set $K = R^n \backslash M$. It is required that this set be closed. Hence, the set M is supposed to be open.

Let us explain the iterations. Replace the original continuous-time dynamics by a discrete one. For all $x \in R^n$ and $u \in P$, introduce the set $G_\varepsilon(x, u)$, which is interpreted in the context of discrete dynamics as a reachable set with respect to $v(\cdot)$ for a fixed control u of the first player on some small time interval of the length ε. The discrete dynamics is chosen in a way that the set $G_\varepsilon(x, u)$ contains an analogous reachable set of the original system. The following sequence of sets is introduced:

$$\begin{cases} K_\varepsilon^0 = K, \\ K_\varepsilon^{i+1} = \{x \in K_\varepsilon^i : \forall u \in P,\ G_\varepsilon(x, u) \cap K_\varepsilon^i \ne \varnothing\}, \quad i = 0, 1, \ldots \end{cases} \tag{22.22}$$

The set K_ε^{i+1} is the maximal subset of the set K_ε^i, from any point x of which the first player, by showing his constant control $u \in P$, cannot steer the system away from the set K_ε^i at the end of the interval ε. It is proved that the sequence K_ε^i converges from above to some set $\overrightarrow{Disc}_{G_\varepsilon}(K)$, which is called a *discrete discriminating kernel*.

This set possesses the following *viability* (stability) property: if the first player shows his constant control u for the interval ε in advance, then, in frames of discrete approximating dynamics, the second player holds the motion in the set $\overrightarrow{Disc}_{G_\varepsilon}(K)$ for infinite time. For points $x \in K \backslash \overrightarrow{Disc}_{G_\varepsilon}(K)$, on the contrary, there exists a positional control method for choosing control u with the step ε, such that the trajectory will approach the set M within a finite time for any actions of the second player.

We described solving the game of kind for discrete approximating dynamics for a fixed ε. Then grids K_h and P_h on the sets K and P with the step size h are introduced, and a grid approximation $\Gamma_{\varepsilon,h}(x_h, u_h)$ of the reachable set of the original system with respect to $v(\cdot)$ is considered. The approximation $\Gamma_{\varepsilon,h}$ depends on the parameters ε and h. For $i \to \infty$, the iterations $K^i_{\varepsilon,h}$ similar to (22.22) give a *fully discrete discriminating kernel* $\overrightarrow{Disc}_{\Gamma_{\varepsilon,h}}(K_h) \subset K_h$. This set converges (under certain relation between ε and h) to the ideal set $\mathcal{A}^*$ as $\varepsilon \to 0$, $h \to 0$. It is proved that $\mathcal{A} = R^n \setminus \mathcal{A}^*$.

In numerical implementations, essential difficulties arise, when the grid K_h cannot be chosen bounded. Therefore, revealing cases where the grid K_h (as well as P_h) can be taken finite is of great interest.

In time-optimal games (games of degree), the set M is assumed to be closed. The first player, using the control u, minimizes the time of approaching M, the second player has the opposite objective. The time t is considered as an additional state variable (even for a stationary system). A discrete approximating dynamics defined by the parameter ε and grids with the step width h in variables $t \geq 0$ and x in R^n as well as a grid Q_h in the set Q are introduced. Let Z_h be a grid obtained in $\{t \geq 0\} \times R^n$. A grid approximation $\Gamma_{\varepsilon,h}(t_h, x_h, v_h)$ of the reachable set with respect to $u(\cdot)$ for the original system is considered, depending on the parameters ε and h.

The following decreasing sequence of sets is introduced:

$$\begin{cases} K^0_{\varepsilon,h} = Z_h, \\ K^{i+1}_{\varepsilon,h} = \{(t,x) \in K^i_{\varepsilon,h} : \forall v_h \in Q_h,\ \Gamma_{\varepsilon,h}(t_h, x_h, v_h) \cap K^i_{\varepsilon,h} \neq \varnothing\}, \quad i = 0, 1, \dots \end{cases}$$

It is proved that $\lim\limits_{i\to\infty} K^i_{\varepsilon,h}$ exists. This limit is denoted as $\overrightarrow{Disc}_{\Gamma_{\varepsilon,h}}(Z_h)$. It is shown that the set $\overrightarrow{Disc}_{\Gamma_{\varepsilon,h}}(Z_h)$ converges to the epigraph of the value function of time-optimal game as $\varepsilon \to 0$, $h \to 0$ (under consistent relation between ε and h). In so doing, the problem of constructing the value function of time-optimal game is solved.

An interesting observation is done in paper (Botkin 1993). Consider a differential game $\dot{x} = f(x, u, v)$, where f globally possesses standard properties formulated in Krasovskii and Subbotin (1988). Let a compact $K \subset R^n$ be the state constraint, and $N = (-\infty, T] \times K$. A subset $Z \subset N$ is called stable if the player "v" can ensure the inclusion $(t, x(t)) \in Z$, $t \in [t_0, t_0 + \varepsilon]$, whenever $\big(t_0, x(t_0)\big) \in Z$, and the player "$u$" shows his constant control on the time interval $[t_0, t_0 + \varepsilon]$ (cf. the case of stable subsets in R^n). If there exists at least one stable subset of N, then there exists a maximal stable subset, say W, of N. If all time sections, $W(t)$, $t \in (-\infty, T]$, are nonempty, then $W(t)$ converges to the discriminating kernel $\overrightarrow{Disc}(K)$ in the Hausdorff metric as $t \to -\infty$. This gives rise (see Botkin 1993) to a recurrent algorithm resembling (22.22).

2.4 Families of Semipermeable Curves

When solving differential games in the plane, it is useful (as it was mentioned in Sect. 2.3) to carry out a preliminary study of families of smooth semipermeable curves that are determined by the dynamics of the controlled system. The knowledge of these families for time-optimal problems allows to verify the correctness of the construction of barrier lines on which the value function is discontinuous.

A smooth semipermeable curve is a line with the following preventing property: one of the players can prevent crossing the curve from the positive side to the negative one, the other player can prevent crossing the curve from the negative side to the positive one.

Let us explain the meaning of semipermeable curves. Introduce the minimax Hamiltonian of the game as follows:

$$H(\ell, z) = \min_u \max_v \ell' f(z, u, v) = \max_v \min_u \ell' f(z, u, v),\ z = (x, y)' \in R^2, \ell \in R^2.$$

Here $f(z, u, v) = p(z)u + v + g$, $p(z) = (-y, x)'$, $g = (0, -1)'$. Fix $z \in R^2$ and consider ℓ such that $H(\ell, z) = 0$. Denote $u^* = \arg\min_u \ell' f(z, u, v)$, $v^* = \arg\max_v \ell' f(z, u, v)$. It holds: $\ell' f(z, u^*, v) \le 0$ for any $v \in Q$ and $\ell' f(z, u, v^*) \ge 0$ for any $u \in [-1, 1]$. This means that the direction $f(z, u^*, v^*)$ which is orthogonal to ℓ separates the vectograms $\Phi(v^*) = \bigcup_{u \in [-1,1]} f(z, u, v^*)$ and $\Xi(u^*) = \bigcup_{v \in Q} f(z, u^*, v)$ of the first and the second players (Fig. 22.4). Such a direction is called semipermeable. Thus, the semipermeable directions are defined by the roots of the equation $H(\ell, z) = 0$. We will distinguish the roots from "–" to "+" and the roots from "+" to "–". When defining these roots, we will suppose that $\ell \in E$, where E is a closed polygonal line around the origin. We say that ℓ_* is a root from "–" to "+" if $H(\ell_*, z) = 0$ and $H(\ell, z) < 0$ $(H(\ell, z) > 0)$ for $\ell \prec \ell_*$ $(\ell \succ \ell_*)$ that are sufficiently close to ℓ_*. The notation $\ell \prec \ell_*$ means that the direction of the vector ℓ can be obtained from the direction of the vector ℓ_* using the counterclockwise rotation by the angle not exceeding π. The roots from "–" to

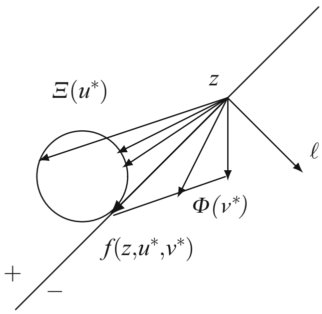


Fig. 22.4 Semipermeable direction

"+" and the roots from "+" to "−" are called the roots of the first and of the second type, respectively.

One can prove that, in the game considered, the equation $H(\ell, z) = 0$ has at least one root of the first type and one root of the second type. Moreover, it has two roots of the first type and two roots of the second type at most. We denote the roots of the first type by $\ell^{(1),i}(z)$ and the roots of the second type by $\ell^{(2),i}(z)$. One can find the domains of the functions $\ell^{(j),i}(\cdot)$, $j = 1, 2$, $i = 1, 2$. The form of these domains is shown in Fig. 22.5.

It can be proved that the function $\ell^{(j),i}(\cdot)$ satisfies the Lipschitz condition in any closed subset of its domain. So, we can consider the following differential equation:

$$\frac{dz}{dt} = \Pi \ell^{(j),i}(z), \tag{22.23}$$

where Π is the matrix of rotation by the angle $\pi/2$ (the rotation's direction depends on j). Since the tangent at each point of phase trajectories of this equation is a semipermeable direction, the trajectories are semipermeable curves. It means that the first player can keep one side of the curve (say, positive side) and the second player can keep another side (negative side). So, the equation (22.23) specifies a family $\Lambda^{(j),i}$ of semipermeable curves.

In Fig. 22.6, the families of semipermeable curves for dynamics (22.5) are presented. There are families $\Lambda^{(1),1}$ and $\Lambda^{(1),2}$ of the first type and families $\Lambda^{(2),1}$ and $\Lambda^{(2),2}$ of the second type. The second upper index in the notation $\Lambda^{(j),i}$ indicates those of two extremal values of control u that corresponds to this family: $i = 1$ is related to curves which are trajectories for $u = 1$; $i = 2$ is related to curves which are trajectories for $u = -1$. The arrows show the direction of motion in reverse time. Due to symmetry properties of the dynamics, all families can be obtained from only one of them (e.g., $\Lambda^{(1),1}$) by means of reflections about the horizontal and vertical axes.

The construction of mentioned four families of smooth semipermeable curves can be explained as follows.

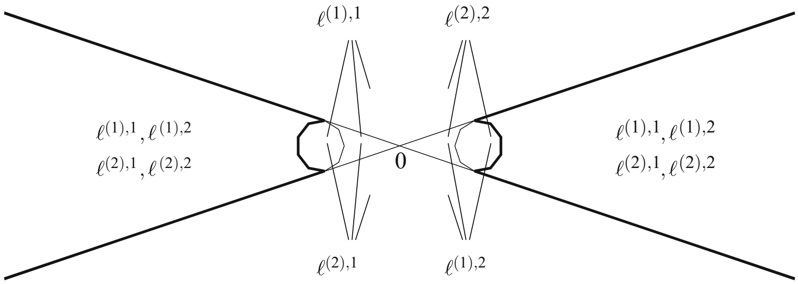


Fig. 22.5 Domains of $\ell^{(j),i}$

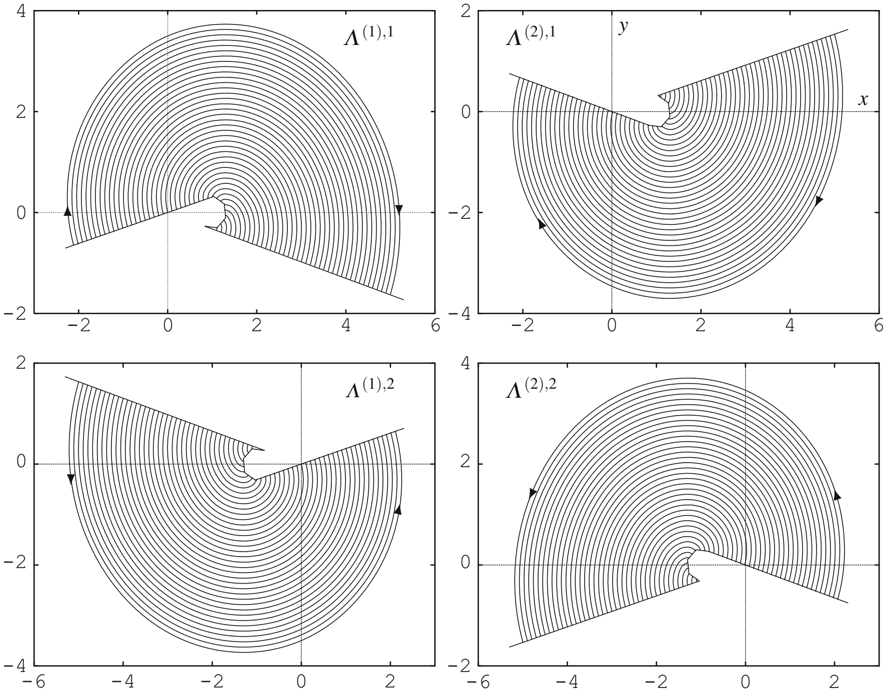


Fig. 22.6 Families of smooth semipermeable curves for the classical homicidal chauffeur dynamics

Assign the set

$$B_* = \{(x, y) : -y + v_x = 0,\ x - 1 + v_y = 0,\ v \in Q\}$$

to the control $u = 1$, and the set

$$A_* = \{(x, y) : y + v_x = 0,\ -x - 1 + v_y = 0,\ v \in Q\}$$

to the control $u = -1$. Hence, B_* is the set of all points in the plane x, y such that the velocity vector of system (22.5) vanishes at $u = 1$ and some $v \in Q$. We have $A_* = -B_*$.

Consider two tangents to the sets A_*, B_* passing through the origin (see Fig. 22.7), and mark arcs $a_1a_2a_3$ and $b_1b_2b_3$ on ∂A_* and ∂B_*, respectively.

Attach an inextensible string of a fixed length to the point b_1 and wind it up on the arc $b_1b_2b_3$. Then wind the string down keeping it taut in the clockwise direction. The end of the string traces an involute, which is a semipermeable curve of the family $\Lambda^{(1),1}$. The complete family $\Lambda^{(1),1}$ is obtained by changing the length of the string.

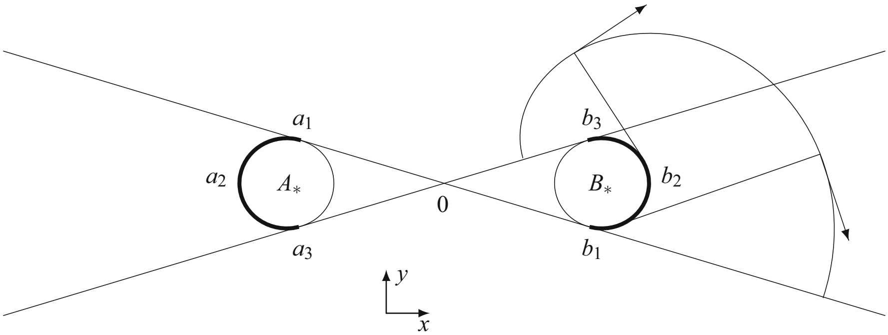


Fig. 22.7 Auxiliary arcs generating the families of smooth semipermeable curves for dynamics (22.5)

The family $\Lambda^{(2),2}$ is obtained as the collection of the counterclockwise involutes of the arc $a_1a_2a_3$ by attaching the string to the point a_3 .

The family $\Lambda^{(2),1}$ is generated by the clockwise involutes of the arc $b_1b_2b_3$ provided the string is attached to the point b_3.

The family $\Lambda^{(1),2}$ is composed of the counterclockwise involutes of the arc $a_1a_2a_3$ provided the string is attached to the point a_1.

The curves of different families belonging to the same type can be sewed in some cases so that the semipermeability property will be preserved. The procedure for computing the solvability set of the game of kind is based on the issuing two semipermeable curves (which are faced each to other with positive sides) from end points of M's usable part, on the analysis of their mutual disposition, and on the sewing semipermeable curves of different families belonging to the same type.

Families of semipermeable curves corresponding to dynamics (22.6) are arranged in a more complicated way (see Patsko and Turova 2009), and we do not present them here.

2.5 Classical Homicidal Chauffeur Problem

In the homicidal chauffeur problem described in the book by R. Isaacs, player P strives as soon as possible to bring the state vector of system (22.5) to a given closed bounded set M, whereas player E strives to prevent this.

2.5.1 Statement by R. Isaacs

Isaacs supposed that the terminal set M is a circle of radius r with the center at the origin. Thus, the description of the problem involves two independent parameters ν and r.

R. Isaacs investigated the problem for some parameter values using his method for solving differential games. The basis of the method is the backward computation

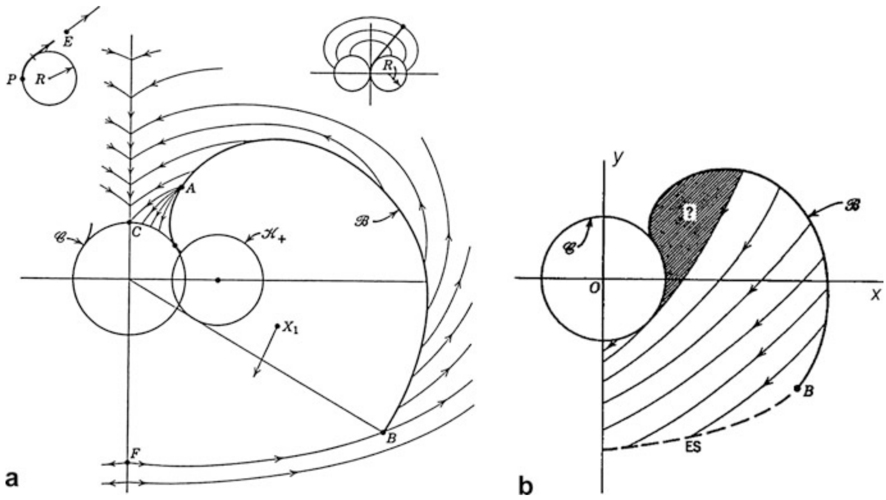


Fig. 22.8 Pictures by R. Isaacs from Isaacs (1965) explaining the solution to the homicidal chauffeur game

of characteristics for an appropriate partial differential equation. First, some primary region is filled out with regular characteristics, then secondary region is filled out, and so on. The final characteristics in the plane of state variables coincide with optimal trajectories.

Figure 22.8a shows a drawing from the book (Isaacs 1965) by R. Isaacs. The solution is symmetric with respect to the vertical axis. The upper part of the vertical axis is a singular line. Forward time-optimal trajectories meet this line at some angle and then go along it toward the target set M. According to the terminology by R. Isaacs, the line is called universal. The part of the vertical axis adjoining the target set from below is also a universal singular line. Optimal trajectories go down along it. The rest of the vertical axis below this universal part is dispersal: two optimal paths emanate from every point of it. On the barrier line $\mathcal{B}$, the value function is discontinuous. The side of the barrier line where the value of the game is smaller will be called positive. The opposite side is negative. One can see in Fig. 22.8a that the barrier line is a semipermeable curve of the first type. There is a similar line, but of the second type, in the left symmetric part.

The equivocal singular line ES emanates tangentially from the terminal point of the barrier (Fig. 22.8b). It separates two regular regions. Optimal trajectories that come to the equivocal curve split into two paths: the first one goes along the curve, and the second one leaves it and comes to the regular region on the right (optimal trajectories in this region are shown in Fig. 22.8a).

The equivocal curve is described through a differential equation which cannot be integrated explicitly. Therefore, any explicit description of the value function in the region between the equivocal and barrier lines is absent. The most difficult for the investigation is the "rear" part (Fig. 22.8b, shaded region) denoted by R. Isaacs with a question mark. He could not obtain a solution for this region.

The arising of singular lines (resp. singular surfaces in the case of higher dimension) that have no explicit description is a typical difficulty when investigating particular differential games. Having explicit formulas on initial stages of the backward construction of solution, we cannot come over a singular line and continue explicit description. Only qualitative investigation is possible. Because of that, development of numerical methods even for problems in the plane is necessary.

2.5.2 Brief Description of Numerical Algorithm for the Construction of Level Sets of the Value Function

In this subsection, we give a schematic description of the algorithm for the backward construction of level sets of the time-optimal value function. One can also say that, using the backward procedure, we construct the solvability sets of the game by a given time on some time grid. The equivalent term occurring in the literature is the backward (guaranteed) reachable set.

Primarily, the algorithm was developed (Patsko and Turova 1995, 1997) for linear time-optimal game problems in the plane. However, it turned out that the linearity of dynamics for time-optimal differential games does not give essential advantages in construction of level sets of the value function, since the level sets are often non-convex, which is typical for nonlinear dynamics. Therefore, for time-optimal games, nonlinear case is equivalent in difficulty to linear one. Hence, after some modernization, the algorithm have been used for nonlinear dynamics too (Patsko and Turova 2001).

The basic idea of the algorithm for approximate construction of the level sets $W_M(\tau) = \{(x, y) : V(x, y) \leq \tau\}$ of the value function V is explained in the following.

We replace the set M by its polygonal approximation. Similarly, the geometric constraint Q on the control v is substituted by a polygon. The set $W_M(\tau)$ is formed via step-by-step backward procedure giving a sequence of embedded sets:

$$W_M(\varDelta) \subset W_M(2\varDelta) \subset W_M(3\varDelta) \subset \ldots \subset W_M(i\varDelta) \subset \ldots \subset W_M(\tau). \qquad (22.24)$$

Here $\varDelta$ is the step of the backward procedure. Each set $W_M(i\varDelta)$ consists of all initial points such that the first player brings system (22.5) into the set $W_M((i-1)\varDelta)$ within the time duration $\varDelta$ (we put $W_M(0) = M$).

The crucial point of the algorithm is the computation of "fronts." The front F_i (Fig. 22.9) is the set of all points of $\partial W_M(i\varDelta)$ for which the minimum guaranteeing time of reaching $W_M((i-1)\varDelta)$ is equal to $\varDelta$. For other points of $\partial W_M(i\varDelta)$ the optimal time is less than $\varDelta$. The line $\partial W_M(i\varDelta) \setminus F_i$ possesses the properties of the barrier. The front F_i is constructed using the previous front F_{i-1}. For the first step of the backward procedure, F_0 coincides with the usable part $\varGamma_0$ of the boundary of M.

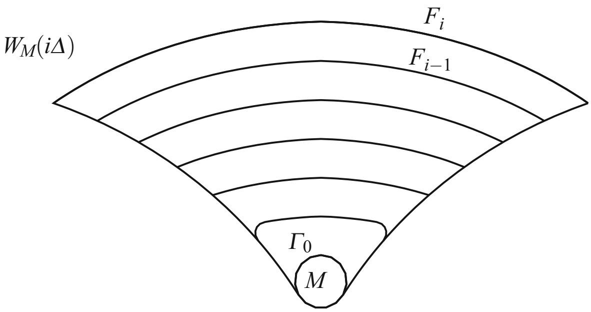


Fig. 22.9 Construction of the sets $W_M(i\Delta)$

Let us explain how the fronts can be constructed. Suppose the front F_{i-1} is a smooth curve. Let z_* be an arbitrary point of F_{i-1} and ℓ is the normal vector to the front at z_*. Let $u^\circ = \arg\min_{|u|\leq 1} \ell' p(z_*)u, \quad v^\circ = \arg\max_{v\in Q} \ell' v$.

We call u°, v° the extremal controls. The controls u° and v° are chosen from the conditions of minimizing and, respectively, maximizing the projection of the velocity vector of (22.5) onto the direction ℓ. If the vector pointed to z_* is collinear to ℓ, then any control $u \in [-1, 1]$ is extremal. If Q is a polygon in the plane, and ℓ is collinear to some normal vector to an edge $[q_1, q_2]$ of Q, then any control $q \in [q_1, q_2]$ is extremal.

After computing the extremal controls, the extremal trajectories of system (22.5) issued from the front's points backward in time are considered: $z(\tau) = z_* - \tau(p(z_*)u^\circ + v^\circ + g)$. The ends of these trajectories at $\tau = \Delta$ form the next front F_i. If the extremal control u° is not unique at some point $z_* \in F_{i-1}$, then the segment $\Phi(z_*) = \{ \bigcup_{u^\circ\in[-1,1]} (z_* - \Delta\,(p(z_*)u^\circ + v^\circ + g))\}$ is considered instead of the single point. Similarly, if the extremal control v° is not unique, the segment $\Xi(z_*) = \{ \bigcup_{v^\circ\in[q_1,q_2]} (z_* - \Delta\,(p(z_*)u^\circ + v^\circ + g))\}$ is considered.

For each front, we distinguish points of the local convexity and points of the local concavity. In Fig. 22.10, d is a point of the local convexity, and e is a point of the local concavity. If z_* is a point of the local convexity and the extremal control u° is not unique, we obtain a local picture like the one shown in Fig. 22.11a after issuing the extremal trajectories from the point z_*. Here, the additional segment $\Phi(z_*)$ appears on the new front F_i. If the extremal control v° is not unique, we obtain a local picture similar to the one shown in Fig. 22.11b: the "swallow tail" $\beta_1\xi\beta_2$ does not belong to the new front F_i and it is taken away. For points of the local concavity, there is an inverse situation: if the extremal control u° is not unique, a swallow tail that should be removed appears; if the extremal control v° is not unique, an additional segment $\Xi(z_*)$ appears on the new front F_i. If both u° and v° are nonunique, the insertion or the swallow tail arises depending on which of segments $\Phi(z_*)$ or $\Xi(z_*)$ is greater.

In the course of numerical computations, we operate with polygonal lines instead of smooth curves. Two normal vectors to the links $[a, b]$, $[b, c]$ of the polygonal

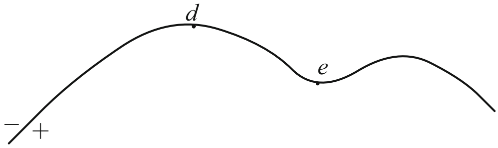


Fig. 22.10 Local convexity and concavity

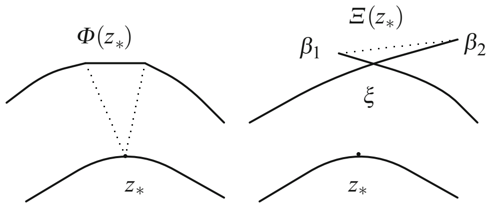


Fig. 22.11 Nonuniqueness of extremal controls in the case of local convexity

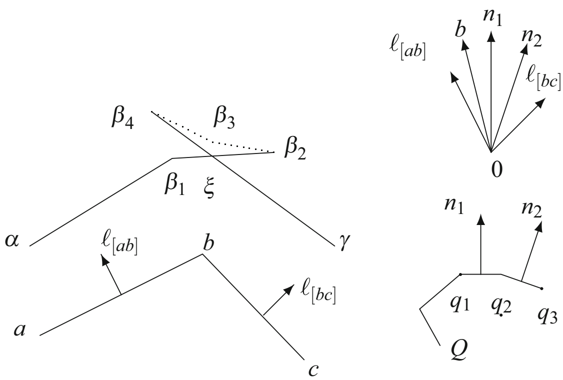


Fig. 22.12 Example of local constructions

line are considered at each vertex b (Fig. 22.12). The algorithm treats all possible variants of disposition of the vectors $\ell_{[ab]}, \ell_{[bc]}$, normals to the edges of Q, and the vector b. In Fig. 22.12, for instance, the case is shown where the vector b is between the vectors $\ell_{[ab]}, \ell_{[bc]}$, and the normals $n_1,\ n_2$ to the set Q are between the vectors b and $\ell_{[bc]}$. The extremal controls of the players are computed for each of these vectors, and the extremal trajectories are issued from the points a, b, c. The ends of these trajectories computed at $\tau = \Delta$ give a local picture shown in Fig. 22.12. In the case considered, four extremal trajectories were issued from the point b. Their ends are $\beta_1, \beta_2, \beta_3$, and β_4. The segment $[\beta_1, \beta_2]$ appears due to nonuniqueness of the extremal control u° for the vector b. The segments $[\beta_2, \beta_3]$ and $[\beta_3, \beta_4]$ arise due to nonuniqueness of the extremal control v° for the vectors n_1 and n_2. After removing the swallow tail $\xi\beta_4\beta_3\beta_2\xi$, the polygonal line $\alpha\beta_1\xi\gamma$ is obtained to be a fragment of the next front.

2.5.3 Two Examples of Numerical Solution of Classical Problem

Figure 22.13 shows level sets $W_M(\tau) = \{(x, y) : V(x, y) \le \tau\}$ of the value function $V(x, y)$ for $\nu = 0.3, r = 0.3$. The numerical results presented in Fig. 22.13

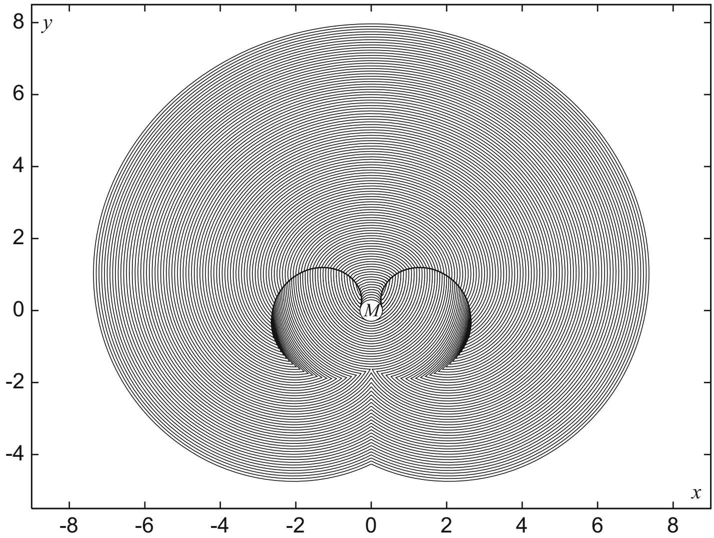


Fig. 22.13 Level sets of the value function for the classical problem; game parameters $\nu = 0.3$ and $r = 0.3$; backward computation is done till the time $\tau_f = 10.3$ with the time step $\Delta = 0.01$, output step for fronts $\delta = 0.1$

are obtained using the algorithm, which is described in the previous subsection. The lines on the boundary of the sets $W_M(\tau)$, $\tau > 0$, consisting of points (x, y) where the equality $V(x, y) = \tau$ holds, are fronts (isochrones). For the visualization of graphs of the value function in time-optimal differential games, a special computer program has been developed (Averbukh et al. 2000).

The computation for Fig. 22.13 is done with the time step $\Delta = 0.01$ till the time $\tau_f = 10.3$. The output step for fronts is $\delta = 0.1$. The set M is approximated by an inscribed regular 20-polygon and the set Q by a 24-polygon. Figure 22.14 presents the graph of the value function. The value function is discontinuous on the two barrier lines and on a part of the boundary of the target set. The barrier lines are arcs of semipermeable curves of the families $\Lambda^{(1),1}$ and $\Lambda^{(2),2}$. In the case considered, the value function is smooth in the abovementioned rear region.

If the center of the target circle is shifted from the y-axis, the symmetry of the solution with respect to y-axis is destroyed. The arising front structure to the negative side of barrier lines can be very complicated. One of such examples is presented in Fig. 22.15. The target circle of radius 0.075 is centered at the point with the coordinates $m_x = 1, m_y = 1.5$. The computation time step $\Delta = 0.01$. The maximal value of the game for computed fronts is 9.5, and it is attained at a point in the second quadrant. The fronts are depicted with the time step $\delta = 0.08$. Figure 22.16 presents the graph of the value function.

The real situation corresponding to a shifted target set M may be the following. Assume that the pursuer is able to create a small circular killing zone in some

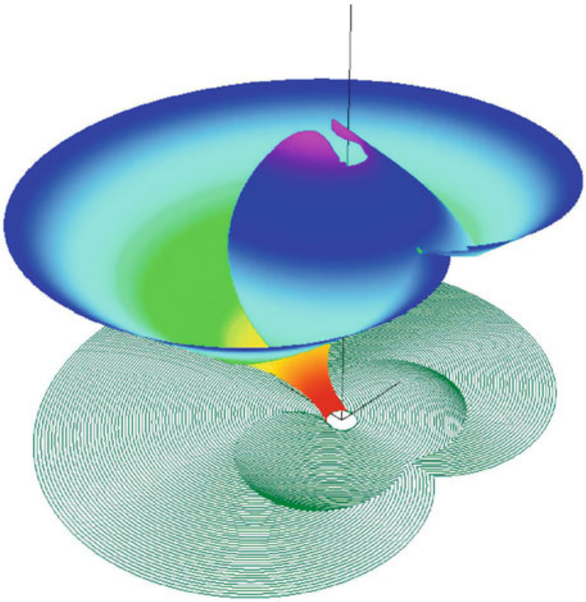


Fig. 22.14 Graph of the value function; $\nu = 0.3$, $r = 0.3$

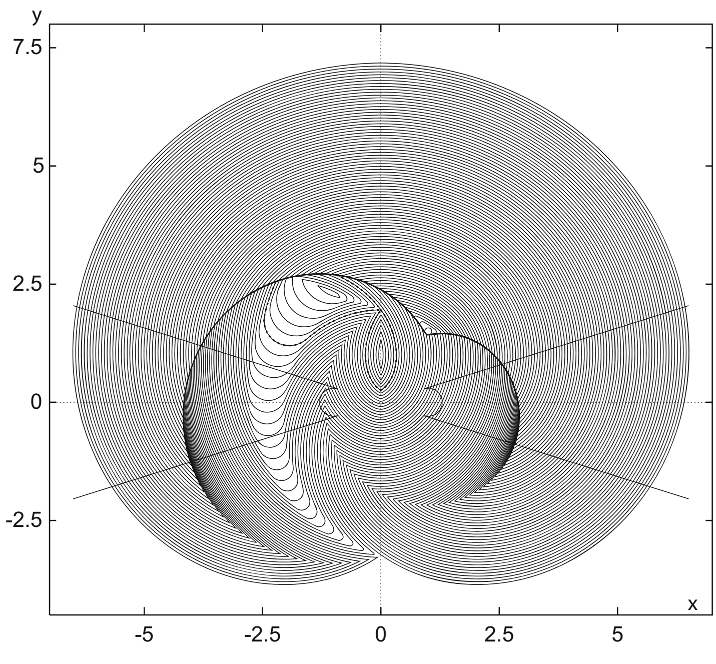


Fig. 22.15 Nontrivial structure of fronts for shifted target circle; $\nu = 0.3$, $\tau_f = 9.5$, $\Delta = 0.01$, $\delta = 0.08$. Target set is a circle of radius $r = 0.075$ centered at point $(1, 1.5)$

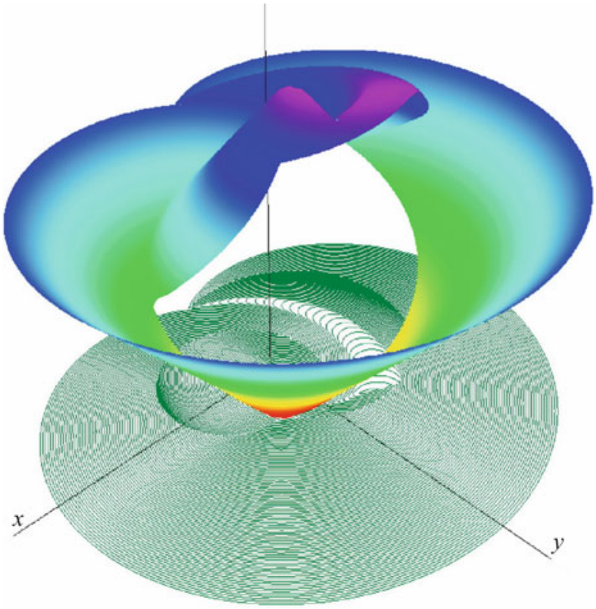


Fig. 22.16 Graph of the value function for shifted target circle; $\nu = 0.3$, $r = 0.075$

distance and at some angle to the velocity vector direction in front of himself. Then the pursuer minimizes the time of capture of the evader in such killing zone.

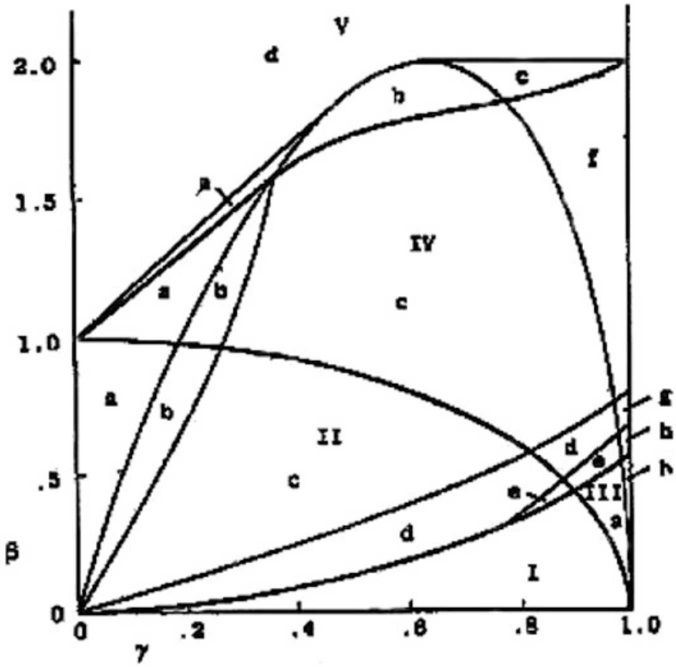
2.5.4 Investigations by J. Breakwell and A. Merz

J. Breakwell and A. Merz continued the investigation of the homicidal chauffeur game in the setting by R. Isaacs. Their results are partly and very briefly described in the papers (Breakwell and Merz 1969; Merz 1974). A complete solution is obtained by A. Merz in his PhD thesis (1971) at Stanford University.

A. Merz divided the two-dimensional parameter space into 20 subregions. He investigated the qualitative structure of the optimal paths and the type of singular lines for every subregion. All types of singular curves (dispersal, universal, equivocal, and switch lines) described by R. Isaacs for differential games in the plane appear in the homicidal chauffeur game for certain values of parameters. In his thesis, A. Merz suggested to distinguish some new types of singular lines and consider them separately. Namely, he introduced the notion of focal singular lines which are universal ones but with tangential approach of optimal paths. The value function is non-differentiable on the focal lines.

Figure 22.17 presents a picture and a table from the thesis by A. Merz that demonstrate the partition of two-dimensional parameter space ν, r into subregions with certain system of singular lines (A. Merz used symbols γ, β for the notation of parameters. He called singular lines as exceptional lines).

The thesis contains many pictures explaining the type of singular lines and the structure of optimal paths. By studying them, one can easily detect tendencies in the behavior of the solution depending on the change of the parameters.



DISTRIBUTION OF EXCEPTIONAL LINES

Region	Sub-region	Exceptional Lines Present in Each Subregion											
		S	UL	PDL_y	PDL	EDL_a	EDL_c	ID^+	ID^-	EL	SL	SE	FL
I		x	x										
II	a	x	x	x					x		x		
	b	x	x	x					x	x	x		
	c	x	x	x					x	x			
	d	x	x	x					x	x		x	
	e	x	x	x					x	x		x	x
III	a	x	x			x		x					
	b	x	x				x	x					
IV	a		x	x	x			x	x		x		
	b		x	x	x			x	x	x	x		
	c		x	x	x	x		x	x	x			
	d		x	x	x	x		x	x	x		x	
	e		x	x	x	x		x	x	x		x	x
	f		x	x	x		x	x	x	x			
	g		x	x	x		x	x	x	x		x	
	h		x	x	x		x	x	x	x		x	x
V	a		x	x				x					
	b		x	x		x		x					
	c		x	x			x	x					
	d		x	x									

Fig. 22.17 Decomposition of two-dimensional parameter space into subregions

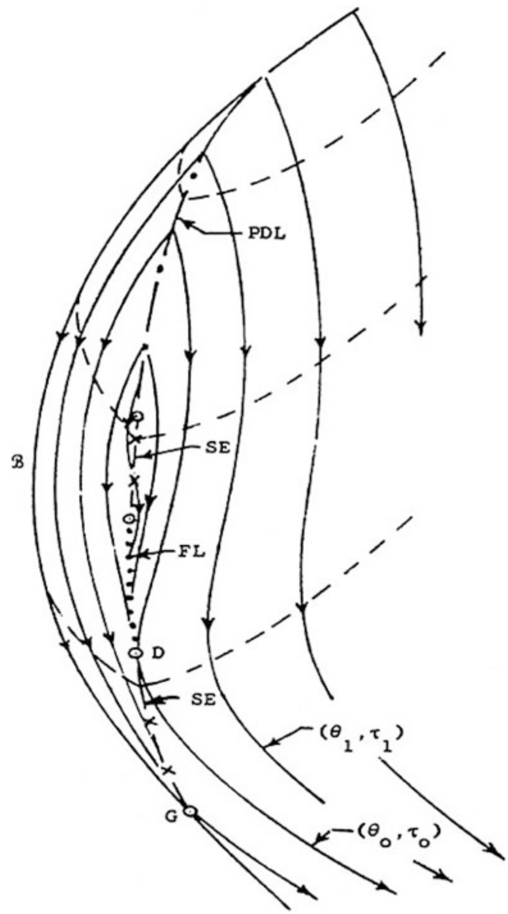


Fig. 22.18 Structure of optimal paths in the rear part for subregion IIe

In Fig. 22.18, the structure of optimal paths in that part of the plane that adjoins the negative side of the barrier is shown for the parameters corresponding to subregion IIe. This is the rear part denoted by R. Isaacs with a question mark. For subregion IIe, very complicated situation takes place.

PDL denotes the dispersal line controlled by player P. Two optimal trajectories emanate from every point of this line. Player P controls the choice of the side to which trajectories come down. Singular curve SE (the switch envelope) is specified as follows. Optimal trajectories approach it tangentially. Then one trajectory goes along this curve, and the other (equivalent) one leaves it at some angle. Therefore, line SE is similar to an equivocal singular line. The thesis contains arguments according to which the switch envelope should be better considered as an individual type of singular line.

FL denotes the focal line. The dotted curves mark boundaries of level sets (in other words, isochrones or fronts) of the value function.

The value function is not differentiable on the line composed of the curves PDL, SE, FL, and SE.

The authors of this chapter undertook many efforts to compute the value function for parameters from subregion IIe. But it was not successful, since we could not obtain corner points that must be present on fronts to the negative side of the barrier. One of the possible explanations to this failure can be the following: the effect is so subtle that it cannot be detected even for very fine discretizations. The computation of level sets of the value function for the subregions where the solution structure changes very rapidly, dependent on the parameters, can be considered as a challenge for differential game numerical methods being presently developed by different scientific teams.

Figure 22.19 demonstrates computation results for the case where fronts have corner points in the rear region. However, the values of parameters correspond not to subregion IIe but to subregion IId. For the latter case, singular curve SE remains, but focal line FL disappears.

For some subregions of parameters, barrier lines on which the value function is discontinuous disappear. A. Merz described a very interesting transformation of the barrier line into two close to each other dispersal curves of players P and E. In this case, there exist both optimal paths that go up and those that go down along

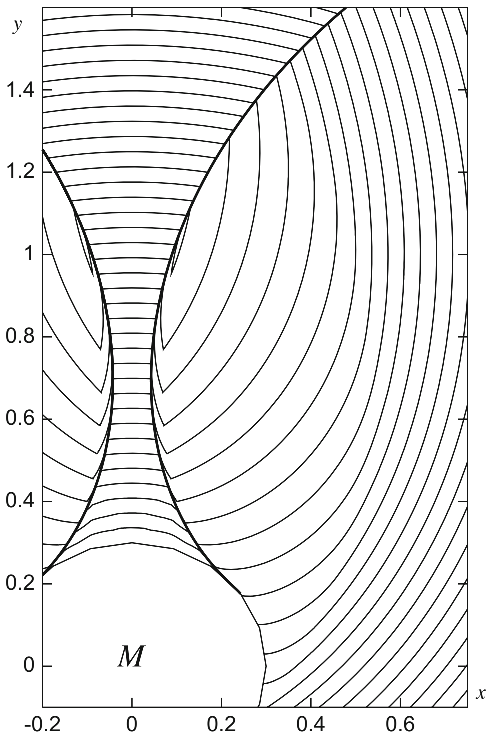


Fig. 22.19 Level sets of the value function for parameters from subregion IId; $\nu = 0.7$, $r = 0.3$; $\tau_f = 35.94$, $\Delta = 0.006$, $\delta = 0.12$

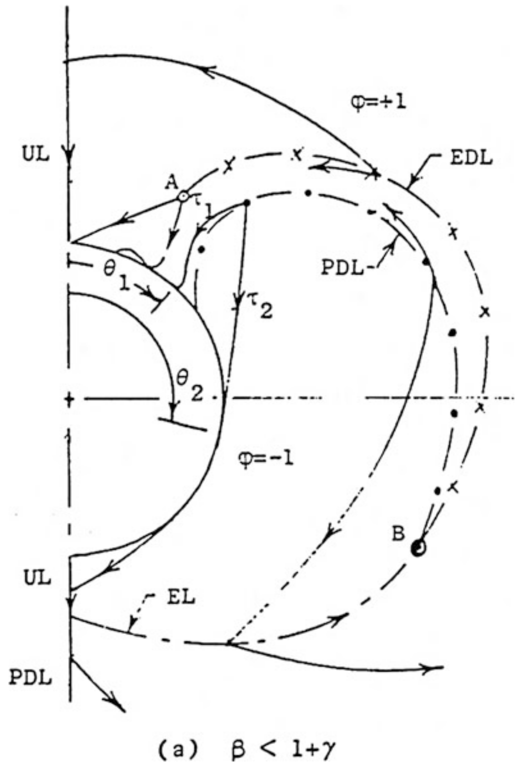


Fig. 22.20 Structure of optimal trajectories in subregion IVc

the boundary of the target set. The investigation of such a phenomenon is of great theoretical interest.

Figure 22.20 presents a picture from the thesis by A. Merz that corresponds to subregion IVc (A. Merz as well as R. Isaacs used the symbol φ for the notation of the control of player P. In this text, the corresponding notation is u). Numerically constructed level sets of the value function are shown in Fig. 22.21. When examining Fig. 22.21, it might seem that some barrier line exists. But this is not true. This again underlines the importance of theoretical investigation of particular differential games. Without the work completed by A. Merz, the presence of some barrier line in that place could erroneously be established based on the numerical outcome. Accounting for the results by A. Merz (obtained mainly analytically) enables refining our numerical constructions. Here, we have exactly the case like the one shown in Fig. 22.20. In Fig. 22.22, an enlarged fragment of Fig. 22.21 is given. The curve consisting of fronts' corner points above the accumulation region of fronts is the dispersal line of player E. The curve composed of corner points below the accumulation region is the dispersal line of player P. The value function is continuous in the accumulation region. To see where (in the considered part of the plane) the point of a maximal value of the game is located, additional fronts are

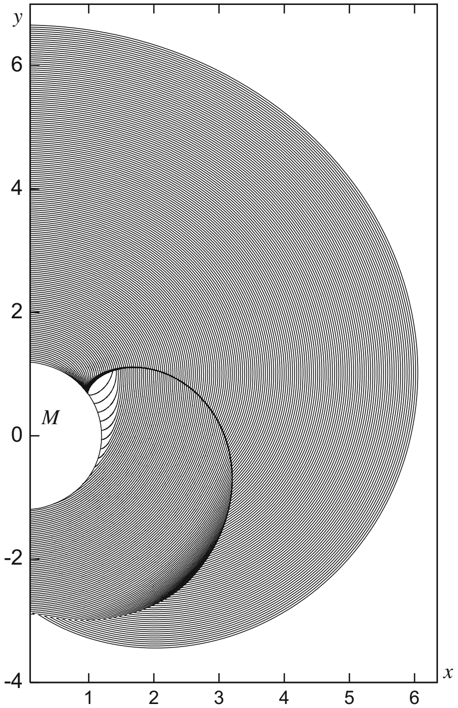


Fig. 22.21 Level sets of the value function; $\nu = 0.7$, $r = 1.2$; $\tau_f = 24.22$, $\Delta = 0.005$, $\delta = 0.1$

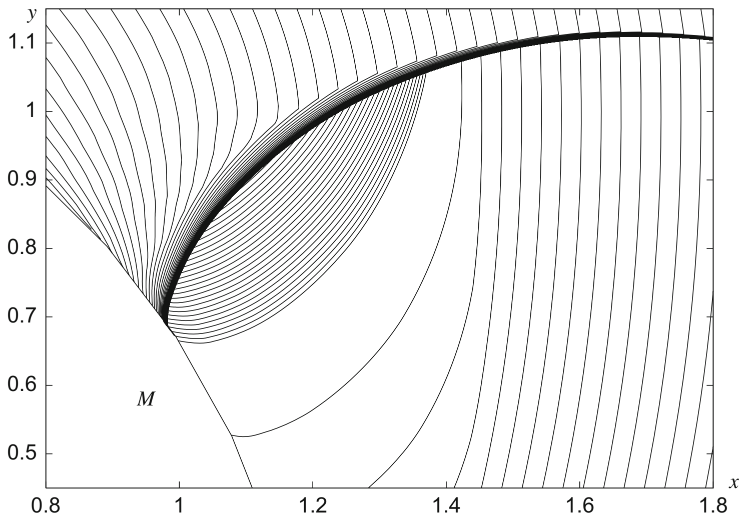


Fig. 22.22 Enlarged fragment of Fig. 22.21; $\tau_f = 24.22$. Output step for fronts close to the time τ_f is decreased up to $\delta = 0.005$

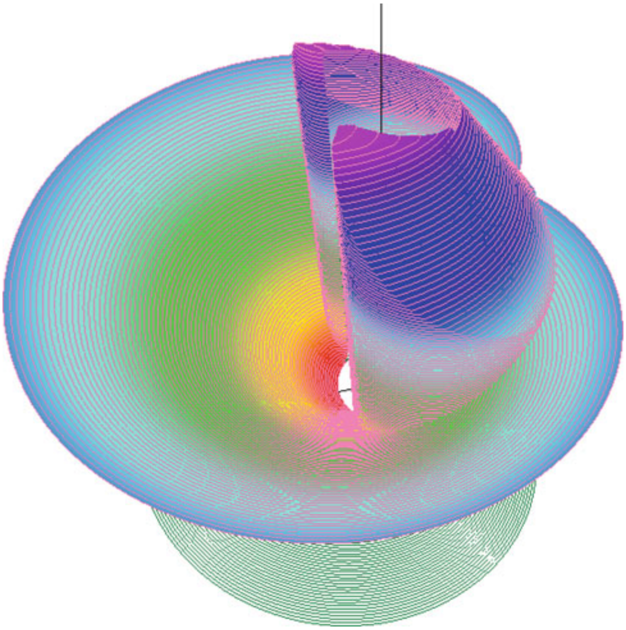


Fig. 22.23 Graph of the value function; $\nu = 0.7$, $r = 1.2$. Level lines are plotted. Salient curve corresponding to the line PDL from Fig. 22.20 is seen

shown. The point of the maximal value has coordinates $x = 1.1$, $y = 0.92$. The value function at this point is equal to 24.22.

The graph of the value function for the example considered is shown in Fig. 22.23. The level lines are plotted to make visible two curves consisting of salient points. Taking into account the symmetry with respect to the y-axis, the curves correspond to the dispersal singular lines EDL and PDL in the plane x, y (see Fig. 22.20).

2.6 Surveillance-Evasion Game

In the PhD thesis by J. Lewin (1973) (performed as well under the supervision of J. Breakwell), in the joint paper by J. Breakwell and J. Lewin (1975), and also in the paper by J. Lewin and G.-J. Olsder (1979), both dynamics and constraints on the controls of the players are the same as in Isaacs' setting but the objectives of the players differ from those in the classic statement. Namely, player E tries to decrease the time of reaching the target set M by the state vector, whereas player P strives to increase that time. In the first and second works, the target set is the complement (with respect to the plane) of an open circle centered at the origin. In the third publication, the target set is the complement of an open cone with the apex at the origin.

The meaning related to the original context concerning two moving objects is the following: player E tries, as soon as possible, to escape from some detection zone attached to the geometric position of player P, whereas player P strives to keep his opponent in the detection zone as long as possible. Such a problem was called the surveillance-evasion game. To solve it, J. Breakwell, J. Lewin, and G.-J. Olsder used Isaacs' method.

Below, level sets of the value function computed for the game with conic terminal set with the backward procedure from Sect. 2.5.2 are presented. Herewith, the extremal controls of the pursuer P and the evader E are determined via the relations $u^\circ = \operatorname{argmax}\{\ell' p(z_*)u : |u| \le 1\}$ and $v^\circ = \operatorname{argmin}\{\ell' v : v \in Q\}$ for every point z_* of local convexity and outer normal ℓ to the front at z_*. For the points of local concavity, the extremal controls of P and E are defined by the formulae $u^\circ = \operatorname{argmin}\{\ell' p(z_*)u : |u| \le 1\}$ and $v^\circ = \operatorname{argmax}\{\ell' v : v \in Q\}$, where ℓ is an inner normal to the front at z_*. So, the local constructions described earlier for the points of local convexity are now true for the points of local concavity and vice versa. In the results presented, the circle constraint of radius $\nu = 0.588$ on the control v of player E is substituted by an inscribed regular hexagon.

In the surveillance-evasion game with the conic target set M (the detection zone is the cone $R^2 \backslash M$ of semi-angle α), examples of transition from finite values of the game to infinite values are of interest and can be easily constructed.

Figure 22.24 shows level sets of the value function for five values of parameter $\alpha = 143°, 136.3°, 130°, 125.6°, 121°$. Since the solution to the problem is symmetric with respect to y-axis, only the right half-plane is shown for four of five figures. The pictures are ordered from greater to smaller α.

In the first picture, the value function is finite in the set that adjoins the target cone and is bounded by the curve $a'b'cba$. This set is filled out with the fronts (isochrones). The value function is zero within the target set. Outside the union of the target set and the set filled out with the fronts, the value function is infinite.

In the third picture, a situation of the accumulation of fronts is presented. Here, the value function is infinite on the line fe and finite on the arc ea. The value function has a finite discontinuity on the arc be.

The second picture demonstrates a transition case from the first to the third picture.

In the fifth picture, the fronts propagate slowly to the right and fill out (outside the target set) the right half-plane as the backward time τ goes to infinity. Figure 22.25 gives a graph of the value function for this case.

The fourth picture shows a transition case between the third and fifth pictures.

Note that all lines on which the value function is discontinuous (barrier lines) are arcs of families of semipermeable curves described in Sect. 2.4.

2.7 Acoustic Game

Let us return to problems where player P minimizes and player E maximizes the time of reaching the target set M. In papers (Cardaliaguet et al. 1995, 1999),

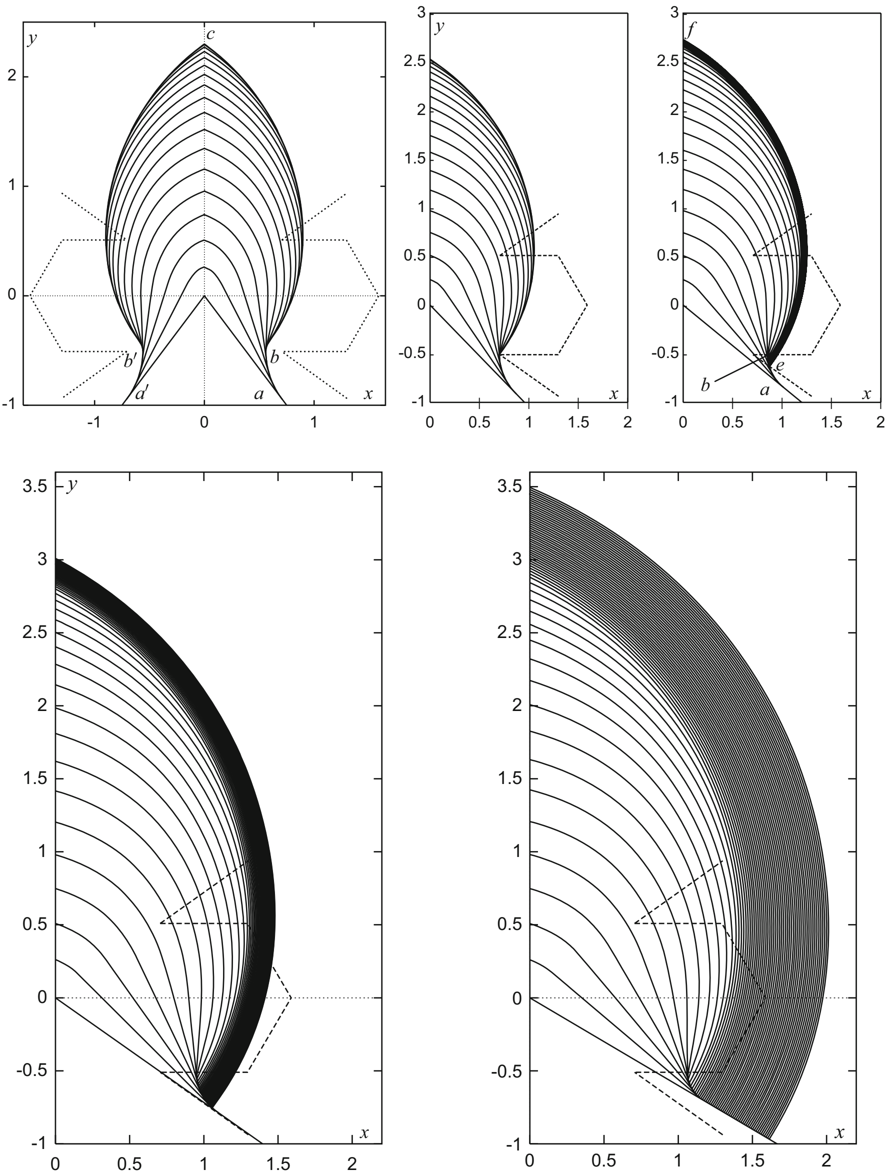


Fig. 22.24 Surveillance-evasion game. Change of the front structure depending on the semi-angle α of the nonconvex detection cone; $\nu = 0.588$, $\Delta = 0.017$, $\delta = 0.17$

P. Cardaliaguet, M. Quincampoix, and P. Saint-Pierre have considered an "acoustic" variant of the homicidal chauffeur problem. It is supposed that the constraint ν on the control of player E depends on the state (x, y). Namely,

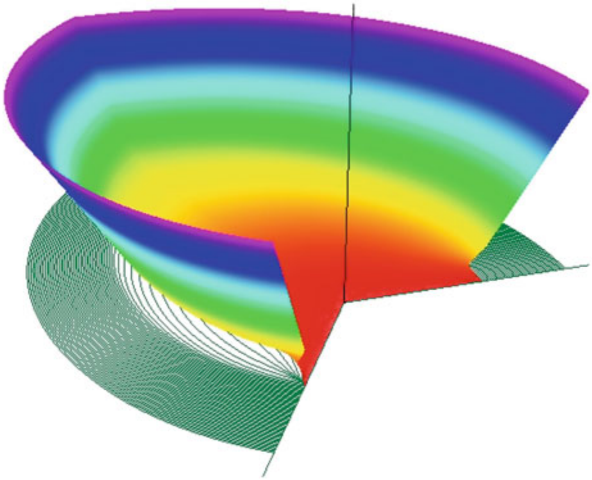


Fig. 22.25 Value function in the surveillance-evasion game: $\nu = 0.588$, $\alpha = 121°$

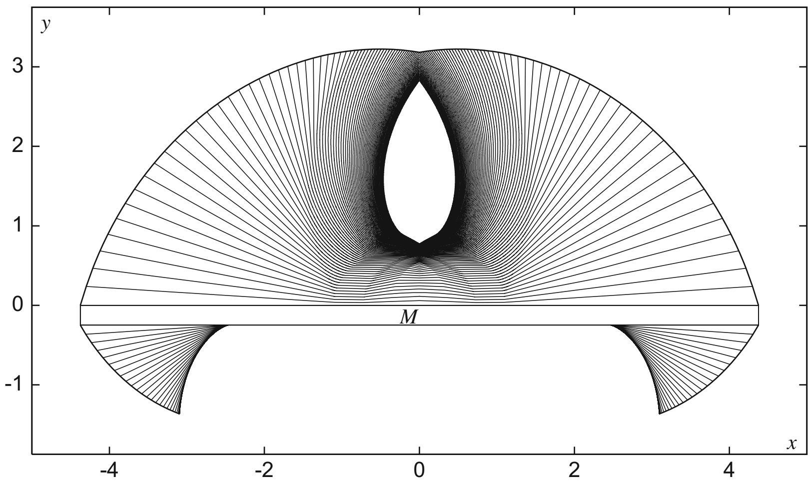


Fig. 22.26 Level sets of the value function in the acoustic problem; $\nu^* = 1.5$, $s = 0.9375$; $\Delta = 0.00625$, $\delta = 0.0625$

$$\nu(x, y) = \nu^* \min\left\{1, \sqrt{x^2 + y^2}/s\right\}, \; s > 0.$$

Here, ν^* and s are the parameters of the problem.

The applied aspect of the acoustic game: object E should not be very loud if the distance between him and object P becomes less than a given value s. Such an applied aspect and its interpretation were suggested by P. Bernhard. Let us cite here work (Bernhard and Larrouturou 1989): "Emitted noise is a function of evader's speed, while perceived noise is also a function of the distance between the opponents."

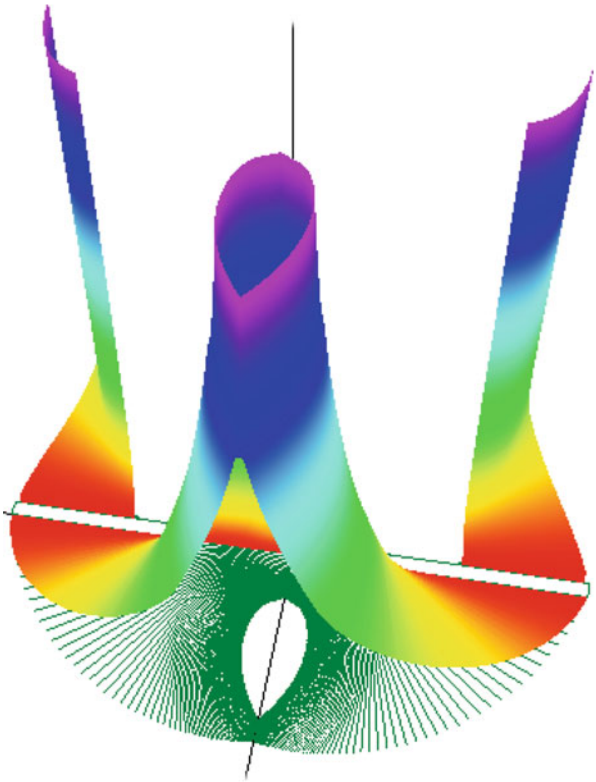


Fig. 22.27 Graph of the value function in the acoustic problem; $\nu^* = 1.5$, $s = 0.9375$

P. Cardaliaguet, M. Quincampoix, and P. Saint-Pierre investigated the acoustic problem using their method for numerical solving of differential games that was briefly described in Sect. 2.3 (item 7). It was revealed that one can choose the values of the parameters in such a way that the set of states where the value function is finite will contain a hole in which points the value function is infinite. Especially easy such a case can be obtained when the target set is a rectangle stretched along the horizontal axis.

Figures 22.26 and 22.27 demonstrate an example of the acoustic problem with the hole. The level sets of the value function and the graph of the value function are shown. The value of the game is infinite outside the set filled out with the fronts.

Let us underline that the abovementioned hole is separated from the target set. In Fig. 22.28, level sets for the parameters $\nu^* = 1.4$, $s = 2.5$ are presented. The graph of the value function is shown in Fig. 22.29. Also here, a hole with infinite magnitudes of the value function arises. But this hole touches the target set, which allows one to compute it easily through the barrier lines emanated from some points on the boundary of the target set.

The acoustic homicidal chauffeur problem is carefully investigated in Patsko and Turova (2001, 2004). These works also contain findings on families of

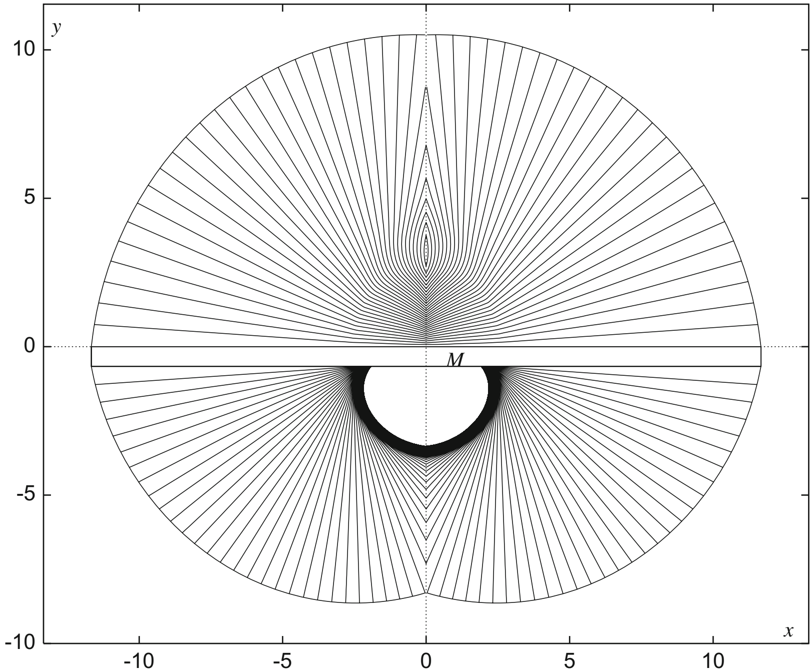
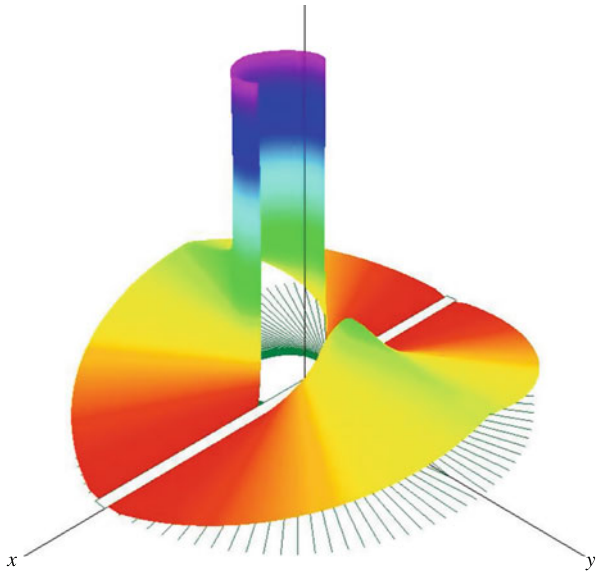


Fig. 22.28 Level sets of the value function in the acoustic problem; $\nu^* = 1.4$, $s = 2.5$; $\Delta = 1/30$, $\delta = 1/6$

Fig. 22.29 Graph of the value function in the acoustic problem; $\nu^* = 1.4$, $s = 2.5$



semipermeable curves of the first and second types arising in this problem under various values of parameters ν^*, s. With the help of semipermeable curves, the appearance of the hole with infinite value of the game (see Figs. 22.26 and 22.27) is explained.

2.8 Game with a More Agile Player *P*

Consider the homicidal chauffeur problem, in which player P controls the car that can change his linear velocity instantaneously. Here, we use dynamics equations (22.6). Accordingly, numerical procedures of Sect. 2.5.2 for the computation of level sets of the value function become more complicated. In more detail, the problem with a more agile player P is investigated in Patsko and Turova (2009).

2.8.1 Level Sets of the Value Function

In Fig. 22.30, the level sets of the value function which correspond to one and the same time $\tau = 3$ but to different values of the parameter a from -1 to 1 are presented. For all computations, the radius of the target set is $r = 0.3$ and the constraint on the control of player E is $\nu = 0.3$. In case $a = -1$, player P controls a Reeds-Shepp's car, and the obtained level set is symmetric with respect to both y-axis and x-axis. If $a = 1$, the level set for the classical homicidal chauffeur game is obtained.

Figure 22.31 shows the level sets of the value function for $a = -0.1$, $\nu = 0.3$, $r = 0.3$. The computation is done backward in time till $\tau_f = 4.89$. Precisely this value of the game corresponds to the last outer front and to the last inner front adjoining to the lower part of the boundary of the target circle M. The front structure is well seen in Fig. 22.32 showing an enlarged fragment of Fig. 22.31. One can see a nontrivial character of changing the fronts near the lower border of the accumulation region. The value function is discontinuous on the arc dhc. It is also discontinuous outside M on two short barrier lines emanating tangentially from the boundary of M. The right barrier is denoted by ce.

2.8.2 Optimal Strategies

When solving time-optimal differential games of the homicidal chauffeur type (with discontinuous value function), the most difficult task is the construction of the optimal (or ε-optimal) strategies of the players. Let us demonstrate such a construction using the last example.

We construct ε-optimal strategies using the extremal aiming procedure (Krasovskii 1985; Krasovskii and Subbotin 1988). The computed control remains unchanged during the next step of the discrete control scheme. The step of the control procedure is a modeling parameter. The strategy of player P (E) is defined using the extremal shift to the nearest point (extremal repulsion from the nearest point) of the corresponding front. If the trajectory comes to a prescribed layer

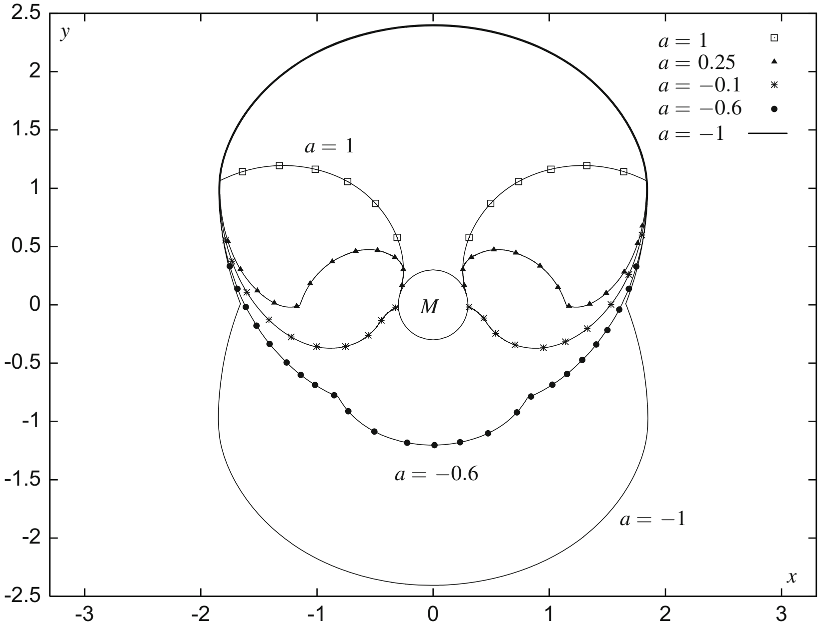


Fig. 22.30 Homicidal chauffeur game with more agile pursuer. Dependence of level sets of the value function on the parameter a for $\tau = 3$; $\nu = 0.3$, $r = 0.3$

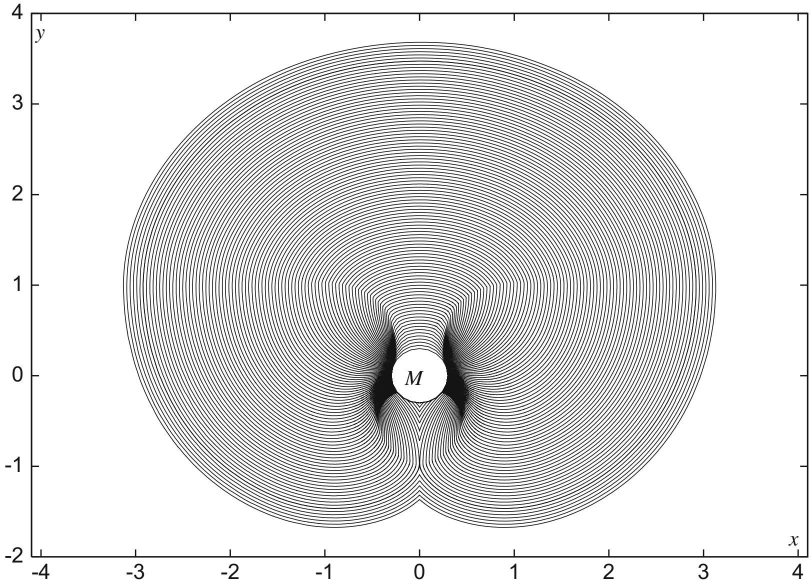


Fig. 22.31 Level sets of the value function in the homicidal chauffeur game with more agile pursuer; $a = -0.1$, $\nu = 0.3$, $r = 0.3$; $\tau_f = 4.89$, $\Delta = 0.002$, $\delta = 0.05$

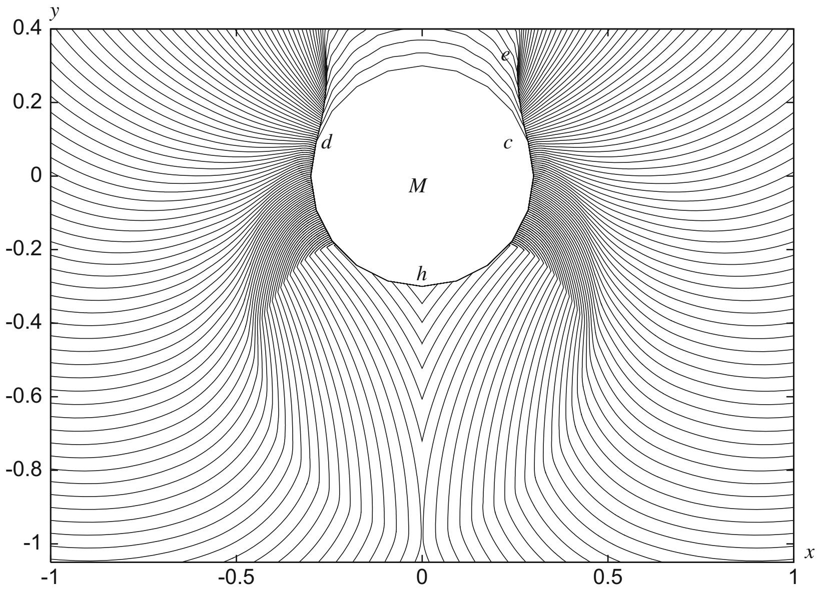


Fig. 22.32 Enlarged fragment of Fig. 22.31

attached to the positive (negative) side of the discontinuity line of the value function, then a control which pushes away from the discontinuity line is utilized.

Let us choose two initial points $a = (0.3, -0.4)$ and $b = (0.29, 0.1)$. The first point is located in the right half-plane below the front accumulation region, the second one is close to the barrier line on its negative side. The values of the game in the points a and b are $V(a) = 4.225$ and $V(b) = 1.918$, respectively.

In Fig. 22.33, the trajectories for ε-optimal strategies of the players are shown. The time step of the control procedure is 0.01. We obtain that the time of reaching the target set M is equal to 4.230 for the point a and 1.860 for the point b. Figure 22.33c demonstrates an enlarged fragment of the trajectory emanating from the initial point b. One can see a sliding mode along the negative side of the barrier.

Figure 22.34 presents trajectories for nonoptimal behavior of player E and optimal behavior of player P. The control of player E is computed using a random number generator (random choice of vertices of the polygon approximating the circle constraint of player E). The reaching time is 2.590 for the point a and 0.300 for the point b. One can see how the second trajectory penetrates the barrier line. In this case, the value of the game calculated along the trajectory drops jump-wise.

In Fig. 22.35, the trajectories for nonoptimal behavior of player P and optimal behavior of player E are shown. The control u of player P acts in optimal way, whereas the control w is nonoptimal. For Fig. 22.35a, $w \equiv 1$. The time of reaching the target set is 7.36. For Fig. 22.35b, c, $w \equiv -1$ until the trajectory comes to the vertical axis, after that $w \equiv 1$. Figure 22.35c demonstrates an enlarged fragment of

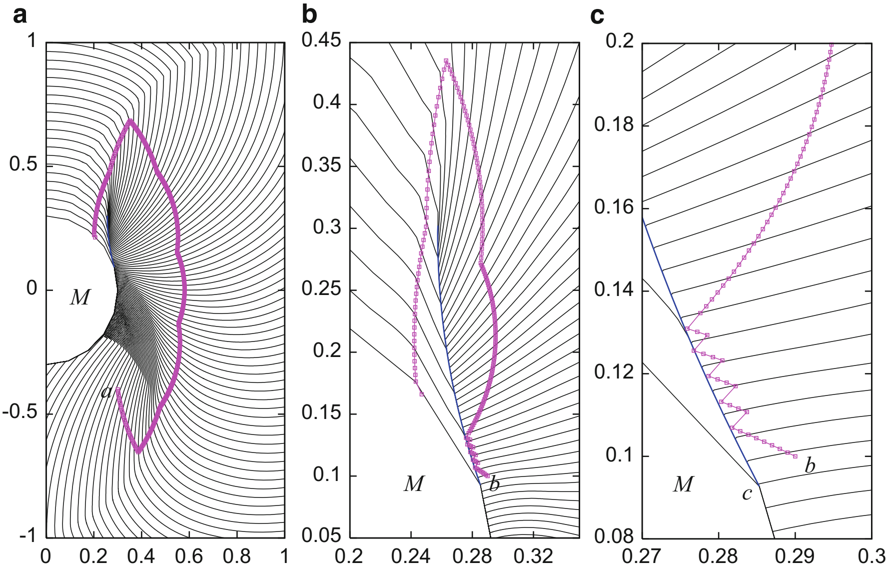


Fig. 22.33 Homicidal chauffeur game with more agile pursuer. Simulation results for optimal motions. (**a**) Initial point $a = (0.3, -0.4)$. (**b**) Initial point $b = (0.29, 0.1)$. (**c**) Enlarged fragment of the trajectory from the point b

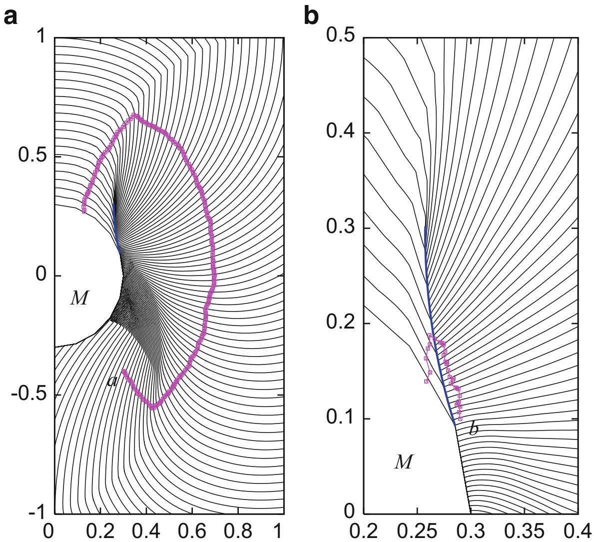


Fig. 22.34 Homicidal chauffeur game with more agile pursuer. Optimal behavior of player P and random action of player E. (**a**) Initial point $a = (0.3, -0.4)$. (**b**) Initial point $b = (0.29, 0.1)$

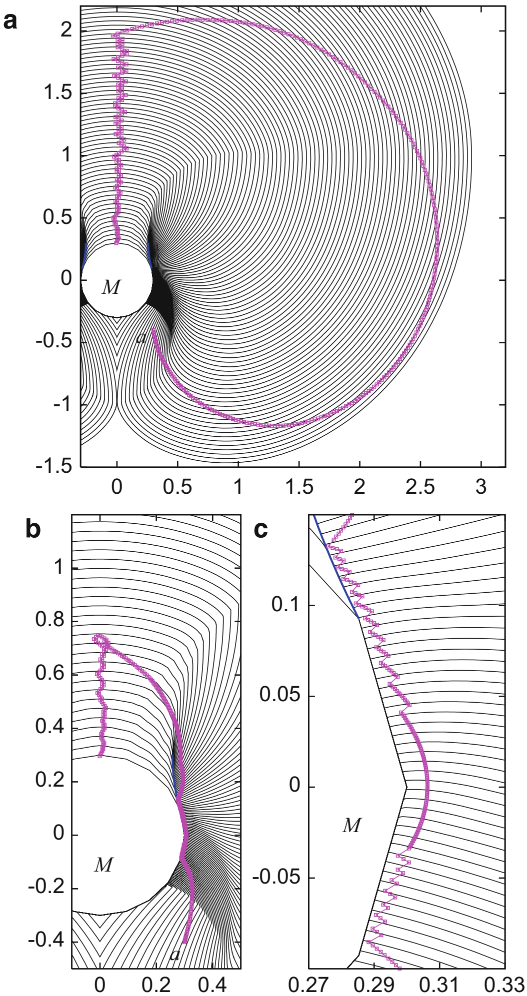


Fig. 22.35 Homicidal chauffeur game with more agile pursuer. Optimal behavior of player E and non-optimal control w of player P. Initial point $a = (0.3, -0.4)$. (**a**) $w \equiv 1$. (**b**) $w = -0.1$ until the trajectory comes to the vertical axis, after that $w = 1$. (**c**) Enlarged fragment of the trajectory on the left

the trajectory from Fig. 22.35b. The trajectory goes very close to the terminal set. The reaching time is 5.06.

2.9 Homicidal Chauffeur Game as a Test Example and a Stimulus for Investigation of New Comprehensive Problems

Presently, numerical methods and algorithms for solving zero-sum differential games are intensively developed. Often, the homicidal chauffeur game is used as a test or demonstration example (see, e.g., Botkin et al. 2011, 2013; Dvurechenskii 2013; Dvurechenskii and Ivanov 2012; Meyer et al. 2005; Mikhalev and Ushakov 2007; Mitchell 2002; Raivio and Ehtamo 2000).

In the reference coordinate frame, the game is of the second order in phase variables. Therefore, one can apply both general algorithms and algorithms taking into account the specifics of the plane. The nontriviality of the dynamics is in that the control u enters the right hand side of the two-dimensional control system as a factor by the state variables, and the constraint on the control v can depend on the phase state. Moreover, the control of player P can be two-dimensional, as it is in the modification discussed in Sect. 2.8.

Speaking about applied problems whose investigation was motivated by time-optimal differential games and, in particular, by the homicidal chauffeur problem, let us note the following:

1. There exist publications (see, e.g., Bakolas and Tsiotras 2012, 2013) related to the minimization of the traveling time of an object in the presence of drift field. In idealized settings, the spatial and velocity characteristics of such field are supposed to be known. A more realistic approach assumes the presence of uncertainties in the representation of drift field. The velocity field can be generated by the wind if the object is moving in the air, or by the undertow if movement of some deep-sea vehicle is considered.

2. Very important are problems in which a controlled object (an aircraft or a ship) should avoid collision with some other moving object. If our information on the movement of the second object is incomplete, we are again in the scope of differential game theory methods. For example, in the paper Exarchos et al. (2015), an object with car-like dynamics should avoid from a given circular neighborhood of an object with dynamics of simple motion. Here, at least locally, on some interval of possible collision, it is appropriate to consider the "homicidal pedestrian" problem.

3. Let a third object known as a "defender" join to the pursuer and the evader. Suppose that an aircraft performing its task is attacked by a missile. At the same time, the second missile begins to defend the aircraft. Hence, on a small time interval, an interaction between the evader (the aircraft), the pursuer (the first missile), and the defender (the second missile) takes place. The evader and the defender can share information on the current position of all objects, completely or in part. Conceptual formulation of such problems and some approaches to their analytic analysis are considered in Shaferman and Shima (2010), Shima (2011), and Pachter et al. (2014). It should be noted that in such problems, it is even not completely clear if the problem can be formulated as a zero-sum differential game and whether the value function of this game (being defined by equating guaranteed result of the evader and defender with the guaranteed result of the pursuer) exists. Moreover, it is desirable to account for the state constraints imposed on the movement of every of the three objects. Here, a new, very interesting direction of research arises. Of course, accurate solving of such problems is impossible without appropriate numerical methods. Issues related to the formulation of such problems and to numerical solution methods are discussed in Fisac and Sastry (2015).

4. In the book Blaquière et al. (1969), two-player differential games with two target sets are considered. Each of the players strives to steer the control system to his own target set prior to his opponent. The applied interpretation of such problems

can be a combat between two aircraft or between two ships (Getz and Pachter 1981). Who is the pursuer and who is the evader, if each of the combatants should hit his opponent? In the context of objects like those ones in the homicidal chauffeur game, the problem is considered in Davidovitz and Shinar (1989), Merz (1985), and Olsder and Breakwell (1974).

5. In addition to the homicidal chauffeur game, the book of R. Isaacs contains, among basic time-optimal problems, the game of "two cars" and the "isotropic rocket" game. The meaning of the first game is clear from its name. In the second game, the pursuer with the dynamics of a material point in the plane in the presence of a drag force (which depends on the magnitude of velocity) pursuers the evader with the dynamics of simple motion. Using some transformation of coordinates, both problems can be reduced to differential games with a three-dimensional state vector. Among publications on the game of two cars, let us mention the works Simakova (1967, 1968), Merz (1972), and Pachter and Getz (1980) and the works Bernhard (1970, 1979) on the isotropic rocket game. Time-optimal problem for a material point (isotropic rocket) in a resistant medium with a partially given drift field is investigated in Selvakumar and Bakolas (2015). However, these problems are not yet completely solved (including dependence of solution on parameters). In this connection, let us stress the necessity of invoking careful numerical modeling. For an example of application of numerical methods to the game of two cars, we refer to the paper Mitchell et al. (2005), in which, for some set of parameters, the game of two cars is used as a fragment of collision-avoidance problem of two aircraft.

6. It should be clearly understood that such problems as the "homicidal chauffeur game," the "game of two cars," and the "isotropic rockets" are problems which reflect, at the model level, typical features of applied problems. One should be careful when considering the opportunity to use solutions of model mathematical problems in real practical applications. We refer to Bolonkin and Murphey (2005) as an example of rational approach to solving the problem of avoiding encounter. In this work, the counteraction of two objects with car-like dynamics is considered. For given initial states, the question being asked is: Can the second object avoid an encounter with the first one? On some time grid, the reachable sets of each object are constructed. Using them, it is analyzed if there exists a motion of the second object, which, at every time grid point, is located outside the corresponding reachable set of the first object. By constructing such a motion, we wish to bring it to the position, starting from which the further avoidance of the second object from the first one becomes evident. If this is possible, the avoidance maneuver is found. Thus, the conclusion on the construction (in approximate, engineering sense) of an avoidance maneuver using an open-loop control is drawn merely on the base of computing reachable sets on a time grid. Fast construction of reachable sets for objects with car-like dynamics is well realizable on current onboard computers. If the problem with one pursuer and several evaders is considered and we are interested in the behavior of the pursuer, then, after the above described analysis, only those evaders for whom the avoidance maneuver is impossible will be left. The pursuer's control law that guarantees the capture of the evader is designed by choosing a single evader from the remaining ones. However, the realization of subtle mathematical results

of the differential games theory related to the construction of optimal feedback controls that guarantee the interception of the second object by the first one is hardly possible at this time. So, we should agree with simplifications dropping the attempt to compute optimal solutions.

7. Mathematical results on the problem, in which the moving car should, as soon as possible, intercept more than one moving object are published in Berdyshev (2002, 2008) and being included into the book Berdyshev (2015). Similar problems but in the game formulation are considered in Shevchenko (2012). As an application, the problem with false targets, in which an attacking vehicle that does not distinct between the true and false targets should visit several moving targets and hit each of them can be mentioned. Very important are also the problems, in which, on the contrary, several moving objects with car-like dynamics (e.g., unmanned vehicles) perform coordinated pursuit of a single moving target (see Shaferman and Shima 2008).

8. Consider one more variant of a pursuit-evasion problem with a false target. There are a pursuer and an evader. Dynamic abilities of the pursuer exceed the ones of the evader. At the initial instant, the pursuer knows only the area where the evader is located. Using this information, the pursuer begins the search. At some instant, the evader creates a false target, and the pursuer encounters with the problem of two targets: the ones is true and another is false. The pursuer's detection means work effectively only when the distance between the pursuer and the observed object is not larger than a given value. Moreover, it is possible to identify the target (true or false) only when the pursuer comes within a given small distance from the target. Generally, the pursuer seeks to minimize the guaranteed capture time of the true target. Some such problems for objects with the dynamics of simple motion in the plane and in three-dimensional space were formulated and investigated in Abramyants et al. (1981, 2007), Shevchenko (1997), Zheleznov et al. (1993). The books Kim (1993) and Petrosyan and Garnaev (1993) are devoted to search problems under uncertainty conditions.

9. Extremely hard are pursuit-evasion problems with two or several groups of interacting objects with car-like (aircraft) dynamics. Here, various problem settings are possible. For instance, which objects can share information on their states and so on. For an example of mathematical works related to this direction, see Stipanović et al. (2010, 2012). Numerical algorithms for such problems are considered in Chen et al. (2015).

10. One of the popular methods in the mathematical control theory is approximation of complex sets by ellipsoids (Chernous'ko 1993; Kurzhanski and Valyi 1997). In Kurzhanski (2015, 2016), ellipsoids whose orientation and size are changed with time are used to describe virtual containers enclosing a group of moving objects. These small objects (little ships) should not collide, whereas the container having them inside should solve his task of transfer from some initial position to a given position, passing through "straits" and rounding "islands." The presence of uncertainties transforms this problem to a very sophisticated pursuit-evasion game.

11. When designing wheel robots, one of the main criteria for the choice of robot dynamics is the following. Assume that the robot should track some

prescribed trajectory in the plane in the presence of disturbances. It turned out that dynamics of simplest car (Dubins' car) are poorly suited for such a problem. The reason is that, for such dynamics, for small time intervals, the reachable set (in geometric coordinates) at fixed time does not contain the initial point. In other words, the reachable set of Dubins' car at fixed final time does not coincide with his time-limited reachable set. For Reeds-Shepp's vehicle, this property is fulfilled. Therefore, it is better to have the dynamics of the robot to be similar to the dynamics of Reeds-Shepp's car. This example shows why comprehensive mathematical control theory and, in, particular, its branch related to time-optimal games finds application in robotics (Laumond 1998; Laumond et al. 2014).

12. Among many contemporary applied problems, in which researchers are trying to apply methods of time-optimal games, it is worthy to mention those ones related to a jamming attack on the communication network of a team of moving objects. These problems are investigated by T. Başar and his collaborators (Bhattacharya and Başar 2012; Han et al. 2011). The simplest task is formulated in the following way. Two aircrafts (maybe, unmanned air vehicles) move in the horizontal plane. Communication requirements demand that the distance between them had to be not larger than a given value. At the initial instant, this condition is fulfilled. After that, the third object, a jammer aircraft, appears. This one tries to jam the connection between the mentioned pair of aircraft. To do this, the jammer uses its special equipment. The time-optimal zero-sum game is formulated as follows: the aerial jammer moves in a way to maximize the connection interruption time, whereas the team of the first and second aircraft tries to minimize this time by changing configuration of their motions.

3 Linear Differential Games with Fixed Termination Instant

Let the dynamics of a controlled object be described by a vector differential equation

$$\dot{z} = A(t)z + B(t)u + C(t)v, \quad u \in P \subset R^p, \ v \in Q \subset R^q. \tag{22.25}$$

Here, $z \in R^n$ is the phase vector, $A(t)$ is a square matrix $n \times n$, u is the vector control of the first player constrained by a compact set P, and v is the vector control of the second player constrained by a compact set Q. The matrices $B(t)$, $C(t)$ have appropriate sizes. All matrix functions are assumed to be continuous in time.

Let the termination instant t_f of the process be fixed and given. A scalar continuous function φ of the terminal payoff is defined. It can depend not on the entire phase vector z, but only on some m its components. The vector consisting of these components is denoted by z_m. The magnitude $\varphi\big(z_m(t_f)\big)$ at the terminal point of a system trajectory is minimized by the first player and maximized by the second one.

Thus, we have a differential game with linear dynamics, fixed termination instant and terminal payoff function. Such game has a continuous value function $(t, z) \mapsto \mathbf{V}(t, z)$.

3.1 Passage to Dynamics Without Phase Variable in the Right-Hand Side

Let $Z(t_f, t)$ be the fundamental Cauchy matrix corresponding to the matrix $A(t)$ participating in system (22.25). The symbol $Z_m(t_f, t)$ denotes a submatrix of the matrix $Z(t_f, t)$ composed of the m rows that correspond to the components of the phase vector z, which define the payoff function φ. The position of system (22.25) computed to the terminal instant t_f from the current instant t and current position $z(t)$ under zero players' controls $u \equiv 0$, $v \equiv 0$ is defined by the formula $Z(t_f, t)z(t)$. Respectively, the forecast value of the chosen m components is computed as $Z_m(t_f, t)z(t)$.

Introduce a new variable $y(t) = Z_m(t_f, t)z(t)$. Then

$$\dot{y}(t) = \frac{d}{dt}\big(Z_m(t_f, t)z(t)\big) = Z_m(t_f, t)B(t)u + Z_m(t_f, t)C(t)v.$$

Denote

$$D(t) = Z_m(t_f, t)B(t), \quad E(t) = Z_m(t_f, t)C(t)$$

and consider a game with the following dynamics

$$\dot{y}(t) = D(t)u + E(t)v, \quad u \in P, \quad v \in Q, \tag{22.26}$$

fixed termination instant t_f and the terminal payoff function φ. If in some interval $[t_*, t_f]$, the players' controls $t \mapsto u(t)$ and $t \mapsto v(t)$ in systems (22.25) and (22.26) are the same and $y(t_*) = Z_m(t_f, t_*)z(t_*)$, then $y(t_f) = z_m(t_f)$. This fact is the basis for the proof of equivalence of games (22.25) and (22.26): the magnitudes of the value function $\mathbf{V}(t, z)$ in game (22.25) and the value function $V(t, y)$ in game (22.26) are connected by relation

$$\mathbf{V}(t, z) = V\big(t, Z_m(t_f, t)z\big).$$

The benefits of dynamics (22.26) are the following:

(1) In the right-hand side of dynamics (22.26), there is no phase variable; this permits to solve problem (22.26) by simpler numerical methods;
(2) The dimension of the phase vector of system (22.26) equals m; if $m < n$, then it can simplify constructions. Often, in applied problems, one has $m = 2$, therefore, $y \in R^2$, and one should make constructions in the plane.

Replacement of dynamics (22.25) by dynamics (22.26) in linear differential games with fixed termination instant is used actively in mathematical and engineering literature since the end of the 1960s (Bryson and Ho 1975; Krasovskii 1970, 1971).

3.2 Lebesgue Sets of Value Function

Let $M_c = \{y : \varphi(y) \leqslant c\}$ be a Lebesgue set (a level set) of the payoff function that corresponds to the number c. Denote $W_c = \{(t, y) : t \leq t_f,\ V(t, y) \leq c\}$. The set W_c is the Lebesgue set of the value function that corresponds to the number c. Let $W_c(t) = \{y : V(t, y) \leq c\}$, $t \leq t_f$, be a time section (a t-section) of the tube W_c at the instant t. A remarkable property of the tube is that W_c is the set maximal by inclusion in the space $t \times y$, which possesses the stability property (weak-invariant set): for any initial position $(t_*, y_*) \in W_c$ for any constant control $t \mapsto v(t) \equiv v^* \in Q$ of the second player, there is a measurable control $t \mapsto u(t)$ of the first player that has its values in P and such that the trajectory $t \mapsto y(t; t_*, y_*, u(\cdot), v^*)$ of system (22.26) at any instant $t \in [t_*, t_f]$ stays in the set $W_c(t)$. A set that possesses such a property is called also u-stable bridge. Thus, W_c is the maximal stable bridge that stops at the set M_c.

Here, we must involve measurable controls of the first player (not piecewise-continuous) because only measurable controls provide closure of the trajectory bundle. From ideological point of view, one can imagine piecewise-continuous controls that are easier to understand.

Discriminating the second player (since he shows his further control in some future time interval), the first player can guide system (22.26) through the section $W_c(t)$, and at the instant t_f, the system is delivered to the set M_c. If to reject the discrimination of the second player, then the positional strategy of the first player extremal to the set W_c (Krasovskii 1985; Krasovskii and Subbotin 1974, 1988) and applied in a discrete control scheme with some sufficiently small time step keeps any trajectory of system (22.26) near the tube W_c up to the instant t_f.

The maximal stable bridge is (Krasovskii and Subbotin 1974, 1988) a closed set. If the set M_c is convex, then any t-section $W_c(t)$ is convex too (see Krasovskii and Subbotin 1988, p. 87) due to linearity of system (22.26).

The closure of the set $\{(t, y) : t \leq t_f\} \setminus W_c$ has analogous property but under discrimination of the first player and is called the maximal v-stable bridge.

Applying analytic or numerical procedures for constructing sets W_c in some grid of values of the parameter c, one can obtain a collection of exact or approximate Lebesgue sets of the value function of game (22.26). On the basis of such a collection, optimal positional controls of the players can be constructed approximately: being at some position (t, y), we choose the first player's control $u(t, y)$ that maximally pulls the system to the closest smaller set $W_c(t)$ and the second player's control $v(t, y)$ that maximally pulls the system from this set.

3.3 Backward Procedure for Constructing Sets $W_c(t)$ in the Convex Case

Now, let us describe the method for constructing level sets of the value function (maximal stable bridges) of games of the type (22.26) in the case of continuous quasiconvex payoff φ. (A function is called quasiconvex if all its Lebesgue sets are convex.) The procedure is of the backward type and can be treated as the dynamic programming principle applied to differential games.

To do the numerical construction, let us take a sequence of instants $t_0 < t_1 < t_2 < \ldots < t_{N-1} < t_N = t_f$ in the time interval $[t_0, t_f]$ of the game. Uniformity of the grid is unessential. For a given constant c, the result of the procedure is a collection of sets, each corresponding to a chosen time instant t_i and approximating the time section $W_c(t_i)$ of the level set $W_c = \{(t, y) : V(t, y) \le c\}$ of the value function V of the game (22.26) at this instant. The symbol $\widetilde{W}_c(t_i)$ will denote the set approximating the original time section $W_c(t_i)$.

Change the dynamics of the game (22.26) by a piecewise-constant dynamics:

$$\dot{y} = \widetilde{D}(t)u + \widetilde{E}(t)v, \quad \widetilde{D}(t) = D(t_i), \ \widetilde{E}(t) = E(t_i), \ t \in [t_i, t_{i+1}). \tag{22.27}$$

Instead of the original constraints P and Q for the controls of the players, let us consider their polyhedral approximations $\widetilde{P}$ and $\widetilde{Q}$. Let $\widetilde{\varphi}$ be the approximating payoff function. It is defined so that for any number c, its level set $\widetilde{M}_c = \{y : \widetilde{\varphi}(y) \le c\}$ is a convex polyhedron close in Hausdorff metrics to the level set M_c of the original payoff function.

The approximating game (22.27) is chosen such that in each step $[t_i, t_{i+1}]$ of the backward procedure we deal with a game with simple motion and polyhedral convex constraints for the players' controls. At the first step $[t_{N-1}, t_N] = [t_{N-1}, t_f]$, a solvability set $\widetilde{W}_c(t_{N-1})$ for a game of homing with target set $\widetilde{W}_c(t_N) = \widetilde{M}_c$ is constructed. Here, the first player tries to guide the system to the set $\widetilde{W}_c(t_N)$ at the instant t_N, and the second one opposes this. Continuing in the same way, a set $\widetilde{W}_c(t_{N-2})$ is constructed on the base of $\widetilde{W}_c(t_{N-1})$, and so on. As a result, we obtain a collection of convex polyhedra $\widetilde{W}_c(t_i)$ approximating (Botkin 1982; Polovinkin et al. 2001; Ponomarev and Rozov 1978) the sections $W_c(t_i)$ of the original level set W_c of the value function of the game (22.26) in Hausdorff metrics.

The procedure of moving from the section $\widetilde{W}_c(t_{i+1})$ to the next one $\widetilde{W}_c(t_i)$ is described in terms of support functions of the sets under consideration. Recall that the value $\rho(l, A)$ of the support function of a bounded closed set A on the vector l is calculated by formula

$$\rho(l, A) = \max_{a \in A} \langle l, a \rangle.$$

Here, the symbol $\langle \cdot, \cdot \rangle$ denotes the dot product of two vectors.

Introduce denotations $\mathscr{P}(t_i) = D(t_i)\widetilde{P}$, $\mathscr{Q}(t_i) = E(t_i)\widetilde{Q}$ for vectograms of the players (i.e., for the sets of velocities, which the players can give to the system). One knows (Pontryagin 1967b; Pschenichnyi and Sagaidak 1970) that in the convex case the set $\widetilde{W}_c(t_i)$ is represented as

$$\widetilde{W}_c(t_i) = \Big(\widetilde{W}_c(t_{i+1}) + (t_{i+1} - t_i) \cdot \big(-\mathscr{P}(t_i)\big)\Big) \overset{*}{-} (t_{i+1} - t_i) \cdot \mathscr{Q}(t_i).$$

Here, the symbol "+" denotes the Minkowski sum (the algebraic sum), and the symbol "$\mathbin{\underline{*}}$" denotes the Minkowski difference (the geometric difference). These operations are defined as

$$\mathscr{A} + \mathscr{B} = \{a + b : a \in \mathscr{A},\, b \in \mathscr{B}\}, \quad \mathscr{A} \mathbin{\underline{*}} \mathscr{B} = \bigcap_{b \in \mathscr{B}} (\mathscr{A} - b).$$

The support function of the Minkowski sum of two sets equals the sum of the support functions of these sets. For two convex compact sets, the support function of their Minkowski difference coincides with the convex hull of the difference of the support functions of these sets. Therefore, $\rho\big(\cdot, \widetilde{W}_c(t_i)\big) = \operatorname{co} \gamma(\cdot, t_i)$, where the value of the function $\gamma(\cdot, t_i)$ on a vector l is defined by the formula

$$\gamma(l, t_i) = \rho\big(l, \widetilde{W}_c(t_{i+1})\big) + (t_{i+1} - t_i) \cdot \rho\big(l, -\mathscr{P}(t_i)\big) - (t_{i+1} - t_i) \cdot \rho\big(l, \mathscr{Q}(t_i)\big). \tag{22.28}$$

The symbol "co" denotes the operation of taking the convex hull of a function.

The function $\gamma(\cdot, t_i)$ is positively homogeneous and piecewise-linear (because the support functions of the polyhedra $\widetilde{W}_c(t_{i+1})$, $-\mathscr{P}(t_i)$ and $\mathscr{Q}(t_i)$ are of this type). The property of local convexity can be violated only on the boundary of the linearity cones of the function $\rho\big(\cdot, \mathscr{Q}(t_i)\big)$, that is, on the boundary of the cones generated by normals to the neighboring faces of the polyhedron $\mathscr{Q}(t_i)$. This can be taken into account during construction of the convex hull of the function $\gamma(\cdot, t_i)$.

As a result of the backward procedure in the interval $[t_0, t_f]$, one obtains a collection of the sets $\widetilde{W}_c(t_i)$ for a value of the parameter c.

3.4 Constructing Sets $\widetilde{W}_c(t)$ in the Two-Dimensional Case

When the phase vector is two-dimensional, constructions described in the previous subsection can be realized very effectively. The procedure of convex hull construction is very fast because we have information about places of possible violation of local convexity. In Fig. 22.36, the structure of the function γ graph is shown schematically. Dash lines point out "corrections" of the function during convexification process. Herewith, this process stops after a few steps.

Below, the argument t_i of the function γ is omitted to simplify denotations.

Cones of linearity of the function γ are defined by the outer normals to edges of the convex polygons $\widetilde{W}_c(t_{i+1})$, $-\mathscr{P}(t_i)$, $\mathscr{Q}(t_i)$. Gathering these normals and ordering them clockwise, we obtain a collection L of vectors. The collection of magnitudes $\gamma(l)$ of the function γ (22.28) on the vectors $l \in L$ is denoted as Φ. The collections L, Φ define completely the function γ (with taking into account its positive homogeneity).

The set of outer normals to edges of the polygon $\mathscr{Q}(t_i)$ ordered clockwise is denoted S. The collection S is called set of "suspicious" vectors. This name is due to the fact that the function γ is certainly convex in any cone whose interior does

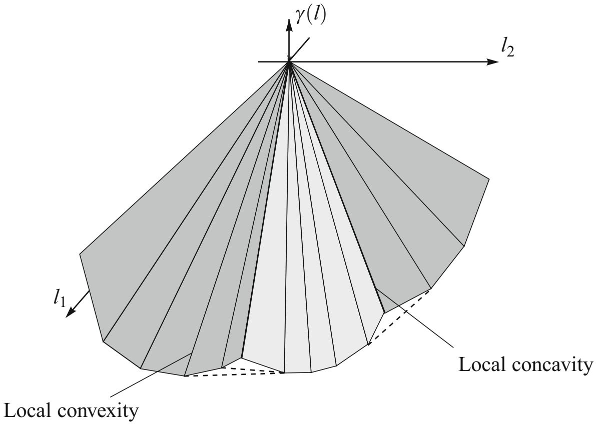


Fig. 22.36 A scheme of constructing convex hull of a positively homogeneous piecewise-linear function

not include any vector from the set S. And, respectively, violation of convexity can happen only in cones that do include at least one of such vectors.

Assume $L^{(1)} = L$, $\Phi^{(1)} = \Phi$, $S^{(1)} = S$. The $(k+1)$th step of the convexification process is made by replacing sets $L^{(k)}$, $\Phi^{(k)}$ by some other collections $L^{(k+1)} \subset L^{(k)}$, $\Phi^{(k+1)} \subset \Phi^{(k)}$. With that, $S^{(k)}$ is also changed by a new collection $S^{(k+1)}$.

Now, let us describe one step of the convexification procedure. Suppose that the angle between any two neighbor vectors from the collection $L^{(k)}$ counted clockwise is less then π. Let $l \mapsto \gamma^{(k)}(l)$ be a piecewise-linear positively homogeneous function described by the collections $L^{(k)}$, $\Phi^{(k)}$. Since

$$L^{(k)} \subset L^{(k-1)} \subset \cdots \subset L^{(1)}, \quad \Phi^{(k)} \subset \Phi^{(k-1)} \subset \cdots \subset \Phi^{(1)},$$

then for any vector $l \in L^{(k)}$ the magnitude $\gamma^{(k)}(l)$ equals $\gamma(l)$.

Take any vector $l_* \in S^{(k)}$ and two its neighboring in $L^{(k)}$ vectors l_- (taken counterclockwise from l_*) and l_+ (taken clockwise from l_*). Check whether the inequality $\langle l_*, y\rangle \leqslant \gamma^{(k)}(l_*)$ is essential in the triple of inequalities $\langle l_-, y\rangle \leqslant \gamma^{(k)}(l_-)$, $\langle l_*, y\rangle \leqslant \gamma^{(k)}(l_*)$, $\langle l_+, y\rangle \leqslant \gamma^{(k)}(l_+)$. The inequality $\langle l_*, y\rangle \leqslant \gamma^{(k)}(l_*)$ is essential, if for the set

$$A = \{y : \langle l_-, y\rangle \leqslant \gamma^{(k)}(l_-),\ \langle l_+, y\rangle \leqslant \gamma^{(k)}(l_+)\}$$

the following relation holds:

$$A \neq A \bigcap \{y : \langle l_*, y\rangle \leqslant \gamma^{(k)}(l_*)\}.$$

From essentiality of the inequality $\langle l_*, y \rangle \leqslant \gamma^{(k)}(l_*)$, it follows that the positively homogeneous function $\gamma^{(k)}$ is locally convex in the cone produced by the vectors l_-, l_*, l_+.

The algorithm for checking essentiality: find the point y_* of intersection of lines $\langle l_-, y \rangle = \gamma^{(k)}(l_-)$ and $\langle l_*, y \rangle = \gamma^{(k)}(l_*)$. Further, check the inequality $\langle l_+, y_* \rangle < \gamma^{(k)}(l_+)$. If it holds, then the middle inequality is essential. Therefore, the local convexity takes place.

If the local convexity holds, the vector l_* is excluded from the collection $S^{(k)}$. The new collection is denoted $S^{(k+1)}$. With that, $L^{(k+1)} = L^{(k)}$, $\Phi^{(k+1)} = \Phi^{(k)}$.

If the local convexity is violated, there are two situations. Let α be the angle between vectors l_- and l_+ counted clockwise. If

- $\alpha < \pi$, then the vector l_* is excluded from the collection $S^{(k)}$ and simultaneously the vectors l_- and l_+ are added there. (One or both can be in the collection $S^{(k)}$.) The collection obtained after these operations is the new set of "suspicious" vectors and is denoted by $S^{(k+1)}$. The new collection $L^{(k+1)}$ is obtained from $L^{(k)}$ by excluding the vector l_*. In the same, to get $\Phi^{(k+1)}$ from $\Phi^{(k)}$ we exclude the value $\gamma^{(k)}(l_*) = \gamma(l_*)$;
- $\alpha \geqslant \pi$, then either the considered triple of inequalities is inconsistent (the convex hull of the initial function does not exist) and, therefore, $\widetilde{W}_c(t_i) = \varnothing$. Or $\widetilde{W}_c(t_i)$ is a degenerated polygon, that is $\widetilde{W}_c(t_i)$ is a point or a segment. In both these cases, further constructions are ceased.

These are the actions made during one step of convex hull construction. The process is stopped at some step j, when $S^{(j)} = \varnothing$, that is, when there is no "suspicious" vectors. This means that the function $\gamma^{(j)}$ defined by the collections $L^{(j)}$ and $\Phi^{(j)}$ is locally convex on all vectors, that is, it is convex. Therefore, it is the convex hull of the initial function γ. Also stop can be caused by ceasing constructions; in this case, either $\widetilde{W}_c(t_i) = \varnothing$ or the polygon $\widetilde{W}_c(t_i)$ is degenerated.

The collection of sets $\widetilde{W}_c(t_i)$ is used by a visualization software to construct a solid tube to be drawn.

The algorithm for constructing t-sections $\widetilde{W}_c(t_i)$ is described in more details in Isakova et al. (1984) and Patsko (1996). The algorithm is very effective because before constructing the convex hull of a function γ we know places where its local convexity can be violated. Proof of convergence of the algorithm and some approximational schemes close to it are given in Botkin (1982) and Ponomarev and Rozov (1978). Convergence of analogous schemes is justified in Ponomarev and Rozov (1978). An algorithm for a posteriori estimating the numerical construction error is developed in Botkin and Zarkh (1984).

Let us give an example of numerical construction of maximal stable bridges for the following game:

$$\dot{z}_1 = z_2 + v, \quad \dot{z}_2 = -z_1 + u, \qquad t \in [0, 8], \\ |u| \leqslant 1, \quad |v| \leqslant 0.9, \quad \varphi(z_1, z_2) = z_1^2 + z_2^2. \tag{22.29}$$

In this case, $n = m = 2$. So, when the value function V is constructed in the coordinates y, we can pass back to the original coordinates z taking into account the relation $z = Z^{-1}(t_f, t)y$.

In Fig. 22.37, three Lebesgue sets of the value function (maximal stable bridges) are shown in the coordinates τ, z_1, z_2. They are computed in the interval [0, 8] of the backward time $\tau = t_f - t$ and correspond to the values $c = 1.05, 1.4, 2.7$ of the payoff function. In Fig. 22.37b, two external bridges are cut off by a plane parallel to the axes z_1, τ. In Fig. 22.37a, all three are cut off.

The break of the internal bridge is sharp. But for other examples, there are magnitudes of the payoff such that the corresponding Lebesgue set of the value function has a degenerated t-section $W_c(\bar{\tau})$ at some instant $\bar{\tau}$ (e.g., the section has no interior). Further, for $\tau > \bar{\tau}$, the sections grow and have interior. Below, tubes of this type are called *critical*. At the beginning of numerical studies of linear differential games, it seemed that critical tubes can appear in rare model examples only, but not in practical problems. But further, it have turned out that such a point of view is incorrect. J. Shinar found that critical tubes and connected to them narrow throats are quite typical for problems of space interception.

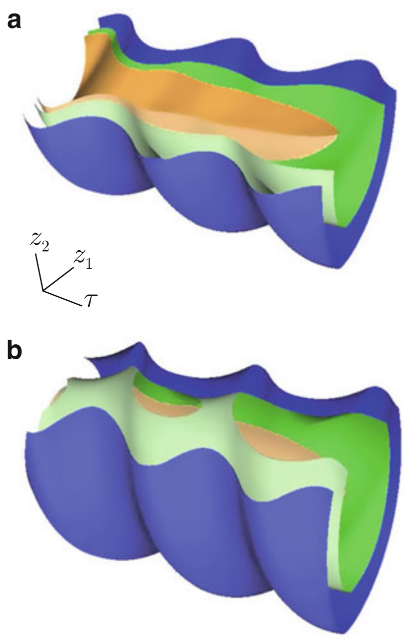


Fig. 22.37 Three maximal stable bridges for game (22.29)

3.5 J. Shinar's Problem of Space Interception

3.5.1 Problem Formulation

In the works Shinar et al. (1984), Shinar and Zarkh (1996), and Melikyan and Shinar (2000), a three-dimensional air-to-air interception problems has been formulated as a pursuit-evasion game by J. Shinar. The study by J. Shinar was based on earlier works Gutman and Leitmann (1975, 1976), Gutman (1979), and Shinar and Gutman (1980).

The pursuer is the interceptor missile, while the evader is a maneuverable aerial target (an aircraft or another missile). The natural payoff function of the game is the distance of closest approach, the miss distance, to be minimized by the pursuer and maximized by the evader. For the sake of simplicity, point mass models with velocities of constant magnitudes V_P, V_E were selected. The lateral accelerations of both players, normal to the respective velocity vectors, have constant bounds a_P and a_E. The evader controls its maneuver with ideal dynamics, while the pursuer's maneuver dynamics is represented by a first-order transfer function with the time constant τ_P.

In Fig. 22.38, the origin of the Cartesian coordinate system is collocated with the pursuer. The direction of the X-axis is along the initial line of sight. The XY plane is the nominal "collision plane" determined by the initial velocity vector of the evader $(V_E)_0$ and the initial line of sight. The Z-axis completes a right-handed coordinate system.

It is assumed the initial conditions are near to a "collision course," defined by

$$V_P \sin(\chi_P)_{\text{col}} = V_E \sin(\chi_E)_0, \tag{22.30}$$

and the actual velocity vector $V_P(t)$ of the pursuer remains close to the collision requirement $(V_P)_{\text{col}}$, satisfying

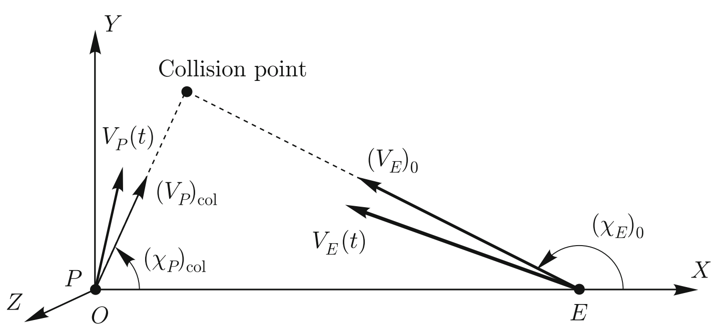


Fig. 22.38 The system of coordinates in the problem of three-dimensional pursuit. The actual vectors of $V_P(t)$ and $V_E(t)$ differ only slightly during the engagement from the nominal vectors $(V_P)_{\text{col}}$ and $(V_E)_0$, respectively

$$\sin\Big(\chi_P(t)-(\chi_P)_{\rm col}\Big) \approx \chi_P(t)-(\chi_P)_{\rm col}, \quad \cos\Big(\chi_P(t)-(\chi_P)_{\rm col}\Big) \approx 1. \qquad (22.31)$$

It is also assumed that the actual velocity vector $V_E(t)$ of the evader will remain close enough to its initial direction satisfying

$$\sin\Big(\chi_E(t)-(\chi_E)_0\Big) \approx \chi_E(t)-(\chi_E)_0, \quad \cos\Big(\chi_E(t)-(\chi_E)_0\Big) \approx 1. \qquad (22.32)$$

Based on the small angle assumptions (22.31) and (22.32), the relative trajectories can be linearized with respect to the initial line of sight. Moreover, the relative motion in the X direction can be considered as uniform. Thus, this coordinate can be replaced by the time, transforming the original three-dimensional motion to a two-dimensional motion in the YZ plane. For a given initial range, the uniform closing velocity determines the final time t_f of the engagement. Therefore, the problem of minimizing (maximizing) the three-dimensional miss distance at a free terminal time can be changed by the minimization (maximization) of the distance in the YZ plane at the fixed terminal time of the nominal collision (two-dimensional miss).

Since in general the velocity vectors $(V_P)_{\rm col}$ and $(V_E)_0$ of the players are not aligned with the initial line of sight, the projections of the originally circular control constraints, normal to the respective velocity vectors, become elliptical.

The equations of motion of the linearized pursuit-evasion game are

$$\begin{array}{ll} \ddot{\xi}_P = F, & \\ \dot{F} = (u-F)/\tau_P, & t \in [0,t_f],\ \xi_P,\ \xi_E \in R^2,\ \ u \in P,\ \ v \in Q, \\ \ddot{\xi}_E = v, & \varphi\big(\xi_P(t_f),\xi_E(t_f)\big) = \big|\xi_P(t_f)-\xi_E(t_f)\big|, \end{array} \qquad (22.33)$$

where ξ_P and ξ_E are the positions of the players in the plane normal to the initial line of sight, and u and v are their respective acceleration command signals.

To reduce dynamics (22.33) to form (22.25), a variable change

$$\begin{array}{lll} z_1 = \xi_{P,1}-\xi_{E,1}, & z_2 = \dot{\xi}_{P,1}-\dot{\xi}_{E,1}, & z_3 = \ddot{\xi}_{P,1}, \\ z_4 = \xi_{P,2}-\xi_{E,2}, & z_5 = \dot{\xi}_{P,2}-\dot{\xi}_{E,2}, & z_6 = \ddot{\xi}_{P,2} \end{array}$$

can be applied, leading to the following standard form of the game

$$\begin{gathered} \dot{z} = Az + Bu + Cv, \\ A = \left[\begin{array}{c|c} A_1 & 0 \\ \hline 0 & A_1 \end{array}\right], \quad A_1 = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & -1/\tau_P \end{bmatrix}, \\ B' = (1/\tau_P)\left[\begin{array}{ccc|ccc} 0 & 0 & 1 & 0 & 0 & 0 \\ \hline 0 & 0 & 0 & 0 & 0 & 1 \end{array}\right], \quad C' = \left[\begin{array}{ccc|ccc} 0 & -1 & 0 & 0 & 0 & 0 \\ \hline 0 & 0 & 0 & 0 & -1 & 0 \end{array}\right] \end{gathered} \qquad (22.34)$$

with constraints for the players' controls taken as ellipses

$$u \in P = \left\{ u \,:\, u' \begin{bmatrix} 1/\cos^2(\chi_P)_{\mathrm{col}} & 0 \\ 0 & 1 \end{bmatrix} u \le a_P^2 \right\},$$
$$v \in Q = \left\{ v \,:\, v' \begin{bmatrix} 1/\cos^2(\chi_E)_0 & 0 \\ 0 & 1 \end{bmatrix} v \le a_E^2 \right\},$$

and the payoff function $\varphi\big(z_1(t_f), z_4(t_f)\big) = \sqrt{z_1^2(t_f) + z_4^2(t_f)}$.

The passage to the equivalent game yields

$$\dot{y} = D(t)u + E(t)v,\ \ t \in [0, t_f],\ \ y \in R^2,\ \ u \in P,\ \ v \in Q,\ \ \varphi\big(y(t_f)\big) = \big|y(t_f)\big|,$$

where

$$D(t) = \zeta(t) \cdot I_2, \quad \zeta(t) = (t_f - t) + \tau_P e^{-(t_f - t)/\tau_P} - \tau_P, \tag{22.35}$$
$$E(t) = \eta(t) \cdot I_2, \quad \eta(t) = -(t_f - t) \tag{22.36}$$

and I_2 is the 2×2 unit matrix.

In order that achieving a small miss distance be feasible, in all realistic pursuit-evasion examples, the pursuer must have some advantage in maximum lateral acceleration in every direction. This means that the control constraint set P of the pursuer has to cover completely the control constraint set Q of the evader. In other words, the inequalities

$$a_P/a_E > 1, \qquad a_P|\cos(\chi_P)_{\mathrm{col}}| > a_E|\cos(\chi_E)_0|, \tag{22.37}$$

describing the relations of the semiaxes of the ellipses P and Q have to be valid. Such an advantage allows reducing an initial launching error and overcoming long duration constant evader maneuvers. However, due to the first-order dynamics of the pursuer's control function and the ideal dynamics of the evader, zero miss distance cannot be achieved against an optimally maneuvering evader.

In Shinar et al. (1984) as well as in Melikyan and Shinar (2000), the parameters of the problem were of an interception of a manned aircraft, assuming that $V_P > V_E$. Thus using (22.30), one obtains $\big|\cos(\chi_P)_{\mathrm{col}}\big| > \big|\cos(\chi_E)_0\big|$.

In Shinar and Zarkh (1996), an interception of a tactical ballistic missile was considered with $V_P < V_E$, leading to $\big|\cos(\chi_P)_{\mathrm{col}}\big| < \big|\cos(\chi_E)_0\big|$.

This difference influences considerably the form of the critical tube in the equivalent game and the associated singular surfaces.

3.5.2 Maximal Stable Bridges: Case of Fast Pursuer

Let us start with the results of numerical investigations for the case when the pursuer's velocity V_P is greater than the velocity V_E of the evader.

Relation (22.30) of the nominal collision and the relation $V_P > V_E$ of the players' velocities yield that the eccentricity of the ellipse P is smaller than the eccentricity of the ellipse Q (see Fig. 22.39).

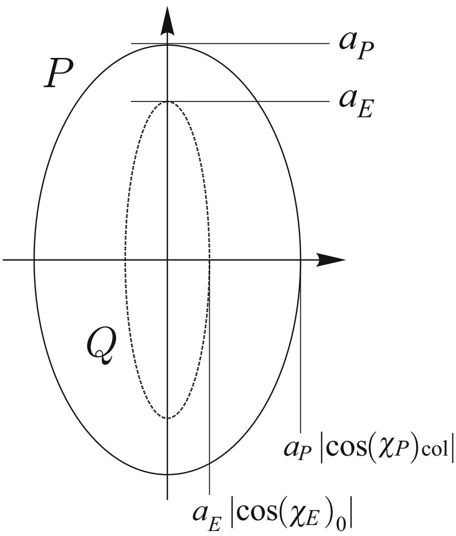


Fig. 22.39 Elliptical constraints of the players' controls in the case of a faster pursuer. The ellipse P is drawn by a *solid line*, and Q is drawn by a *dashed one*. The eccentricity of P is smaller than the eccentricity of Q

In the numerical investigation, the following data were chosen:

$$\frac{V_E}{V_P} = 0.666, \quad \frac{a_P}{a_E} = 5.0, \quad |\cos(\chi_P)_{\rm col}| = 0.87, \quad |\cos(\chi_E)_0| = 0.66, \quad \tau_P = 1\,\text{s}.$$

Consequently, the elliptical control constraints are

$$P = \left\{ u \in R^2 : \frac{u_1^2}{0.87^2} + \frac{u_2^2}{1.00^2} \le 5.0^2 \right\}, \quad Q = \left\{ v \in R^2 : \frac{v_1^2}{0.66^2} + \frac{v_2^2}{1.00^2} \le 1 \right\}.$$

This example has been computed in the interval $\tau \in [0, 2]$ s of backward time. The time step Δ was taken equal to 0.025 s. The circles of the level sets of the payoff function and the ellipses P and Q of constraints for the players' controls were approximated by 100-gons (polygons with 100 vertices).

In Fig. 22.40, the vectogram tubes for this example are shown. The first player's tube is light gray, the second one's is dark gray.

An enlarged part of the previous picture from another point of view can be seen in Fig. 22.41. On the vectogram tube of the second player, the contours of some time sections are shown.

Since $\mathscr{Q}(t) = E(t)Q = \eta(t)I_2Q = \eta(t)Q$, where $\eta(\cdot)$ is described by (22.36), the dark gray tube grows linearly with τ. For the tube $\mathscr{P}$, we have $\mathscr{P}(t) = \zeta(t)P$, where $\zeta(\cdot)$ is taken from (22.35). So, initially (for small values of τ), the light gray tube grows slower than linearly, but for large values of τ, it becomes almost linear and starts to grow faster than the tube $\mathscr{Q}$ does. This faster growth is provided by inequalities (22.37).

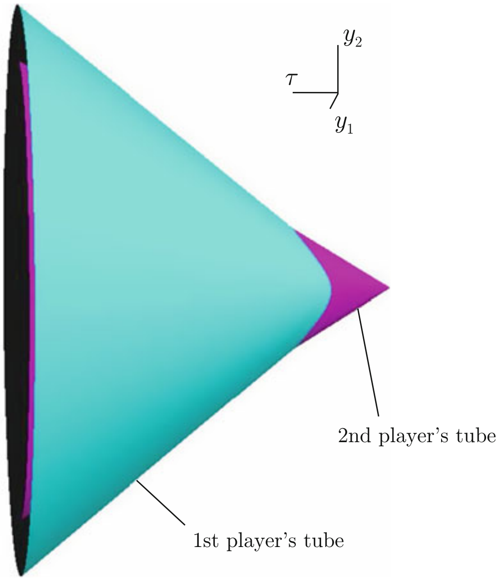


Fig. 22.40 Vectogram tubes for the case of a faster pursuer

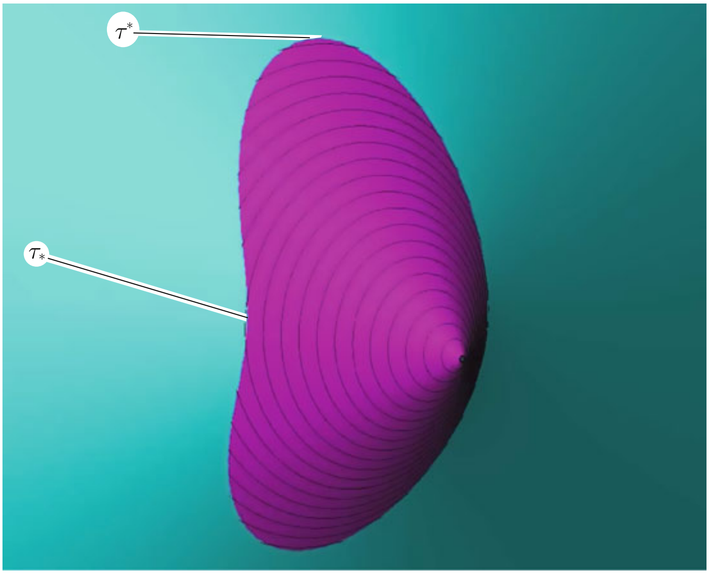


Fig. 22.41 An enlarged fragment of the vectogram tubes. The first player gains advantage in horizontal direction at the instant τ_* and a complete advantage at the instant τ^*

So, for $\tau < \tau_*$, the second player (the maximizer) has a complete advantage, that is, the vectogram $\mathscr{Q}(\tau)$ of the second player completely covers the vectogram $\mathscr{P}(\tau)$ of the first player (Fig. 22.42a). The instant τ_* is characterized by the fact that the horizontal size of the ellipses $\mathscr{P}(\tau_*)$ and $\mathscr{Q}(\tau_*)$ are equal (Fig. 22.42b). In the

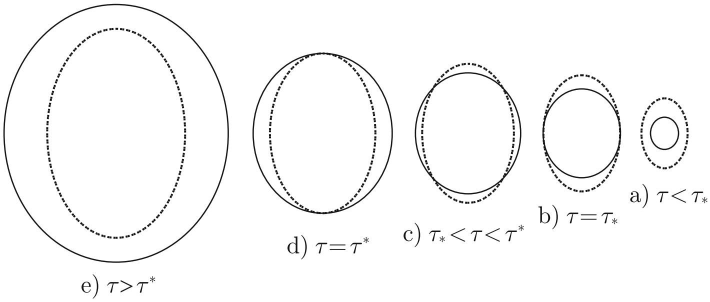


Fig. 22.42 Sections of the vectogram tubes at some time instants. The vectograms of the first player are shown by the *solid lines*, the *dashed lines* denote the vectograms of the second player

interval $\tau_* < \tau < \tau^*$, none of the players has a complete advantage: the first player is stronger in horizontal direction, the second player is stronger in directions near to the vertical (Fig. 22.42c). When $\tau = \tau^*$, the vertical sizes of the ellipses become equal (Fig. 22.42d) and for $\tau > \tau^*$ the first player has complete advantage (Fig. 22.42e). This change in the relationship of the vectograms $\mathscr{P}(\tau)$ and $\mathscr{Q}(\tau)$ can be explained by the difference between the eccentricities and the sizes of the ellipses P and Q and the form of the functions $\zeta(t)$ and $\eta(t)$.

Such a shift of the advantage from the maximizing player to the minimizer leads to creating a narrow throat. Figure 22.43 shows a level set close to the critical one. This level set is computed for $c = 0.141$ m. The narrow throat is located at $\tau^* = 0.925$ s. Contours of some time sections of the level set are shown. One can see that the t-sections $\widetilde{W}_c(t)$ of the level set near the narrow throat are elongated horizontally. This is due to the relation of the players' capabilities. For $\tau < \tau_*$, the second player is stronger in the vertical direction than horizontally. According to this, the sections $\widetilde{W}_c(t)$ are compressed more in the vertical direction. When τ is slightly greater than τ^*, the first player's advantage is stronger in the horizontal direction (Fig. 22.42e), which leads to a horizontal expansion of the sections. For sufficiently large values of τ the first player's advantage in vertical direction becomes greater than in the horizontal one, so, the t-sections start to grow vertically faster than in the horizontal direction, and at some instant, the elongation becomes vertical. For the presented example, this happens outside the time interval of Fig. 22.43.

In Fig. 22.44, scene is given that contains a level set close to the critical one. The τ-axis goes from the right to left, and the axis y_2 is directed upward. The axis y_1 is orthogonal to the sheet. Both vectogram tubes are transparent now. Such an overlap demonstrates clearly the influence of the players' vectograms on the geometry of the level set surface. For example, one can easily see that, when the first player gains complete advantage (at $\tau^* = 0.925$ s), the narrow throat ends (the tube of the level set starts to enlarge). In addition, it is seen that before that instant, the tube contracts due to the advantage of the second player.

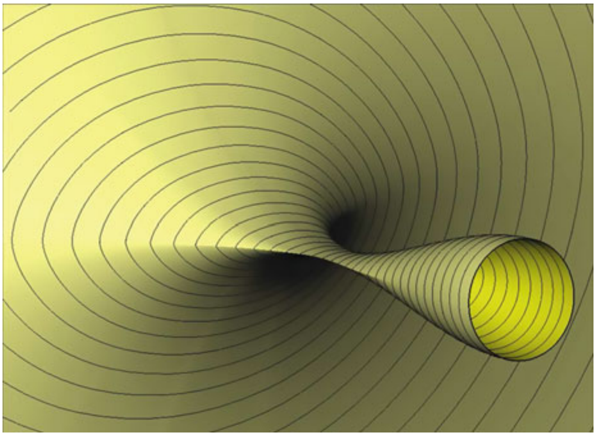


Fig. 22.43 A level set close to the critical one, $c = 0.141$ m. The instant of the most narrow place $\tau^* = 0.925$ s

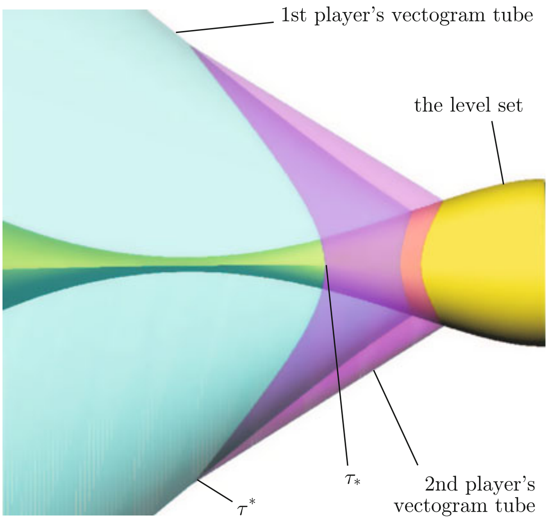


Fig. 22.44 Superposition of the vectogram tubes and the level set close to the critical one. The players' vectogram tubes are transparent

The results shown here agree qualitatively with the ones obtained in an analytical investigation of the problem with a faster pursuer made in Shinar et al. (1984) and Melikyan and Shinar (2000). In these papers, it is shown that in the case of a faster pursuer the geometry of the critical level set is the same for any combination of parameters of the problem.

3.5.3 Maximal Stable Bridges: Case of Slow Pursuer

In this subsection, the results with a slower pursuer $V_P < V_E$ are presented. The eccentricity of the ellipse P is greater than the eccentricity of the ellipse Q (see Fig. 22.45).

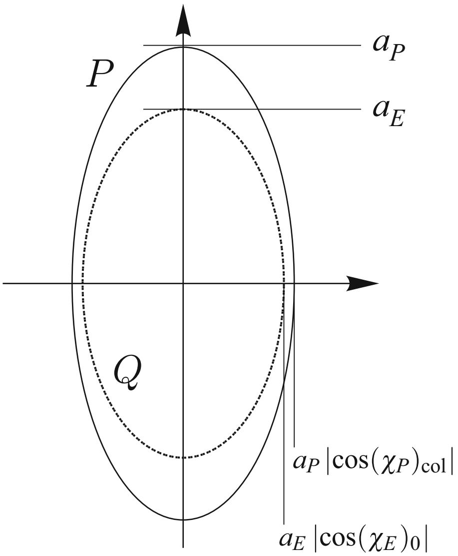


Fig. 22.45 Elliptical control constraints of the players in the case of a slower pursuer. The ellipse P is drawn by a *solid line*, Q is drawn by a *dashed one*. The eccentricity of P is greater than the eccentricity of Q

Based on the data of the original problem

$$\frac{V_E}{V_P} = 1.054, \quad \frac{a_P}{a_E} = 1.3, \quad \left|\cos(\chi_P)_{\text{col}}\right| = 0.67, \quad \left|\cos(\chi_E)_0\right| = 0.71, \quad \tau_P = 1\,\text{s}$$

in the construction, the following data were used:

$$P = \left\{u \in R^2 : \frac{u_1^2}{0.67^2} + \frac{u_2^2}{1.00^2} \le 1.30^2\right\}, \quad Q = \left\{v \in R^2 : \frac{v_1^2}{0.71^2} + \frac{v_2^2}{1.00^2} \le 1\right\}.$$

This example has been computed in the interval $\tau \in [0, 7]$ s with the time step $\Delta = 0.01$ s. The circles of the level sets of the payoff function and the ellipses of the players' control constraints, P and Q, were approximated by 100-gons.

Like in the example of the previous subsection, there is a narrow throat also here. Figure 22.46 shows a general view of the level set $\widetilde{W}_c$ computed for the parameter $c = 1.546$ m, which is slightly greater than the critical one. But unlike the example described above, here the narrow throat has a much more complex structure: the orientation of the t-sections' elongation changes very tricky near the throat. An enlarged view of the throat is shown in Fig. 22.47.

Let us use the players' vectogram tubes for this problem to explain the shape of the level set. The vectogram tubes are shown in Fig. 22.48. The tube of vectograms of the first player ($\mathscr{P}$) is drawn in red, and the tube of the second player ($\mathscr{Q}$) is in green. Here also, the tube $\mathscr{Q}$ grows linearly with τ, whereas the tube $\mathscr{P}$ grows slower than linearly at small values of τ and becomes almost linear later. Eventually, for large values of τ, it will grow faster than the tube $\mathscr{Q}$, because (22.37).

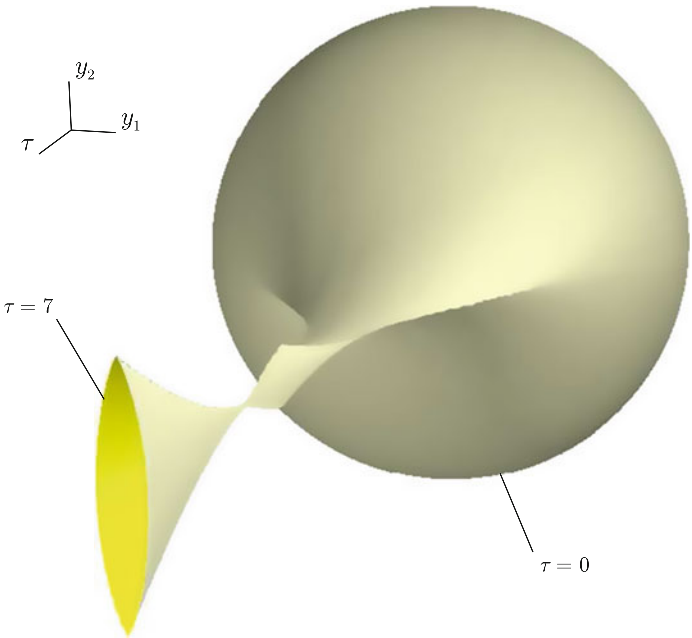


Fig. 22.46 General view of the level set of the value function with a narrow throat

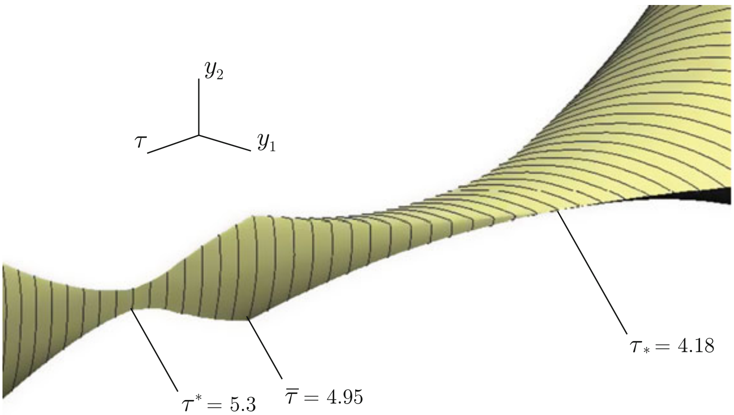
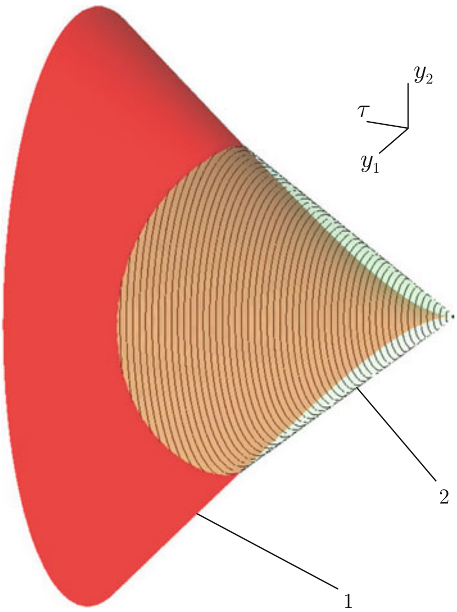


Fig. 22.47 Enlarged view of the narrow throat

Since the ellipses P and Q have different eccentricities, the first player's ellipse $\mathscr{P}(\tau)$ starts to cover the ellipse $\mathscr{Q}(\tau)$ of the second player in different directions at different instants. So, for $\tau < \tau_* = 4.18$ s, the ellipse $\mathscr{Q}(\tau)$ includes the ellipse $\mathscr{P}(\tau)$ completely (see Fig. 22.49a). At $\tau = \tau_*$, the first player's ellipse reaches the ellipse of the second player in the vertical direction (see Fig. 22.49b). In

Fig. 22.48 General view of the vectogram tubes of the first (*1*) and second (*2*) players. The vectogram tube of the second player is transparent, showing contours of some sections



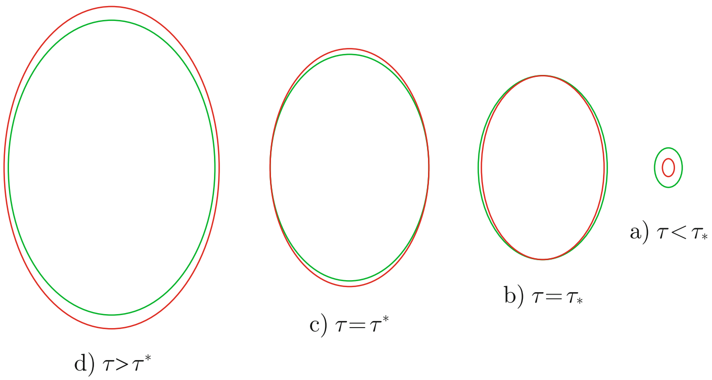


Fig. 22.49 Sections of the vectogram tubes at some time instants. The vectograms of the first player are shown by the *red lines*; the *green lines* denote the vectograms of the second player

the interval $\tau_* < \tau < \tau^*$, the ellipse $\mathscr{P}(\tau)$ covers more and more the ellipse $\mathscr{Q}(\tau)$. Finally, at $\tau = \tau^* = 5.3$ s the set $\mathscr{P}(\tau)$ covers the set $\mathscr{Q}(\tau)$ even in the horizontal direction (see Fig. 22.49c). And for $\tau > \tau_*$, $\mathscr{P}(\tau)$ covers $\mathscr{Q}(\tau)$ completely (see Fig. 22.49d).

The relationship between the players' vectograms leads to an intricate changing of the level set's t-sections near the narrow throat, as it can be seen in Figs. 22.46, 22.47, and 22.50. The latter shows groups of sections in different intervals of τ to demonstrate the different phases of the sections' changing.

For $\tau < \tau_*$, the second player has complete advantage over the first one. Since in backward time the second player tries to contract the sections of the level sets

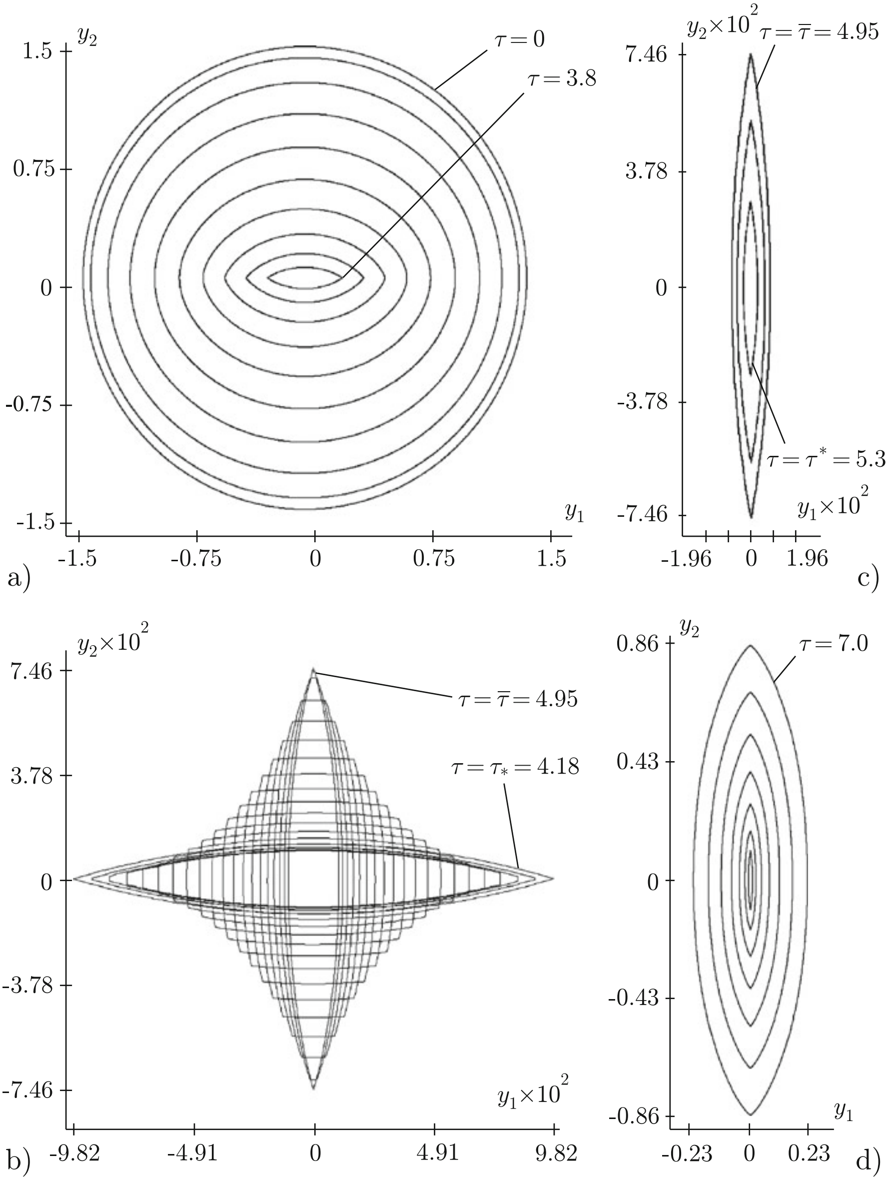


Fig. 22.50 Groups of time sections of a level set close to the critical tube in some intervals of the backward time: (**a**) $\tau \in [0, 3.8]$ s; (**b**) $\tau \in [\tau_*, \bar{\tau}]$; (**c**) $\tau \in [\bar{\tau}, \tau^*]$; (**d**) $\tau \in [5.41, 7.0]$ s

as much as possible, the t-sections of the tube $\widetilde{W}_c$ are reduced in the interval $0 < \tau < \tau_*$. In Fig. 22.50a, the sections are shown in the interval $\tau \in [0, 3.8]$ s. Since the second player's advantage is greater in the vertical direction, the tube starts to contract more in the vertical direction than in the horizontal one. Therefore, at $\tau = \tau_*$ the t-section of the level set is elongated horizontally.

In the interval $\tau_* < \tau < \tau^*$, the first player gains advantage gradually, starting in the vertical direction, while the second player keeps its horizontal advantage. For this reason the t-sections of the level set start expanding vertically while being reduced in the horizontal direction. This interval can be subdivided into two parts.

Between τ_* and $\bar{\tau} = 4.95$ s the time sections have the shape of "curvilinear rectangles" as it can be seen in Fig. 22.50b. Their form is gradually changing from a horizontal elongation to a vertical one.

At $\bar{\tau}$ the horizontal arcs disappear, and the t-sections start having a vertical lens shape. Simultaneously, the vertical expansion becomes a contraction despite of the vertical advantage of the first player (Fig. 22.50c), because the horizontal contraction enforces a contraction due to the lens shape.

Finally, at $\tau = \tau^*$, when the first player gains a complete advantage, one obtains the narrowest section of the throat (Figs. 22.47 and 22.50c) with vertical elongation. For $\tau > \tau^*$ the first player keeps the complete advantage and the t-sections start to expand in all directions monotonically. The rate of expansion is, however, nonuniform, but the direction of elongation remains vertical (see Fig. 22.50d).

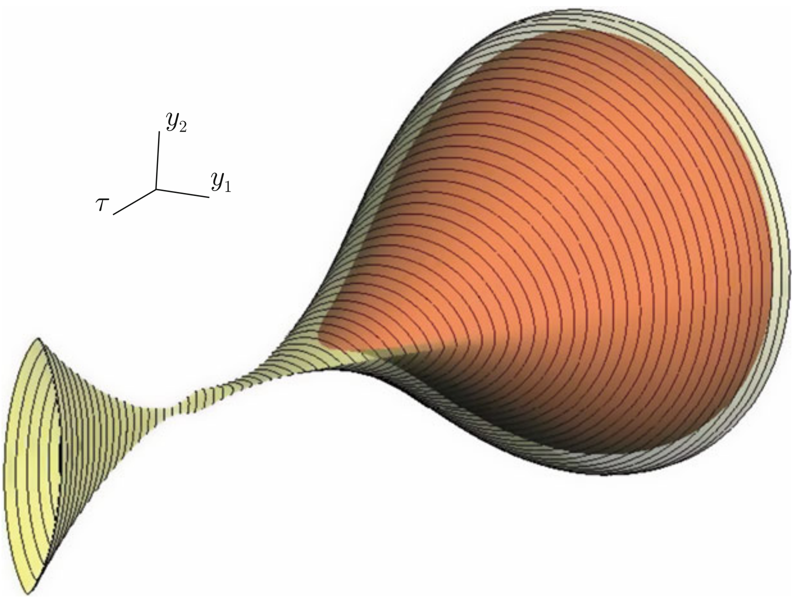


Fig. 22.51 The level set with narrow throat for the parameter $c = 1.546$ m (*yellow transparent*) and the level set for a less value of the parameter $c = 1.48$ m (*red*)

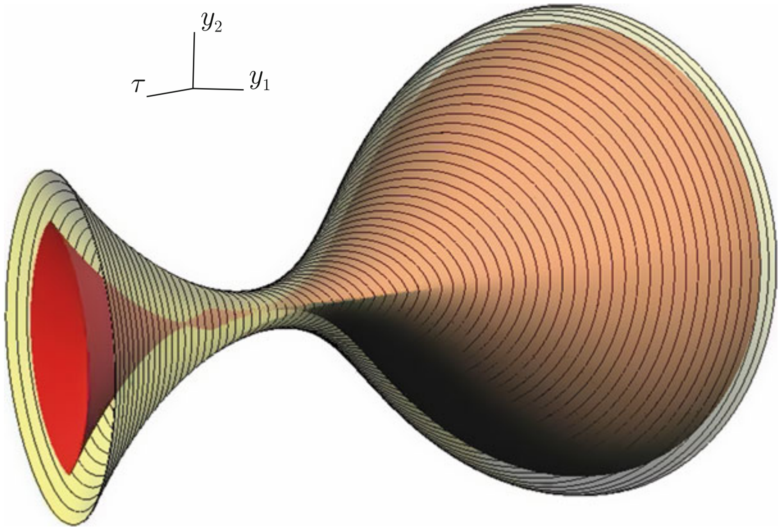


Fig. 22.52 The level set with narrow throat for the parameter $c = 1.546$ m (*red*) and the level set for a greater value of the parameter $c = 1.67$ m (*yellow transparent*)

The following two figures show the critical level set in comparison with level sets close to it. Figure 22.51 shows the critical tube (drawn in transparent yellow) and the tube computed for the value of $c = 1.48$ m of the payoff function, which is less than the critical one. This tube is finite in time and drawn in red. In Fig. 22.52, the critical level set (in red) and the one computed for $c = 1.67$ m (in transparent yellow) are presented. One can see that the latter has smooth boundary. These figures demonstrate that the majority of peculiarities of the value function are found near the narrow throat, emphasizing the necessity of extremely accurate computations near the throat.

The analytical results of the paper Shinar and Zarkh (1996) shows that in the case of a slower pursuer the geometry of the critical tube differs qualitatively for different combinations of the parameters of the problem. The dependence of the critical tube geometry on the parameters of the problem (and how it affects the singular surfaces) is investigated in that paper. The example computed numerically in this subsection corresponds to the case of the most complicated structure of the narrow throat.

3.6 Adaptive Control on the Basis of Differential Game Methods

In the framework of linear differential games, let us describe a method for constructing the first player's control, which can be reasonably called *adaptive*.

Consider a system with linear dynamics

$$\begin{aligned}&\dot{z} = A(t)z + B(t)u + C(t)v,\\&z \in R^n,\ t \in T,\ u \in P \subset R^p,\ v \in R^q,\end{aligned} \tag{22.38}$$

analogous to system (22.25) except that here there is no any compact constraint for the second player's control v. Here, $T = [t_0, t_f]$ is the time interval of process. Assume that the set P contains the origin of the space R^p. Also, assume that the matrix function B is Lipschitzian in the interval T.

The first player tries to guide m chosen components of the phase vector of system (22.38) at the instant t_f to the given terminal set M. The set M is assumed to be a convex compactum in the set of these m coordinates of the phase vector z. Suppose also that the interior of M is not empty and contains the origin, which will be considered as a "center" of M. An additional objective of the first player is to guide these m components of the vector z as closer to the center of M as possible.

As it is described in Sect. 3.1, let us pass to a system without the phase variable in the right-hand side of the dynamics:

$$\begin{aligned}&\dot{y} = D(t)u + E(t)v,\\&y \in R^m,\ t \in T,\ u \in P \subset R^p,\ v \in R^q.\end{aligned} \tag{22.39}$$

The first player tries to guide the phase vector of system (22.39) at the instant t_f to the set M as closer to its center as possible.

All constructions below are for system (22.39). The obtained adaptive control $U(t, y)$ can be applied to system (22.38) too as $U\big(t, Z_m(t_f, t)z\big)$.

3.6.1 System of Stable Bridges

Let the symbol $S(t) = \big\{y \in R^m : (t, y) \in S\big\}$ denote the time section of the set $S \subset T \times R^m$ at the instant $t \in T$. Denote by $O(\varepsilon) = \{y \in R^m : |y| \leqslant \varepsilon\}$ the ball with the radius ε and center at the origin of the space R^m.

Stable bridges. Consider in the interval $[t_0, t_f]$ a zero-sum differential game with a terminal set $\mathbb{M}$ and geometric constraints $\mathbb{P}$, $\mathbb{Q}$ for the players' controls:

$$\begin{aligned}&\dot{y} = D(t)u + E(t)v,\\&y \in R^m,\ t \in T,\ \mathbb{M},\ u \in \mathbb{P},\ v \in \mathbb{Q}.\end{aligned} \tag{22.40}$$

Here, the matrices $D(t)$, $E(t)$ are the same as in system (22.39). The sets $\mathbb{M}$, $\mathbb{P}$, $\mathbb{Q}$ are assumed to be convex compacta. They are regarded as parameters of the game.

Let $u(\cdot)$ and $v(\cdot)$ be measurable functions with their values in the compact sets $\mathbb{P}$ and $\mathbb{Q}$, respectively. A motion of system (22.40) (and, therefore, of system (22.39)) emanated from the point y_* at the instant t_* under controls $u(\cdot)$ and $v(\cdot)$ is denoted by $y\big(\cdot; t_*, y_*, u(\cdot), v(\cdot)\big)$.

Below, the symbol $W_{\mathbb{M}}$ denotes the maximal stable bridge in game (22.40) that stops at the set $\mathbb{M}$ at the instant t_f.

3.6.2 Constructing System of Stable Bridges

1°. Take a set $Q_{\max} \subset R^q$, which is regarded as the maximal constraint for the second player's control can be treated as "reasonable" by the first player when guiding system (22.39) to the terminal set M. Assume that the set $Q_{\max}$ contains the origin of its space. Denote by W_{main} the maximal stable bridge for system (22.40) that corresponds to the parameters $\mathbb{P} = P$, $\mathbb{Q} = Q_{\max}$, $\mathbb{M} = M$. Below, it is called the *main bridge* for brevity.

Suppose additionally that the set $Q_{\max}$ is chosen in such a way that for some $\varepsilon > 0$ for every $t \in T$, the following inclusion holds:

$$O(\varepsilon) \subset W_{\text{main}}(t). \tag{22.41}$$

The value of ε is fixed for further reasonings.

Thus, W_{main} is a closed tube in the space $T \times R^n$ that stops at the set M at the instant t_f. Any t-section $W_{\text{main}}(t)$ is convex and contains the origin of the space R^n with some neighborhood.

2°. Introduce some *additional* closed tube $W_{\text{add}} \subset T \times R^n$ such that any t-section $W_{\text{add}}(t)$ is the reachable set of system (22.40) at the instant t with the initial set $O(\varepsilon)$ taken at the instant t_0. That is, constructing the tube W_{add}, we assume that in dynamics (22.40) the first player is absent ($u \equiv 0$) and the control of the second player is constrained by $Q_{\max}$. One can easily see that W_{add} is the maximal stable bridge for system (22.40) with

$$\mathbb{P} = \{0\}, \quad \mathbb{Q} = Q_{\max}, \quad \mathbb{M} = W_{\text{add}}(t_f).$$

For any $t \in T$, the t-section $W_{\text{add}}(t)$ is convex, and the following inclusion holds:

$$O(\varepsilon) \subset W_{\text{add}}(t). \tag{22.42}$$

3°. Consider a collection of tubes $W_k \subset T \times R^m$, $k \geqslant 0$, whose t-section $W_k(t)$ are defined as

$$W_k(t) = \begin{cases} kW_{\text{main}}(t), & 0 \leqslant k \leqslant 1, \\ W_{\text{main}}(t) + (k-1)W_{\text{add}}(t), & k > 1. \end{cases}$$

The sets $W_k(t)$ are compact and convex. For any numbers $0 \leqslant k_1 < k_2 \leqslant 1 < k_3 < k_4$ due to relations (22.41), (22.42) strict inclusions hold

$$W_{k_1}(t) \subset W_{k_2}(t) \subset W_{k_3}(t) \subset W_{k_4}(t).$$

In works Ganebny et al. (2006, 2007), the following important properties have been justified. A tube W_k for $0 \leqslant k \leqslant 1$ is the maximal stable bridge for system (22.40) that corresponds to a constraint kP for the first player's control, a constraint $kQ_{\max}$ for the second player's control, and a terminal set kM. For $k > 1$, a set W_k is a stable bridge (but, generally speaking, not the maximal one) for the parameters

$$\mathbb{P} = P, \ \ \mathbb{Q} = kQ_{\max}, \ \ \mathbb{M} = M + (k-1)W_{\text{add}}(t_f).$$

Thus, one has a growing system of stable bridges, where each greater bridge corresponds to a greater constraints for the second player's control. This system is generated only by two tubes W_{main} and W_{add} by means of operation of Minkowski sum and multiplication by a nonnegative number parameter.

Feedback control. The adaptive control $(t, y) \mapsto U(t, y)$ itself is constructed in the following way.

Fix a number $r > 0$. Consider a position (t, y). If $|y| \le r$, assume $U(t, y) = 0$. If $|y| > r$, find the minimal number k^* such that the distance between the point y and the t-section $W_{k^*}(t)$ of the bridge W_{k^*} equals r. On the boundary of the set $W_{k^*}(t)$ find the point y^* closest to y. One has $|y^* - y| = r$. Define a vector $u^* \in P_{k^*}$ from the extremum condition

$$(y^* - y)' D(t) u^* = \max \left\{ (y^* - y)' D(t) u : u \in P_{k^*} \right\}.$$

Assume $U(t, y) = u^*$.

Thus, the control U is generated on the basis of the extremal shift rule (i.e., well known in the theory of differential games) and is applied in the *discrete scheme* (Krasovskii 1985; Krasovskii and Subbotin 1974, 1988) with the time step Δ_U. The control is chosen at the beginning of each time step of length Δ_U and kept during the step. In Ganebny et al. (2009), a theorem about the result guaranteed by the control U is formulated and proved.

3.7 Adaptive Control in J. Shinar's Problem

To apply the adaptive method of control, one should introduce an auxiliary constraint $Q_{\max}$. To do this, let us take a reasonable value $a_{E\,\max}$ bounding the lateral acceleration of the evader. This value defines the constraint $Q_{\max}$ as an ellipse

$$Q_{\max} = \left\{ v \in R^2 : \frac{v_1^2}{A_E^2} + \frac{v_2^2}{B_E^2} \le 1 \right\}$$

where the semiaxes A_E, B_E are parallel to the coordinate axes and are computed on the basis of the value $a_{E\,\max}$ and cosine of the angle $(\chi_E)_{\text{nom}}$.

Let us show the simulation results for the case

$$\tau_P = 1.0\,\text{s}, t_f = 10.0\,\text{s}, a_P = 1.3\,\text{m/s}^2, (\chi_P)_{\text{nom}} = 47.94°, (\chi_E)_{\text{nom}} = 45°.$$

The ellipse P, therefore, has the semiaxes equal to $A_P = 1.3$, $B_P = 0.87$. The radius of the terminal circle is taken to be equal to 2.

Let us choose the value $a_{E\ \max} = 1.0\,\text{m/s}^2$. Then, the ellipse $Q_{\max}$ has semiaxes $A_E = 1.0$, $B_E = 0.71$. To construct the adaptive control, one should introduce also the parameter r. Let us take $r = 0.01$. The adaptive control U is applied in the discrete scheme with the time step $\Delta_U = 0.01\,\text{s}$.

The initial phase vector in the difference coordinates is taken as $(z_1^0, z_4^0)' = \xi_P^0 - \xi_E^0 = (-3\,\text{m}, 0\,\text{m})$, $(z_2^0, z_5^0)' = \dot{\xi}_P^0 - \dot{\xi}_E^0 = (0\,\text{m/s}, 2\,\text{m/s})$, $(z_3^0, z_6^0)' = F^0 = 0$. The disturbance control is generated as a piecewise-constant function, which values are in the ellipse $1.5Q_{\max}$ and which stays constant for a random time periods not longer than 3 s. The random procedure for choosing the next value from the ellipse is the following: at first, uniformly we choose an angle from the interval $[0, 2\pi)$, then also uniformly in the radius-vector a point is chosen between the origin and the boundary of ellipse.

In Fig. 22.53a, the phase trajectory of system (22.34) is shown in difference coordinates z_1, z_4. The initial point is denoted by an asterisk and the final one by a black circle. The circle of the terminal set is shown.

In Fig. 22.53b, c, one can see hodographs of the realizations of the controls $u(t)$ and $v(t)$. The hodograph of the control $u(t)$ is inside the ellipse P, the initial and final points are also marked by an asterisk and a black circle. The hodograph of the control $v(t)$ in some time intervals goes outside the ellipse $Q_{\max}$.

Figure 22.54a, b shows graphs of levels of the vector control $u(t)$ with respect to the ellipse P and of the vector disturbance $v(t)$ with respect to the ellipse $Q_{\max}$. There are two intervals of maximality of the useful control: at the beginning of the

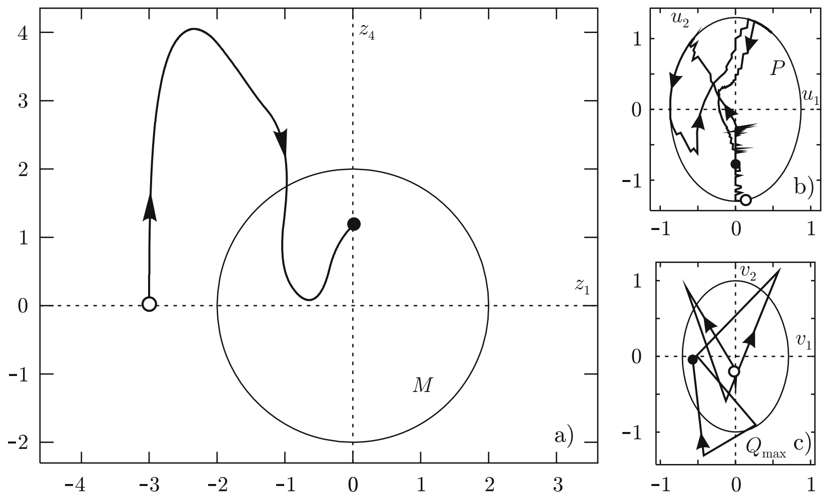


Fig. 22.53 Simulation results: (**a**) the phase trajectory of the system in difference geometric coordinates; (**b**) the hodograph of the useful control u; (**c**) the hodograph of the disturbance v, the ellipse $Q_{\max}$ is shown

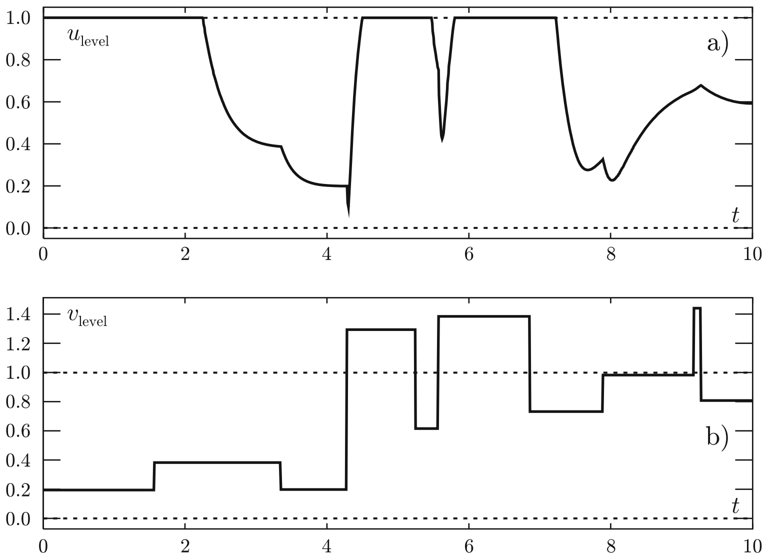


Fig. 22.54 Realizations of controls: (**a**) the graph of the level of the useful control u; (**b**) the graph of the level of the disturbance v with respect to the chosen constraint Q_{max} (the level 1.0)

process, when the initial deviation is diminished, and in the middle of the process, when the disturbance is sufficiently outside the forecasted ellipse Q_{max}. In other intervals, the useful control level is less then maximally possible. One can see how the useful control reaches a level, which corresponds to the level of the disturbance in the next time interval.

Looking at Fig. 22.53a, one can also see that despite the disturbance realization is greater than the chosen level Q_{max}, the process termination is successful: the system is guided inside the terminal set.

3.8 One-Dimensional Differential Games. Linear Problems with Positional Functional. Linear-Quadratic Problems

Differential games with linear dynamics and fixed terminal time permit to reduce the dimension of the phase vector when constructing u-stable bridges; the new phase vector has a dimension equal to the dimension of the target set. Moreover, the t-sections of the bridges in the new coordinates keep the convexity property if the target set is convex. In J. Shinar's problem (Sect. 3.5), a two-dimensional geometric miss is measured at the terminal instant; therefore, the new phase variable is two-dimensional.

1. In problems dealing with space pursuit of one weak-maneuvering object by another one, it is quite reasonable to disjoin the two-dimensional miss in the plane orthogonal to the nominal line-of-sight to two orthogonal components and

to consider two problems each having a one-dimensional miss. In this case, both new problems have one-dimensional phase variable, and maximal stable bridges are constructed in the space *time* $\times$ *phase variable* of dimension 1×1. The construction actually is reduced to one-dimensional integrating to obtain two lines that are upper and lower boundaries of the bridge. Moreover, all upper boundaries as well as lower ones differ from each other by a vertical shift.

J. Shinar and his group used these facts effectively to study dependence of the problem solution on the parameters of the game in the case of one-dimensional miss. Since the maximal u-stable bridge can be regarded as the solvability set of the problem with a given level of the miss (or as a Lebesgue set of the value function), the analysis of its peculiarities is significant both from theoretical and practical points of view.

In particular, in works by J. Shinar and his collaborators, the following questions have been studied. (1) If the object has both air rudders and jet engine, what period of the final stage of the pursuit is more suitable for spending the jet impulse (at the beginning, in the middle, or at the end)? How does it influence to enlargement of the solvability set (Shinar et al. 2012)? (2) What part of the object is more suitable for locating the air rudders (head, middle or rear part) (Shima 2005; Shima and Golan 2006)?

The shape of the solvability set depends significantly on the character of transient processes in the servomechanisms. A study of this question is made in Shinar et al. (2013).

2. In applied problems, the miss distance is often measured not at the terminal instant only but also at some prescribed intermediate ones. With that, the first player minimizes and the second one maximizes some functional that depends on all these misses. The differential games that involve such nonterminal functionals are studied in book A.N.Krasovskii and N.N.Krasovskii (1995). In this book, the concept of positional functional is introduced for which the optimal strategies of the first and second players do not depend on the history of the process as it is for the case of the terminal payoff. For systems with linear dynamics with convex positional functional, the book suggests a constructive method for computing the value function based on a convex hull construction operation. In works Lukoyanov (1998), Kornev (2012), and Gomoyunov et al. (2015), the main stages of this method, proofs, and corresponding numerical procedures are considered in details.

3. In Sect. 3, we consider differential games with linear dynamics, fixed terminal instant, and geometric constraints on the players' controls. Taking into account such "hard-bounded control limits" complicates sufficiently optimal feedback control laws in comparison with linear-quadratic (LQ) formulations, where constraints on the players' controls are introduced "softly" by means of integral quadratic functionals. Linear feedback control laws obtained in the framework of LQ formulations are very popular in engineering practice. Among a large number of works, we would like to mention book (1998) by J.Z. Ben-Asher and I. Yaesh. This book is oriented to students and engineers specializing in missile guidance. It is written in informal mathematical style. There are a lot of topics and examples from the simplest LQ classical optimal control problems to game LQ methods under inexact parameters

or inexact time-to-go of the control process. The book includes listings of MAPLE and MATLAB subroutines for certain problems of missile guidance.

4 Conclusion

The theory of differential games is very lucky because its first author and founder R. Isaacs being an excellent mathematician (the number theory was his first area of interests in young age) turned to the new work at the beginning of the 1950s having before his eyes a bundle of practical problems. His book "Differential Games" is an enthusiastic anthem to problems of this type. With that, as well as L.S. Pontryagin in the theory of optimal control, he started to investigate one of the most difficult classes, the time-optimal problems. The object that R. Isaacs called "car" is now extremely widespread in works on aviation, robotics, sea navigation, etc.

Just a time-optimal game including exactly this object (the "homicidal chauffeur" game) is chosen by us as the central one for the first part of this chapter (Sect. 2). We have collected some applied problems (first of all, those have been considered in works by J.V. Breakwell and A.W. Merz) to show how deeply they can be studied and how various they can be.

In the second part of this chapter (Sect. 3), we have concentrated on differential games with linear dynamics and fixed terminal instant. J. Shinar showed plenty of problems from the aerospace navigation, in which linearization of the dynamics and fixation of the terminal instant are relevant. The presence of both these factors allows one to pass to an equivalent differential game with the phase variable having the dimension, possibly, sufficiently lower than in the original game. Moreover, if the terminal payoff is convex, then the level sets of the value function (the solvability sets, the maximal stable bridges) of such a problem have convex t-sections. This simplifies greatly the solution of the game.

Choosing such topics for our chapter, we have tried to solve numerically several problems taken from these areas. The numerical method used for solving time-optimal problems in the first part is heuristic in many respects. Computation of each problem needs an "individual supervision" of the program execution. That is why the method is set forth very schematically, only its main idea is described. Our major objective is to give quite exact pictures of the value function, which is often discontinuous and, generally speaking, has non-convex level sets. Also, there can be places of fast change of the value function that are expressed in condensation of contour curves, which is a difficult situation from the computational point of view. Our results have been checked on the basis of many examples computed by other methods and by other authors.

Numerical constructions in the second part are mainly connected with convexification operation (construction of the convex hull). If the problems are of optimal control type, that is, if they do not include the second player, then in the situation of convex payoff, the t-sections of the level sets would be convex automatically without any additional convexifications. In the game problems, even if the payoff is convex, at each step we should involve the convex hull construction procedure that complicates the algorithm. At the same time, our procedure is much faster in comparison with general convex hull construction methods because in the framework of our algorithm we have information about places of possible local convexity violation of the processed function. If the phase variable of the game is two-dimensional, this information can be effectively used. The convexification algorithm is given in sufficient details, but descriptively, not in a strict programmatic manner. As in the first part, we use our own visualization software when demonstrating the results of numerical constructions.

In the last subsections of both parts, we touch upon some publications on applied problems close to those have been considered but much more difficult. The corresponding text includes short descriptions of these problems and some remarks on them.

Acknowledgements The authors are sincerely thankful to E. Bakolas, J. Fisac, V. Glizer, J.-P. Laumond, E.P. Maslov, S. Le Ménec, M. Pachter, H.J. Pesch, L.A. Petrosyan, V. Shaferman, I.I. Shevchenko, T. Shima, P. Tsiotras, and V. Turetsky for their useful consultations.

The authors are grateful to the reviewer for the very helpful remarks, suggestions, and corrections.

The third author acknowledges support by DFG, grant TU427/2-1.

The authors gratefully acknowledge the photographs in Figs. 1 and 2 which were kindly provided by Ellen Sara Isaacs, Antony Willitz Merz, John Alexander Breakwell, and Noa Shinar-Mulder.

References

Abramyants TG, Maslov EP, Yakhno VP (2007) Evasion from detection in the three-dimensional space. J Comput Syst Sci Int 46(5):675–680

Abramyants TG, Maslov EP, Rubinovich EYa (1981) A simplest differential game of alternate pursuit. Autom Remote Control 41(8):1043–1052

Aubin JP (1991) Viability theory. Birkhäuser, Basel

Averbukh VL, Ismagilov TR, Patsko VS, Pykhteev OA, Turova VL (2000) Visualization of value function in time-optimal differential games. In: Handlovičová A, Komorníková M, Mikula K, Ševčovič D (eds) Algoritmy 2000, 15th Conference on Scientific Computing, Vysoke Tatry – Podbanske, 10–15 Sept 2000, pp 207–216

Başar T, Olsder GJ (1995) Dynamic noncooperative game theory. Academic Press, London

Bakolas E, Tsiotras P (2012) Feedback navigation in an uncertain flow-field and connections with pursuit strategies. J Guid Control Dyn 35(4):1268–1279

Bakolas E, Tsiotras P (2013) Optimal synthesis of the Zermelo-Markov-Dubins problem in a constant drift field. JOTA 156:469–492

Ben-Asher JZ, Yaesh I (1998) Advances in missile guidance theory. Progress in astronautics and aeronautics. AIAA, New York

Berdyshev YI (2002) A problem of the sequential approach to a group of moving points by a third-order non-linear control system. J Appl Math Mech 66(5):709–718

Berdyshev YI (2008) On a nonlinear problem of a sequential control with a parameter. J Comput Syst Sci Int 47(3):380–385

Berdyshev YI (2015) Nonlinear problems of sequential control and their application. UrO RAN, Ekaterinburg (in Russian)

Berkovitz LD (1994) A theory of differential games. In: Başar T, Haurie A (eds) Annals of the international society of dynamic games. Advances in dynamic games and applications, vol 1. Birkhäuser, Boston, pp 3–22

Bernhard P (1970) Linear differential games and the isotropic rocket. Ph.D. thesis, Stanford University

Bernhard P (1977) Singular surfaces in differential games, an introduction. In: Haggedorn P, Knobloch HV, Olsder GJ (eds) Differential games and applications. Lecture notes in control and information sciences. Springer, Berlin, pp 1–33

Bernhard P (1979) Contribution à l'étude des jeux différentiels à deux joueurs, somme nulle et information parfaite. thèse de doctorat d'État

Bernhard P, Larrouturou B (1989) Etude de la barriere pour un probleme de fuite optimale dans le plan. rapport de recherche

Bernhard P, Pourtallier O (1994) Pursuit evasion game with costly information. Dyn Control 4(4):365–382

Bhattacharya S, Başar T (2012) Differential game-theoretic approach to a spatial jamming problem. In: Cardaliaguet P, Cressman R (eds) Advances in dynamic game: theory, applications and numerical methods for differential and stochastic games. Annals of the international society of dynamic games, vol 12. Birkhäuser, Boston, pp 245–268

Blagodatskih AI, Petrov NN (2009) Conflict interaction controlled objects groups. Udmurt State University, Izhevsk (in Russian)

Blaquière A, Gérard F, Leitmann G (1969) Quantitative and qualitative differential games. Academic Press, New York

Bolonkin AA, Murphey RA (2005) Geometry-based parametric modeling for single-pursuer multiple evader problems. J Guid Control Dyn 28(1):145–149

Botkin ND (1982) Evaluation of numerical construction error in differential game with fixed terminal time. Probl Control Inf Theor 11:283–295

Botkin ND (1993) Asymptotic behavior of solution in differential games. viability domains of differential inclusions. Russian Acad Sci Dokl Math 46(1):8–11

Botkin ND, Hoffmann KH, Mayer N, Turova VL (2013) Computation of value functions in nonlinear differential games with state constraints. In: Hömberg D, Tröltzsch F (eds) Proceedings of the 25th IFIP TC7 Conference on System Modeling and Optimization, pp 235–244

Botkin ND, Hoffmann KH, Turova VL (2011) Stable numerical schemes for solving Hamilton-Jacobi-Bellmann-Isaacs equations. SIAM J Sci Comput 33(2):992–1007

Botkin ND, Kein VM, Patsko VS (1984) The model problem of controlling the lateral motion of an aircraft during landing. J Appl Math Mech 48(4):395–400

Botkin ND, Zarkh MA. Estimation of error for numerical constructing absorption set in linear differential game. In: Subbotin and Patsko (1984), pp 39–80 (in Russian)

Breakwell JV, Merz AW (1969) Toward a complete solution of the homicidal chauffeur game. In: Proceedings of the 1st International Conference on the Theory and Application of Differential Games, Amherst, pp III-1–III-5

Breitner MH (2005) The genesis of differential games in light of Isaacs' contributions. JOTA 124(3):523–559

Bryson AE, Ho YC (1975) Applied optimal control. Optimization, estimation and control. Hemisphere Publishing Corporation/John Wiley and Sons, New York

Bulirsch R, Montrone F, Pesch HJ (1991) Abort landing in the presence of windshear as a minimax optimal control problem, Part 1: necessary conditions. JOTA 70(1):1–23

Bulirsch R, Montrone F, Pesch HJ (1991) Abort landing in the presence of windshear as a minimax optimal control problem, Part 2: multiple shooting and homotopy. JOTA 70(2):223–254

Cardaliaguet P, Quincampoix M, Saint-Pierre P (1995) Numerical methods for optimal control and differential games. Technical report, Ceremade CNRS URA 749, University of Paris – Dauphine

Cardaliaguet P, Quincampoix M, Saint-Pierre P (1999) Set-valued numerical analysis for optimal control and differential games. In: Bardi M, Raghavan TES, Parthasarathy T (eds) Stochastic and differential games: theory and numerical methods. Annals of the international society of dynamic games, vol 4. Birkhäuser, Boston, pp 177–247

Chen M, Fisac J, Sastry S, Tomlin C (2015) Safe sequential path planning of multi-vehicle systems via double-obstacle Hamilton-Jacobi-Isaacs variational inequality. In: Proceedings of the 14th European Control Conference, pp 3304–3309
Chentsov AG (1976) On a game problem of converging at a given instant of time. Math USSR-Sb 99(141)(3):394–420
Chentsov AG (1978) An iterative program construction for a differential game with fixed termination time. Soviet Math Dokl 19(3):559–562
Chentsov AG (1978) On a game problem of converging at a given instant of time. Math USSR-Sb 19(3):559–562
Chernous'ko FL (1976) A problem of evasion from many pursuers. J Appl Math Mech 4(1):11–20
Chernous'ko FL (1993) State estimation for dynamic systems. CRC Press, Boca Raton
Chernous'ko FL, Melikyan AA (1978) Game problems of control and search. Nauka, Moscow (in Russian)
Chikrii AA (1997) Conflict-controlled processes. Mathematics and its applications, vol 405. Kluwer Academic Publishers Group, Dordrecht
Courant R (1962) Partial differential equations. InterScience, New York
Crandall MG, Evans LC, Lions PL (1984) Some properties of viscosity solutions of Hamilton-Jacobi equations. Trans Am Math Soc 282(1):487–502
Crandall MG, Lions PL (1983) Viscosity solutions of Hamilton-Jacobi equations. Trans Am Math Soc 277(1):1–42
Davidovitz A, Shinar J (1989) Two-target game model of an air combat with fire-and-forget all-aspect missiles. JOTA 63(2):133–165
Dubins LE (1957) On curves of minimal length with a constraint on average curvature and with prescribed initial and terminal positions and tangents. Am J Math 79:497–516
Dvurechenskii PE (2013) Algorithms for construction of ε-optimal strategies in nonlinear differential games in the plane. Ph.D. thesis, Moscow Institute of Physics and Technology, Moscow (in Russian)
Dvurechenskii PE, Ivanov GE (2012) An algorithm for constructing optimal strategy in nonlinear differential game with nonfixed time of termination. Proc Moscow Inst Phys Tech 4(4):51–61 (in Russian)
Evans LC (1998) Partial differential equations. AMS, New York
Exarchos I, Tsiotras P, Pachter M (2015) On the suicidal pedestrian differential game. Dyn Games Appl 5(3):297–317
Falcone M (2006) Numerical methods for differential games based on partial differential equations. Int Game Theory Rev 8(2):231–272
Fisac JF, Sastry S (2015) The pursuit-evasion-defense differential game in dynamic constrained environments. In: Proceedings of the IEEE 54th Annual Conference on Decision and Control (CDC), pp 4549–4556
Ganebny SA, Kumkov SS, Patsko VS (2006) Control design in problems with an unknown level of dynamic disturbance. J Appl Math Mech 70(5):680–695
Ganebny SA, Kumkov SS, Patsko VS (2009) Extremal aiming in problems with an unknown level of dynamic disturbance. J Appl Math Mech 73:411–420
Ganebny SA, Kumkov SS, Patsko VS, Pyatko SG (2007) Constructing robust control in differential games: application to aircraft control during landing. In: Jørgensen S, Quincampoix M, Vincent T (eds) Annals of the international society of dynamic games, vol 9. Birkhäuser, Boston, pp 69–92
Getz WM, Pachter M (1981) Two-target pursuit-evasion differential games in the plane. JOTA 34(3):383–403
Gomoyunov MI, Kornev DV, Lukoyanov NY (2015) On the numerical solution of a minimax control problem with a positional functional. Proc Steklov Inst Math 291(1):S77–S95
Grigorenko NL (1989) The pursuit problem in n-person differential games. Math USSR-Sb 63(1):35–45
Grigorenko NL (1990) Mathematical methods of control of multiple dynamic processes. Moscow State University, Moscow (in Russian)

Grigor'eva SV, Pakhotinskikh VYu, Uspenskii AA, Ushakov VN (2005) Construction of solutions in some differential games with phase constraints. Sb Math 196(4):513–539
Grigor'eva SV, Tarasyev AM, Uspenskii AA, Ushakov VN (2000) Constructions of the differential game theory for solving the Hamilton-Jacobi equations. Proc Steklov Inst Math (Suppl Issues) Suppl 2:S38–S53
Gutman S (1979) On optimal guidance for homing missile. J Guid Control 3(4):296–300
Gutman S, Leitmann G (1975) On a class of linear differential games. JOTA 17(5–6):511–522
Gutman S, Leitmann G (1976) Optimal strategies in the neighborhood of a collision course. AIAA J 14(9):1210–1212
Hagedorn P, Breakwell JV (1976) A differential game with two pursuers and one evader. JOTA 18(2):15–29
Han Z, Niyato D, Saad W, Başar T, Hjorungnes A (2011) Game theory in wireless and communication networks: theory, models, and applications. Cambridge University Press, Cambridge/New York
Isaacs R (1951) Games of pursuit. Scientific Report of the RAND Corporation. Technical report, RAND Corporation, Santa Monica
Isaacs R (1965) Differential games. John Wiley and Sons, New York
Isakova EA, Logunova GV, Patsko VS Computation of stable bridges for linear differential games with fixed time of termination. In: Subbotin and Patsko (1984), pp 127–158 (in Russian)
Kim DP (1993) Methods of search and pursuit of mobile objects. Nauka, Moscow (in Russian)
Kornev DV (2012) On numerical solution of positional differential game with nonterminal payoff. Autom Remote Control 73:1808–1821
Krasovskii AN, Krasovskii NN (1995) Control under lack of information. Birkhäuser, Berlin
Krasovskii NN (1970) Game-theoretic problems on the encounter of motions. Nauka, Moscow (in Russian)
Krasovskii NN (1971) Rendezvous game problems. National Technical Information Service, Springfield
Krasovskii NN (1985) Control of a dynamic system. The minimum problem of a guaranteed result. Nauka, Moscow (in Russian)
Krasovskii NN, Subbotin AI (1974) Positional differential games. Nauka, Moscow (in Russian)
Krasovskii NN, Subbotin AI (1988) Game-theoretical control problems. Springer, New York
Kryazhimskiy AV, Osipov YS (2010) Idealized program packages and problems of positional control with incomplete information. Proc Steklov Inst Math 268:155–174
Kumkov SS, Le Ménec S, Patsko VS (2017) Zero-sum pursuit-evasion differential games with many objects: survey of publications. Dyn Games Appl 7(4):609–633
Kurzhanski AB (2015) On a team control problem under obstacles. Proc Steklov Inst Math 291:128–142
Kurzhanski AB (2016) Problem of collision avoidance for a team motion with obstacles. Proc Steklov Inst Math 293:120–136
Kurzhanski AB, Valyi I (1997) Ellipsoidal calculus for estimation and control. Birkhäuser, Boston
Kurzhanski AB (1977) Control and observation under uncertainty. Nauka, Moscow (in Russian)
Kurzhanski AB (2004) The problem of measurement feedback control. J Appl Math Mech 68(4):487–501
Laumond JP (ed) (1998) Robot motion planning and control. Lecture notes in control and information science, vol 229. Springer, New York
Laumond JP, Mansard N, Lasserre JB (2014) Optimality in robot motion: optimal versus optimized motion. Commun ACM 57(9):82–89
Leitmann G, Pandey S (1991) Aircraft control for flight in an uncertain environment: take-off in windshear. JOTA 70(1):25–55
Levchenkov AY, Pashkov AG (1990) Differential game of optimal approach of two inertial pursuers to a noninertial evader. JOTA 65(3):501–518
Lewin J (1973) Decoy in pursuit-evasion games. Ph.D. thesis, Stanford University
Lewin J (1994) Differential games. Springer, London
Lewin J, Breakwell JV (1975) The surveillance-evasion game of degree. JOTA 16(3–4):339–353

Lewin J, Olsder GJ (1979) Conic surveillance evasion. JOTA 27(1):107–125
Lions PL (1982) Generalized solutions of Hamilton-Jacobi equations. Research notes in mathematics, vol 69. Pitman, Boston
Lukoyanov NY (1998) The problem of computing the value of a differential game for a positional functional. J Appl Math Mech 62(2):177–186
Markov AA (1889) Some examples of the solution of a special kind of problem on greatest and least quantities. Soobscenija Charkovskogo matematiceskogo obscestva 2, 1(5, 6):250–276 (in Russian)
Melikian AA (1975) On optimum selection of noise intervals in differential games of encounter. J Appl Math Mech 39(2):195–203
Melikyan AA (1973) On minimal observations in a game of encounter. J Appl Math Mech 37(3):407–414
Melikyan AA (1998) Generalized characteristics of first order PDEs. Birkhäuser, Boston
Melikyan AA, Shinar J (2000) Identification and construction of singular surface in pursuit-evasion games. In: Filar JA, Gaitsgory V, Mizukami K (eds) Annals of the international society of dynamic games, vol 5. Birkhäuser, Boston, pp 151–176
Merz AW (1971) The homicidal chauffeur – a differential game. Ph.D. thesis, Stanford University
Merz AW (1972) The game of two identical cars. JOTA 9(5):324–343
Merz AW (1974) The homicidal chauffeur. AIAA J 12(3):259–260
Merz AW (1985) To pursue or to evade – that is the question. J Guid Control Dyn 8(2): 161–166
Meyer A, Breitner MH, Kriesell M (2005) A pictured memorandum on synthesis phenomena occurring in the homicidal chauffeur game. In: Martin-Herran G, Zaccour G (eds) Proceedings of the Fifth International ISDG Workshop, Segovia, 21–24 Sept 2005, pp 17–32
Miele A, Wang T, Melvin WW (1986) Optimal take-off trajectories in the presence of windshear. JOTA 49(1):1–45
Miele A, Wang T, Tzeng CY, Melvin WW (1987) Optimal abort landing trajectories in the presence of windshear. JOTA 55(2):165–202
Miele A, Wang T, Wang H, Melvin WW (1988) Optimal penetration landing trajectories in the presence of windshear. JOTA 57(1):1–40
Mikhalev DK, Ushakov VN (2007) Two algorithms for approximate construction of the set of positional absorption in the game problem of pursuit. Autom Remote Control 68(11): 2056–2070
Mishchenko EF, Nikol'skii MS, Satimov NYu (1977) The problem of avoiding encounter in n-person differential games. Trudy Mat Inst Steklov 143:105–128
Mitchell I (2002) Application of level set methods to control and reachability problems in continuous and hybrid systems. Ph.D. thesis, Stanford University
Mitchell I, Bayen A, Tomlin C (2005) A time-dependent Hamilton-Jacobi formulation of reachable sets for continuous dynamic games. IEEE Trans Autom Control 50(7):947–957
Neveu D, Pignon JP, Raimondo A, Nicolas JM, Pourtallier O (1995) Pursuit games with costly information: application to the ASW helicopter versus submarine game. In: Olsder GJ (ed) New trends in dynamic games and applications. Annals of the international society of dynamic games, vol 3. Birkhäuser, Boston, pp 247–257
Olsder GJ, Breakwell JV (1974) Role determination in aerial dogfight. Int J Game Theory 3:47–66
Olsder GJ, Pourtallier O (1995) Optimal selection of observation times in a costly information game. In: Olsder GJ (ed) New trends in dynamic games and applications. Annals of the interantional society of dynamic games, vol 3. Birkhäuser, Boston, pp 227–246
Osipov YS (2006) Program packages: an approach to the solution of positional control problems with incomplete information. Russ Math Surv 61(4):611–661
Pachter M, Garcia E, Casbeer D (2014) Active target defense differential game. In: Proceedings of the 52nd Allerton Conference on Communication, Control and Computing, Monticello, pp 46–53
Pachter M, Getz W (1980) The geometry of the barrier in the game of two cars. Optim Control Appl Met 1(2):103–118

Patsko VS (1975) Problem of quality in linear differential games of the second order. In: Kurzhanski AB (ed) Differential games and control problems. UrSC AS USSR, Sverdlovsk, pp 167–227 (in Russian)
Patsko VS (1996) Special aspects of convex hull constructing in linear differential games of small dimension. In: Control applications of optimization 1995. A Postprint Volume from the IFAC Workshop, Haifa, 19–21 Dec 1995. Pergamon, New York, pp 19–24
Patsko VS, Botkin ND, Kein VM, Turova VL, Zarkh MA (1994) Control of an aircraft landing in windshear. JOTA 83(2):237–267
Patsko VS, Turova VL (1995) Numerical solution of two-dimensional differential games. Technical report, UrO RAN. Institute of Mathematics and Mechanics, Ekaterinburg, Ekaterinburg
Patsko VS, Turova VL (1997) Numerical solutions to the minimum-time problem for linear second-order conflict-controlled systems. In: Bainov D (ed) Proceedings of the Seventh International Colloquium on Differential Equations, Plovdiv, 18–23 Aug 1996, pp 329–338
Patsko VS, Turova VL (2001) Level sets of the value function in differential games with the homicidal chauffeur dynamics. IGTR 3(1):67–112
Patsko VS, Turova VL (2004) Families of semipermeable curves in differential games with the homicidal chauffeur dynamics. Automatica 40(12):2059–2068
Patsko VS, Turova VL (2009) Numerical investigation of the value function for the homicidal chauffeur problem with a more agile pursuer. In: Bernhard P, Gaitsgory V, Pourtallier O (eds) Advances in dynamic games and their applications: analytical and numerical developments. Annals of the international society of dynamic games, vol 10. Birkhäuser, Boston, pp 231–258
Petrosjan LA (1965) A family of differential survival games in the space R^n. Soviet Math Dokl 6(2):377–380
Petrosyan LA (1966) Pursuit lifeline games with many participants. Proc Armen Acad Sci Ser Math 1:331–340
Petrosyan LA (1970) Differential games with incomplete information. Soviet Math Dokl 11(6):1524–1528
Petrosyan LA (1977) Differential games of pursuit. Leningrad State University, Leningrad, USSR (in Russian)
Petrosyan LA (1993) Differential games of pursuit. World Scientific Publisher, London
Petrosyan LA, Dutkevich YG (1972) Games with a "lifeline". The case of l-capture. SIAM J Control 10(1):40–47
Petrosyan LA, Garnaev AY (1993) Search games. St. Petersburg Gos. University, St.-Petersburg (in Russian)
Petrosyan LA, Shiryaev VD (1980) Group pursuit by one pursuer of many evaders. Viestn Leningr Univ 13(3):50–59
Petrov NN (1988) A group pursuit problem with phase constraints. J Appl Math Mech 52(6): 803–806
Polovinkin ES, Ivanov GE, Balashov MV, Konstantiov RV, Khorev AV (2001) An algorithm for the numerical solution of linear differential games. Sb Math 192(10):1515–1542
Ponomarev AP, Rozov NH (1978) Stability and convergence of alternating Pontrjagin sums. Univ, Ser. 15: Vycisl. Mat. i Kibernet 1:82–90 (in Russian)
Pontryagin LS, Mischenko EF (1971) The problem of evasion in linear differential games. Diff Urav 7(2):436–445 (in Russian)
Pontryagin LS (1967) Linear differential games, 1. Soviet Math Dokl 8:769–771
Pontryagin LS (1967) Linear differential games, 2. Soviet Math Dokl 8:910–912
Pontryagin LS (1971) A linear differential escape game. Proc Steklov Inst Math 112:27–60
Pschenichnyi BN (1976) Simple pursuit by several objects. Cybern Syst Anal 12(3):484–485
Pschenichnyi BN, Chikrii AA, Rappoport IS (1981) An efficient method of solving differential games with many pursuers. Soviet Math Dokl 23(1):104–109
Pschenichnyi BN, Sagaidak MI (1970) Differential games of prescribed duration. Cybernetics 6(2):72–83
Raivio T, Ehtamo H (2000) On the numerical solution of a class of pursuit-evasion games. In: Filar JA, Gaitsgory V, Mizukami K (eds) Advances in dynamic games and applications. Annals of the international society of dynamic games, vol 5. Birkhäuser, Boston, pp 177–192

Reeds JA, Shepp LA (1990) Optimal paths for a car that goes both forwards and backwards. Pac J Math 145(2):367–393
Selvakumar J, Bakolas E (2015) Optimal guidance of the isotropic rocket in a partially uncertain flow. In: Proceedings of European Control Conference, Linz, pp 3328–3333
Shaferman V, Shima T (2008) Unmanned aerial vehicles cooperative tracking of moving ground target in urban environments. J Guid Control Dyn 31(5):1360–1371
Shaferman V, Shima T (2010) Cooperative multiple model adaptive guidance for an aircraft defending missile. J Guid Control Dyn 33(6):1801–1813
Shevchenko I (1997) Successive pursuit with a bounded detection domain. JOTA 95(1):25–48
Shevchenko I (2012) Locally optimizing strategies for approaching the furthest evader. In: Petrosyan LA, Zenkevich NA (eds) Contributions to game theory and management, vol 5. Graduate School of Management, St. Petersburg State University, St.Petersburg, pp 293–303
Shima T (2005) Capture conditions in a pursuit-evasion game between players with biproper dynamics. JOTA 126(3):503–528. https://doi.org/10.1007/s10957-005-5495-3
Shima T (2011) Optimal cooperative pursuit and evasion strategies against a homing missile. J Guid Control Dyn 34(2):414–425
Shima T, Golan OM (2006) Bounded differential games guidance law for dual-controlled missiles. IEEE Trans Control Syst Tech 14(4):719–724
Shinar J, Glizer VY, Turetsky V (2012) Complete solution of a pursuit-evasion differential game with hybrid evader dynamics. IGTR 14(3):1–31
Shinar J, Glizer VY, Turetsky V (2013) The effect of pursuer dynamics on the value of linear pursuit-evasion games with bounded controls. In: Krivan V, Zaccour G (eds) Annals of the international society of dynamic games, vol 13. Dynamic games – theory, applications, and numerical methods. Birkhäuser, Basel, pp 313–350. https://doi.org/10.1007/978-3-319-02690-9_15
Shinar J, Gutman S (1980) Three-dimensional optimal pursuit and evasion with bounded controls. IEEE Trans Autom Control AC-25(3):492–496
Shinar J, Medinah M, Biton M (1984) Singular surfaces in a linear pursuit-evasion game with elliptical vectograms. J Optim Theory Appl 43(3):431–456
Shinar J, Zarkh M (1996) Pursuit of a faster evader – a linear game with elliptical vectograms. In: Proceedings of the Seventh International Symposium on Dynamic Games, Yokosuka, pp 855–868
Simakova EN (1967) On one differential pursuit game. Autom Remote Control 2:173–181
Simakova EN (1968) Concerning certain problem of persuit on the plane. Autom Remote Control 29(7):1031–1034
Stipanović DM, Melikyan AA, Hovakimyan N (2010) Guaranteed strategies for nonlinear multi-player pursuit-evasion games. IGTR 12(1):1–17
Stipanović DM, Tomlin C, Leitmann G (2012) Monotone approximations of minimum and maximum functions and multi-objective problems. Appl Math Opt 66(3):455–473
Subbotin AI (1980) A generalization of the basic equation of the theory of differential games. Soviet Math Dokl 22:358–362
Subbotin AI (1984) Generalization of the main equation of differential game theory. JOTA 43(1):103–133
Subbotin AI (1995) Generalized solutions of first-order PDEs. The dynamical optimization perspective. Birkhäuser, Boston
Subbotin AI, Chentsov AG (1981) Optimization of the guarantee in control problems. Nauka, Moscow (in Russian)
Subbotin AI, Patsko VS (eds) (1984) Algorithms and programs for solving linear differential games. Ural Scientific Center, Institute of Mathematics and Mechanics, Academy of Sciences of USSR, Sverdlovsk (in Russian)
Taras'yev AM, Ushakov VN, Khripunov AP (1988) On a computational algorithm for solving game control problems. J Appl Math Mech 51(2):167–172
Turova VL Linear differential game of quality. In: Subbotin and Patsko (1984), pp 191–248 (in Russian)

Ushakov VN (1998) Construction of solutions in differential games of pursuit-evasion. In: Differential inclusions and optimal control. Lecture notes in nonlinear analysis, vol 2. Nicholas Copernicus University, Torun, pp 269–281

Yavin Y, Pachter M, Rodin EY (eds) (1987) Pursuit-Evasion differential games. International series in modern applied mathematics and computer science, vol 14. Pergamon Press, Oxford

Zheleznov VS, Ivanov MN, Kurskii EA, Maslov EP (1993) Avoidance of detection in 3D. Comput Math Appl 26(6):55–66

Biology and Evolutionary Games

23

Mark Broom and Vlastimil Křivan

Contents

M. Broom (✉)
Department of Mathematics, City, University London, London, UK
e-mail: Mark.Broom@city.ac.uk

V. Křivan
Biology Center, Czech Academy of Sciences, České Budějovice, Czech Republic

Faculty of Science, University of South Bohemia, České Budějovice, Czech Republic
e-mail: vlastimil.krivan@gmail.com

T. Başar, G. Zaccour (eds.), *Handbook of Dynamic Game Theory*,
https://doi.org/10.1007/978-3-319-44374-4_31

Abstract

This chapter surveys some evolutionary games used in biological sciences. These include the Hawk–Dove game, the Prisoner's Dilemma, Rock–Paper–Scissors, the war of attrition, the Habitat Selection game, predator–prey games, and signaling games.

Keywords

Battle of the Sexes · Foraging games · Habitat Selection game · Hawk–Dove game · Prisoner's Dilemma · Rock–Paper–Scissors · Signaling games · War of attrition

1 Introduction

Evolutionarily game theory (EGT) as conceived by Maynard Smith and Price (1973) was motivated by evolution. Several authors (e.g., Lorenz 1963; Wynne-Edwards 1962) at that time argued that animal behavior patterns were "for the good of the species" and that natural selection acts at the group level. This point of view was at odds with the Darwinian viewpoint where natural selection operates on the individual level. In particular, adaptive mechanisms that maximize a group benefit do not necessarily maximize individual benefit. This led Maynard Smith and Price (1973) to develop a mathematical model of animal behavior, called the Hawk–Dove game, that clearly shows the difference between group selection and individual selection. We thus start this chapter with the Hawk–Dove game.

Today, evolutionary game theory is one of the milestones of evolutionary ecology as it put the concept of Darwinian evolution on solid mathematical grounds. Evolutionary game theory has spread quickly in behavioral and evolutionary biology with many influential models that change the way that scientists look at evolution today. As evolutionary game theory is noncooperative, where each individual maximizes its own fitness, it seemed that it cannot explain cooperative or altruistic behavior that was easy to explain on the grounds of the group selection argument. Perhaps the most influential model in this respect is the Prisoner's Dilemma (Poundstone 1992), where the evolutionarily stable strategy leads to a collective payoff that is lower than the maximal payoff the two individuals can achieve if they were cooperating. Several models within evolutionary game theory have been developed that show how mutual cooperation can evolve. We discuss some of these models in Sect. 3. A popular game played by human players across the world, which can also be used to model some biological populations, is the Rock–Paper–Scissors game (RPS; Sect. 4). All of these games are single-species matrix games, so that their payoffs are linear, with a finite number of strategies. An example of a game that cannot be described by a matrix and that has a continuum of strategies is the war of attrition in Sect. 5.1 (or alternatively the Sir Philip Sidney game mentioned in Sect. 10). A game with nonlinear payoffs which examines an important biological phenomenon is the sex-ratio game in Sect. 5.2.

Although evolutionary game theory started with consideration of a single species, it was soon extended to two interacting species. This extension was not straightforward, because the crucial mechanism of a (single-species) EGT, that is, negative frequency dependence that stabilizes phenotype frequencies at an equilibrium, is missing if individuals of one species interact with individuals of another species. These games are asymmetric, because the two contestants are in different roles (such asymmetric games also occur within a single species). Such games that can be described by two matrices are called bimatrix games. Representative examples include the Battle of the Sexes (Sect. 6.1) and the Owner–Intruder game (Sect. 6.2). Animal spatial distribution that is evolutionarily stable is called the Ideal Free Distribution (Sect. 7). We discuss first the IFD for a single species and then for two species. The resulting model is described by four matrices, so it is no longer a bimatrix game. The IFD, as an outcome of animal dispersal, is related to the question of under which conditions animal dispersal can evolve (Sect. 8). Section 9 focuses on foraging games. We discuss two models that use EGT. The first model, that uses decision trees, is used to derive the diet selection model of optimal foraging. This model asks what the optimal diet of a generalist predator is in an environment that has two (or more) prey species. We show that this problem can be solved using the so-called agent normal form of an extensive game. We then consider a game between prey individuals that try to avoid their predators and predators that aim to capture prey individuals. The last game we consider in some detail is a signaling game of mate quality, which was developed to help explain the presence of costly ornaments, such as the peacock's tail.

We conclude with a brief section discussing a few other areas where evolutionary game theory has been applied. However, a large variety of models that use EGT have been developed in the literature, and it is virtually impossible to survey all of them.

2 The Hawk–Dove Game: Selection at the Individual Level vs. Selection at the Population Level

One of the first evolutionary games was introduced to understand evolution of aggressiveness among animals (Maynard Smith and Price 1973). Although many species have strong weapons (e.g., teeth or horns), it is a puzzling observation that in many cases antagonistic encounters do not result in a fight. In fact, such encounters often result in a complicated series of behaviors, but without causing serious injuries. For example, in contests between two male red deer, the contestants first approach each other, and provided one does not withdraw, the contest escalates to a roaring contest and then to the so-called parallel walk. Only if this does not lead to the withdrawal of one deer does a fight follow. It was observed (Maynard Smith

1982) that out of 50 encounters, only 14 resulted in a fight. The obvious question is why animals do not always end up in a fight? As it is good for an individual to get the resource (in the case of the deer, the resources are females for mating), Darwinian selection seems to suggest that individuals should fight whenever possible. One possible answer why this is not the case is that such a behavior is for the good of the species, because any species following this aggressive strategy would die out quickly. If so, then we should accept that the unit of selection is not an individual and abandon H. Spencer's "survival of the fittest" (Spencer 1864).

The Hawk–Dove model explains animal contest behavior from the Darwinian point of view. The model considers interactions between two individuals from the same population that meet in a pairwise contest. Each individual uses one of the two strategies called Hawk and Dove. An individual playing Hawk is ready to fight when meeting an opponent, while an individual playing Dove does not escalate the conflict. The game is characterized by two positive parameters where V denotes the value of the contested resource and C is the cost of the fight measured as the damage one individual can cause to his opponent. The payoffs for the row player describe the increase/decrease in the player's fitness after an encounter with an opponent. The game matrix is

$$\begin{array}{c} \\ \text{Hawk} \\ \text{Dove} \end{array} \begin{array}{c} \begin{array}{cc} \text{Hawk} & \text{Dove} \end{array} \\ \begin{pmatrix} \frac{V-C}{2} & V \\ 0 & \frac{V}{2} \end{pmatrix} \end{array}$$

and the model predicts that when the cost of the fight is lower than the reward obtained from getting the resource, $C < V$, all individuals should play the Hawk strategy that is the strict Nash equilibrium (NE) (thus an evolutionarily stable strategy (ESS)) of the game. When the cost of a fight is larger than the reward obtained from getting the resource, $C > V$, then $p = V/C$ $(0 < p < 1)$ is the corresponding monomorphic ESS. In other words, each individual will play Hawk when encountering an opponent with probability p and Dove with probability $1-p$. Thus, the model predicts that aggressiveness in the population decreases with the cost of fighting. In other words, the species that possess strong weapons (e.g., antlers in deer) should solve conflicts with very little fighting.

Can individuals obtain a higher fitness when using a different strategy? In a monomorphic population where all individuals use a mixed strategy $0 < p < 1$, the individual fitness and the average fitness in the population are the same and equal to

$$E(p, p) = \frac{V}{2} - \frac{C}{2} p^2.$$

This fitness is maximized for $p = 0$, i.e., when the level of aggressiveness in the population is zero, all individuals play the Dove strategy, and individual fitness equals $V/2$. Thus, if selection operated on a population or a species level, all individuals should be phenotypically Doves who never fight. However, the strategy

$p = 0$ cannot be an equilibrium from an evolutionary point of view, because in a Dove-only population, Hawks will always have a higher fitness (V) than Doves ($V/2$) and will invade. In other words, the Dove strategy is not resistant to invasion by Hawkish individuals. Thus, securing all individuals to play the strategy D, which is beneficial from the population point of view, requires some higher organizational level that promotes cooperation between animals (Dugatkin and Reeve 1998, see also Sect. 3).

On the contrary, at the evolutionarily stable equilibrium $p^* = V/C$, individual fitness

$$E(p^*, p^*) = \frac{V}{2}\left(1 - \frac{V}{C}\right)$$

is always lower than $V/2$. However, the ESS cannot be invaded by any other single mutant strategy.

Darwinism assumes that selection operates at the level of an individual, which is then consistent with noncooperative game theory. However, this is not the only possibility. Some biologists (e.g., Gilpin 1975) postulated that selection operates on a larger unit, a group (e.g., a population, a species etc.), maximizing the benefit of this unit. This approach was termed group selection. Alternatively, Dawkins (1976) suggested that selection operates on a gene level. The Hawk–Dove game allows us to show clearly the difference between the group and Darwinian selections.

Group selection vs. individual selection also nicely illustrates the so-called tragedy of the commons (Hardin 1968) (based on an example given by the English economist William Forster Lloyd) that predicts deterioration of the environment, measured by fitness, in an unconstrained situation where each individual maximizes its profit. For example, when a common resource (e.g., fish) is over-harvested, the whole fishery collapses. To maintain a sustainable yield, regulation is needed that prevents over-exploitation (i.e., which does not allow Hawks that would over-exploit the resource to enter). Effectively, such a regulatory body keeps p at zero (or close to it), to maximize the benefits for all fishermen. Without such a regulatory body, Hawks would invade and necessarily decrease the profit for all. In fact, as the cost C increases (due to scarcity of resources), fitness at the ESS decreases, and when C equals V, fitness is zero.

2.1 Replicator Dynamics for the Hawk–Dove Game

In the previous section, we have assumed that all individuals play the same strategy, either pure or mixed. If the strategy is mixed, each individual randomly chooses one of its elementary strategies on any given encounter according to some given probability. In this monomorphic interpretation of the game, the population mean strategy coincides with the individual strategy. Now we will consider a distinct situation where n phenotypes exist in the population. In this polymorphic setting we say that a population is in state $\mathbf{p} \in \Delta_n$ (where $\Delta_n = \{\mathbf{p} \in \mathbb{R}^n \mid p_i \geq 0, p_1 +$

$\cdots + p_n = 1\}$ is a probability simplex) if p_i is the proportion of the population using strategy i. As opposed to the monomorphic case, in this polymorphic interpretation, the individual strategies and the mean population strategy are different, because the mean strategy characterizes the population, not a single individual.

The ESS definition does not provide us with a mechanistic description of phenotype frequency dynamics that would converge to an ESS. One of the frequency dynamics often used in evolutionary game theory is the replicator dynamics (Taylor and Jonker 1978). Replicator dynamics assume that the population growth rate of each phenotype is given by its fitness, and they focus on changes in phenotypic frequencies in the population (see Volume I, ▸ Chap. 6, "Evolutionary Game Theory"). Let us consider the replicator equation for the Hawk–Dove game. Let x be the frequency of Hawks in the population. The fitness of a Hawk is

$$E(H, x) = \frac{V - C}{2} x + V(1 - x)$$

and, similarly, the fitness of a Dove is

$$E(D, x) = (1 - x)\frac{V}{2}.$$

Then the average fitness in the population is

$$E(x, x) = xE(H, x) + (1 - x)E(D, x) = \frac{V - Cx^2}{2},$$

and the replicator equation is

$$\frac{dx}{dt} = x\,(E(H, x) - E(x, x)) = \frac{1}{2}x(1 - x)(V - Cx).$$

Assuming $C > V$, we remark that the interior distribution equilibrium of this equation, $x = V/C$, corresponds to the mixed ESS for the underlying game. In this example phenotypes correspond to elementary strategies of the game. It may be that phenotypes also correspond to mixed strategies.

3 The Prisoner's Dilemma and the Evolution of Cooperation

The Prisoner's Dilemma (see Flood 1952; Poundstone 1992) is perhaps the most famous game in all of game theory, with applications from areas including economics, biology, and psychology. Two players play a game where they can Defect or Cooperate, yielding the payoff matrix

$$\begin{array}{c} \\ \text{Cooperate} \\ \text{Defect} \end{array} \begin{array}{c} \text{Cooperate} \quad \text{Defect} \\ \begin{pmatrix} R & S \\ T & P \end{pmatrix} \end{array}.$$

These abbreviations are derived from Reward (reward for cooperating), Temptation (temptation for defecting when the other player cooperates), Sucker (paying the cost of cooperation when the other player defects), and Punishment (paying the cost of defecting). The rewards satisfy the conditions $T > R > P > S$. Thus while Cooperate is Pareto efficient (in the sense that it is impossible to make any of the two players better off without making the other player worse off), Defect row dominates Cooperate and so is the unique ESS, even though mutual cooperation would yield the greater payoff. Real human (and animal) populations, however, involve a lot of cooperation; how is that enforced?

There are many mechanisms for enabling cooperation, see for example Nowak (2006). These can be divided into six types as follows:

1. Kin selection, that occurs when the donor and recipient of some apparently altruistic act are genetic relatives.
2. Direct reciprocity, requiring repeated encounters between two individuals.
3. Indirect reciprocity, based upon reputation. An altruistic individual gains a good reputation, which means in turn that others are more willing to help that individual.
4. Punishment, as a way to enforce cooperation.
5. Network reciprocity, where there is not random mixing in the population and cooperators are more likely to interact with other cooperators.
6. Multi-level selection, alternatively called group selection, where evolution occurs on more than one level.

We discuss some of these concepts below.

3.1 Direct Reciprocity

Direct reciprocity requires repeated interaction and can be modeled by the Iterated Prisoner's Dilemma (IPD). The IPD involves playing the Prisoner's Dilemma over a (usually large) number of rounds and thus being able to condition choices in later rounds on what the other player played before. This game was popularized by Axelrod (1981, 1984) who held two tournaments where individuals could submit computer programs to play the IPD. The winner of both tournaments was the simplest program submitted, called Tit for Tat (TFT), which simply cooperates on the first move and then copies its opponent's previous move.

TFT here has three important properties: it is *nice* so it never defects first; it is *retaliatory* so it meets defection with a defection next move; it is *forgiving* so even after previous defections, it meets cooperation with cooperation next move. TFT effectively has a memory of one place, and it was shown in Axelrod and Hamilton (1981) that TFT can resist invasion by any strategy that is not nice if it can resist both Always Defect ALLD and Alternative ALT, which defects (cooperates) on odd (even) moves. However, this does not mean that TFT is an ESS, because nice strategies can invade by drift as they receive identical payoffs to TFT in a TFT population (Bendorf and Swistak 1995). We note that TFT is not the only strategy that can promote cooperation in the IPD; others include Tit for Two Tats (TF2T which defects only after two successive defections of its opponent), Grim (which defects on all moves after its opponent's first defection), and win stay/lose shift (which changes its choice if and only if its opponent defected on the previous move).

Games between TFT, ALLD, and ALT against TFT have the following sequence of moves:

$$\begin{array}{c|l} TFT & CCCCCC\ldots \\ TFT & CCCCCC\ldots \\ \hline ALLD & DDDDDD\ldots \\ TFT & CDDDDD\ldots \\ \hline ALT & DCDCDC\ldots \\ TFT & CDCDCD\ldots \end{array} \tag{23.1}$$

We assume that the number of rounds is not fixed and that there is always the possibility of a later round (otherwise the game can be solved by backwards induction, yielding ALLD as the unique NE strategy). At each stage, there is a further round with probability w (as in the second computer tournament); the payoffs are then

$$E(TFT, TFT) = R + Rw + Rw^2 + \ldots = \frac{R}{1-w}, \tag{23.2}$$

$$E(ALLD, TFT) = T + Pw + Pw^2 + \ldots = T + \frac{Pw}{1-w}, \tag{23.3}$$

$$E(ALT, TFT) = T + Sw + Tw^2 + Sw^3 + \ldots = \frac{T + Sw}{1-w^2}. \tag{23.4}$$

Thus TFT resists invasion if and only if

$$\frac{R}{1-w} > \max\left(T + \frac{Pw}{1-w}, \frac{T+Sw}{1-w^2}\right)$$

i.e., if and only if

$$w > \max\left(\frac{T-R}{T-P}, \frac{T-R}{R-S}\right), \tag{23.5}$$

i.e., when the probability of another contest is sufficiently large (Axelrod 1981, 1984). We thus see that for cooperation to evolve here, the extra condition $2R >$

$S + T$ is required, since otherwise the right-hand side of inequality (23.5) would be at least 1.

While TFT proved successful at promoting cooperation above, what if errors occur, so that an intention to cooperate becomes a defection (or is perceived as such)? After a single mistake, a pair of interacting TFT players will be locked in an alternating cycle of Defect versus Cooperation and then mutual defection after a second mistake when C was intended. Under such circumstances, TF2T can maintain cooperation, whereas TFT cannot. In fact a better strategy (in the sense that it maintains cooperation when playing against itself but resists invasion from defecting strategies) is GTFT (generous tit for tat; see Komorita et al. 1968), which combines pure cooperation with TFT by cooperating after a cooperation, but meeting a defection with a defection with probability

$$\min\left(1 - \frac{T-R}{R-S}, \frac{R-P}{T-P}\right).$$

3.2 Kin Selection and Hamilton's Rule

In most evolutionary game theoretical models, the aim of each individual is to maximize its own fitness, irrespective of the fitness of other individuals. However, if individuals are related, then the fitnesses of others should be taken into account.

Let us consider two interacting individuals, with coefficient of relatedness r, which is the probability that they share a copy of a given allele. For example, father and son will have $r = 1/2$. One individual acts as a potential donor, the other as a recipient, which receives a benefit b from the donor at the donor's cost c. The donating individual pays the full cost but also indirectly receives the benefit b multiplied by the above factor r. Thus donation is worthwhile provided that

$$rb > c \text{ i.e., } r > \frac{c}{b}$$

which is known as *Hamilton's rule* (Hamilton 1964).

Note that this condition is analogous to the condition for cooperation to resist invasion in the IPD above, where a commonly used special class of the PD matrix has payoffs representing cooperation as making a donation and defecting as not. Then TFT resists invasion when $w > c/b$.

3.3 Indirect Reciprocity and Punishment

The IPD is an example of direct reciprocity. Suppose now we have a population of individuals who play many contests, but these are not in long sequences against a single "opponent" as above? If faced with a series of single-shot games, how can cooperation be achieved?

Such situations are often investigated by the use of public goods games involving experiments with groups of real people, as in the work of Fehr and Gachter (2002). In these experiments individuals play a series of games, each game involving a new group. In each game there were four individuals, each of them receiving an initial endowment of 20 dollars, and each had to choose a level of investment into a common pool. Any money that was invested increased by a factor of 1.6 and was then shared between the four individuals, meaning that the return for each dollar invested was 40 cents to each of the players. In particular the individual making the investment of one dollar only receives 40 cents and so makes a loss of 60 cents. Thus, like the Prisoner's Dilemma, it is clear that the best strategy is to make no investment but simply to share rewards from the investments of other players. In these experiments, investment levels began reasonably high, but slowly declined, as players saw others cheat.

In later experiments, each game was played over two rounds, an investment round and a punishment round, where players were allowed to punish others. In particular every dollar "invested" in punishment levied a fine of three dollars on the target of the punishment. This led to investments which increased from their initial level, as punishment brought cheating individuals into line. It should be noted that in a population of individuals many, but not all of whom, punish, optimal play for individuals in this case should not be to punish, but to be a second-order free rider who invests but does not punish, and therefore saves the punishment fee. Such a population would collapse down to no investment after some number of rounds. Thus it is clear that the people in the experiments were not behaving completely rationally.

Thus we could develop the game to have repeated rounds of punishment. An aggressive punishing strategy would then in round 1, punish all defectors; in round 2, punish all cooperators who did not punish defectors in round 1; in round 3, punish all cooperators who did not punish in round 2 as above; and so on. Thus such players not only punish cheats, but anyone who does not play exactly as they do. Imagine a group of m individuals with k cooperators (who invest and punish), ℓ defectors and $m-k-\ell-1$ investors (who do not punish). This game, with this available set of strategies, requires two rounds of punishment as described above. The rewards to our focal individual in this case will be

$$R = \begin{cases} \frac{(m-\ell)cV}{m} - kP & \text{if an investor,} \\ \frac{(m-\ell)cV}{m} - (m-k-1) & \text{if a cooperator,} \\ \frac{(m-\ell-1)cV}{m} + V - kP & \text{if a defector,} \end{cases}$$

where V is the initial level of resources of each individual, $c < m$ is the return on investment (every dollar becomes $1+c$ dollars), and P is the punishment multiple (every dollar invested in punishment generates a fine of P dollars). The optimal play for our focal individual is

$$\begin{array}{ll} \text{Defect} & \text{if } V\left(1-\frac{c}{m}\right) > kP - (m-k-1), \\ \text{Cooperate} & \text{otherwise.} \end{array}$$

Thus defect is always stable and invest and punish is stable if $V(1 - c/m) < (m - 1)P$.

We note that there are still issues on how such punishment can emerge in the first place (Sigmund 2007).

4 The Rock–Paper–Scissors Game

The Rock–Paper–Scissors game is a three-strategy matrix game, which people commonly play recreationally. In human competition, the game dates back at least to seventeenth-century China. There is a lot of potential psychology involved in playing the game, and there are numerous tournaments involving it. The important feature of the game is that Rock beats Scissors, Scissors beats Paper, and Paper beats Rock. The payoff matrix is

$$\begin{array}{c} \\ \text{Rock} \\ \text{Scissors} \\ \text{Paper} \end{array} \begin{array}{c} \begin{array}{ccc} \text{Rock} & \text{Scissors} & \text{Paper} \end{array} \\ \begin{pmatrix} 0 & a_3 & -b_2 \\ -b_3 & 0 & a_1 \\ a_2 & -b_1 & 0 \end{pmatrix} \end{array},$$

where all a's and b's are positive. For the conventional game played between people $a_i = b_i = 1$ for $i = 1, 2, 3$.

There is a unique internal NE of the above game given by the vector

$$\mathbf{p} = \frac{1}{K}(a_1a_3 + b_1b_2 + a_1b_1, a_1a_2 + b_2b_3 + a_2b_2, a_2a_3 + b_1b_3 + a_3b_3),$$

where the constant K is just the sum of the three terms to ensure that $\mathbf{p}$ is a probability vector. In addition, $\mathbf{p}$ is a globally asymptotically stable equilibrium of the replicator dynamics if and only if $a_1a_2a_3 > b_1b_2b_3$. It is an ESS if and only if $a_1 - b_1, a_2 - b_2$, and $a_3 - b_3$ are all positive, and the largest of their square roots is smaller than the sum of the other two square roots (Hofbauer and Sigmund 1998). Thus if $\mathbf{p}$ is an ESS of the RPS game, then it is globally asymptotically stable under the replicator dynamics. However, since the converse is not true, the RPS game provides an example illustrating that while all internal ESSs are global attractors of the replicator dynamics, not all global attractors are ESSs.

We note that the case when $a_1a_2a_3 = b_1b_2b_3$ (including the conventional game with $a_i = b_i = 1$) leads to closed orbits of the replicator dynamics, and a stable (but not asymptotically stable) internal equilibrium. This is an example of a *nongeneric* game, where minor perturbations of the parameter values can lead to large changes in the nature of the game solution.

This game is a good representation for a number of real populations. The most well known of these is among the common side-blotched lizard *Uta stansburiana*. This lizard has three types of distinctive throat coloration, which correspond to very different types of behavior. Males with orange throats are very aggressive and have large territories which they defend against intruders. Males with dark blue throats are less aggressive and hold smaller territories. Males with yellow stripes do not have a territory at all but bear a strong resemblance to females and use a sneaky mating strategy. It was observed in Sinervo and Lively (1996) that if the Blue strategy is the most prevalent, Orange can invade; if Yellow is prevalent, Blue can invade; and if Orange is prevalent, then Yellow can invade.

An alternative real scenario is that of *Escherichia coli* bacteria, involving three strains of bacteria (Kerr et al. 2002). One strain produces the antibiotic colicin. This strain is immune to it, as is a second strain, but the third is not. When only the first two strains are present, the second strain outcompetes the first, since it forgoes the cost of colicin production. Similarly the third outcompetes the second, as it forgoes costly immunity, which without the first strain is unnecessary. Finally, the first strain outcompetes the third, as the latter has no immunity to the colicin.

5 Non-matrix Games

We have seen that matrix games involve a finite number of strategies with a payoff function that is linear in the strategy of both the focal player and that of the population. This leads to a number of important simplifying results (see Volume I, ▶ Chap. 6, "Evolutionary Game Theory"). All of the ESSs of a matrix can be found in a straightforward way using the procedure of Haigh (1975). Further, adding a constant to all entries in a column of a payoff matrix leaves the collection of ESSs (and the trajectories of the replicator dynamics) of the matrix unchanged. Haigh's procedure can potentially be shortened, using the important Bishop–Cannings theorem (Bishop and Cannings 1976), a consequence of which is that if $\mathbf{p_1}$ is an ESS, no strategy $\mathbf{p_2}$ whose support is either a superset or a subset of the support of $\mathbf{p_1}$ can be an ESS.

However, there are a number of ways that games can involve nonlinear payoff functions. Firstly, playing the field games yield payoffs that are linear in the focal player but not in the population (e.g., see Sects. 7.1 and 9.1). Another way this can happen is to have individual games of the matrix type, but where opponents are not selected with equal probability from the population, for instance, if there is some spatial element. Thirdly, the payoffs can be nonlinear in both components. Here strategies do not refer to a probabilistic mix of pure strategies, but a unique trait, such as the height of a tree as in Kokko (2007) or a volume of sperm; see, e.g., Ball and Parker (2007). This happens in particular in the context of adaptive dynamics (see Volume I, ▶ Chap. 6, "Evolutionary Game Theory").

Alternatively a non-matrix game can involve linear payoffs, but this time with a continuum of strategies (we note that the cases with nonlinear payoffs above can also involve such a continuum, especially the third type). A classical example of this is the war of attrition (Maynard Smith 1974).

5.1 The War of Attrition

We consider a Hawk–Dove game, where both individuals play Dove, but that instead of the reward being allocated instantly, they become involved in a potentially long displaying contest where a winner will be decided by one player conceding, and there is a cost proportional to the length of the contest. An individual's strategy is thus the length of time it is prepared to wait. Pure strategies are all values of t on the non-negative part of the real line, and mixed strategies are corresponding probability distributions. These kinds of contests are for example observed in dung flies (Parker and Thompson 1980).

Choosing the cost to be simply the length of time spent, the payoff for a game between two pure strategies S_t (wait until time t) and S_s (wait until time s) for the player that uses strategy S_t is

$$E(S_t, S_s) = \begin{cases} V - s & t > s, \\ V/2 - t & t = s, \\ -t & t < s. \end{cases}$$

and the corresponding payoff from a game involving two mixed strategists playing the probability distributions $f(t)$ and $g(s)$ to the $f(t)$ player is

$$\int_0^\infty \int_0^\infty f(t)g(s)E(S_t, S_s)dtds.$$

It is clear that no pure strategy can be an ESS, since S_s is invaded by S_t (i.e., $E(S_t, S_s) > E(S_s, S_s)$) for any $t > s$, or any positive $t < s - V/2$. There is a unique ESS which is found by first considering (analogous to the Bishop–Cannings theorem; see Volume I, ▸ Chap. 6, "Evolutionary Game Theory") a probability distribution $p(s)$ that gives equal payoffs to all pure strategies that could be played by an opponent. This is required, since otherwise some potential invading strategies could do better than others, and since $p(s)$ is simply a weighted average of such strategies, it would then be invaded by at least one type of opponent. Payoff of a pure strategy S_t played against a mixed strategy $S_{p(s)}$ given by a probability distribution $p(s)$ over the time interval is

$$E(S_t, S_{p(s)}) = \int_0^t (V - s)p(s)ds + \int_t^\infty (-t)p(s)ds. \tag{23.6}$$

Differentiating equation (23.6) with respect to t (assuming that such a derivative exists) gives

$$(V - t)p(t) - \int_t^\infty p(s)ds + tp(t) = 0. \tag{23.7}$$

If $P(t)$ is the associated distribution function, so that $p(t) = P'(t)$ for all $t \geq 0$, then Eq. (23.7) becomes

$$VP'(t) + P(t) - 1 = 0$$

and we obtain

$$p(t) = \frac{1}{V} \exp\left(-\frac{t}{V}\right). \tag{23.8}$$

It should be noted that we have glossed over certain issues in the above, for example, consideration of strategies without full support or with atoms of probability. This is discussed in more detail in Broom and Rychtar (2013). The above solution was shown to be an ESS in Bishop and Cannings (1976).

5.2 The Sex-Ratio Game

Why is it that the sex ratio in most animals is close to a half? This was the first problem to be considered using evolutionary game theory (Hamilton 1967), and its consideration, including the essential nature of the solution, dates right back to Darwin (1871). To maximize the overall birth rate of the species, in most animals there should be far more females than males, given that females usually make a much more significant investment in bringing up offspring than males. This, as mentioned before, is the wrong perspective, and we need to consider the problem from the viewpoint of the individual.

Suppose that in a given population, an individual female will have a fixed number of offspring, but that she can allocate the proportion of these that are male. This proportion is thus the strategy of our individual. As each female (irrespective of its strategy) has the same number of offspring, this number does not help us in deciding which strategy is the best. The effect of a given strategy can be measured as the number of grandchildren of the focal female. Assume that the number of individuals in a large population in the next generation is N_1 and in the following generation is N_2. Further assume that all other females in the population play the strategy m and that our focal individual plays strategy p.

As N_1 is large, the total number of males in the next generation is mN_1 and so the total number of females is $(1 - m)N_1$. We shall assume that all females (males) are equally likely to be the mother (father) of any particular member of the following generation of N_2 individuals. This means that a female offspring will be the mother of $N_2/((1 - m)N_1)$ of the following generation of N_2 individuals, and a male offspring will be the father of $N_2/(mN_1)$ of these individuals. Thus our focal individual will have the following number of grandchildren

$$E(p, m) = p\frac{N_2}{mN_1} + (1 - p)\frac{N_2}{(1 - m)N_1} = \frac{N_2}{N_1}\left(\frac{p}{m} + \frac{1 - p}{1 - m}\right). \tag{23.9}$$

To find the best p, we maximize $E(p, m)$. For $m < 0.5$ the best response is $p = 1$, and for $m > 0.5$ we obtain $p = 0$. Thus $m = 0.5$ is the interior NE at which all values of p obtain the same payoff. This NE satisfies the stability condition in that $E(0.5, m') - E(m', m') > 0$ for all $m' \neq 0.5$ (Broom and Rychtar 2013).

Thus from the individual perspective, it is best to have half your offspring as male. In real populations, it is often the case that relatively few males are the parents of many individuals, for instance, in social groups often only the dominant male fathers offspring. Sometimes other males are actually excluded from the group; lion prides generally consist of a number of females, but only one or two males, for example. From a group perspective, these extra males perform no function, but there is a chance that any male will become the father of many.

6 Asymmetric Games

The games we have considered above all involve populations of identical individuals. What if individuals are not identical? Maynard Smith and Parker (1976) considered two main types of difference between individuals. The first type was *correlated asymmetries* where there were real differences between them, for instance, in strength or need for resources, which would mean their probability of success, cost levels, valuation of rewards, set of available strategies, etc., may be different, i.e., the payoffs "correlate" with the type of the player. Examples of such games are the predator–prey games of Sect. 9.2 and the Battle of the Sexes below in Sect. 6.1.

The second type, *uncorrelated asymmetries*, occurred when the individuals were physically identical, but nevertheless occupied different roles; for example, one was the owner of a territory and the other was an intruder, which we shall see in Sect. 6.2. For uncorrelated asymmetries, even though individuals do not have different payoff matrices, it is possible to base their strategy upon the role that they occupy. As we shall see, this completely changes the character of the solutions that we obtain.

We note that the allocation of distinct roles can apply to games in general; for example, there has been significant work on the asymmetric war of attrition (see, e.g., Hammerstein and Parker 1982; Maynard Smith and Parker 1976), involving cases with both correlated and uncorrelated asymmetries.

The ESS was defined for a single population only, and the stability condition of the original definition cannot be easily extended for bimatrix games. This is because bimatrix games assume that individuals of one species interact with individuals of the other species only, so there is no frequency-dependent mechanism that could prevent mutants of one population from invading residents of that population at the two-species NE. In fact, it was shown (Selten 1980) that requiring the stability condition of the ESS definition to hold in bimatrix games restricts the ESSs to strict NEs, i.e., to pairs of pure strategies. Two key assumptions behind Selten's theorem are that the probability that an individual occupies a given role is not affected by the strategies that it employs, and that payoffs within a given role are linear, as in matrix games. If either of these assumptions are violated, then mixed strategy ESSs can result (see, e.g., Broom and Rychtar 2013; Webb et al. 1999).

There are interior NE in bimatrix games that deserve to be called "stable," albeit in a weaker sense than was used in the (single-species) ESS definition. For example, some of the NEs are stable with respect to some evolutionary dynamics (e.g., with respect to the replicator dynamics, or the best response dynamics). A static concept that captures such stability that proved useful for bimatrix games is the Nash–Pareto equilibrium (Hofbauer and Sigmund 1998). The Nash-Pareto equilibrium is an NE which satisfies an additional condition that says that it is impossible for both players to increase their fitness by deviating from this equilibrium. For two-species games that cannot be described by a bimatrix (e.g., see Sect. 7.4), this concept of two-species evolutionary stability was generalized by Cressman (2003) (see Volume I, ▸ Chap. 6, "Evolutionary Game Theory") who defined a two-species ESS (p^*, q^*) as an NE such that, if the population distributions of the two species are slightly perturbed, then an individual in at least one species does better by playing its ESS strategy than by playing the slightly perturbed strategy of this species. We illustrate these concepts in the next section.

6.1 The Battle of the Sexes

A classical example of an asymmetric game is the *Battle of the Sexes* (Dawkins 1976), where a population contains females with two strategies, Coy and Fast, and males with two strategies, Faithful and Philanderer. A Coy female needs a period of courtship, whereas a Fast female will mate with a male as soon as they meet. Faithful males are willing to engage in long courtships and after mating will care for the offspring. A Philanderer male will not engage in courtship and so cannot mate with a Coy female and also leaves immediately after mating with a Fast female.

Clearly in this case, any particular individual always occupies a given role (i.e., male or female) and cannot switch roles as is the case in the Owner–Intruder game in Sect. 6.2 below. Thus, males and females each have their own payoff matrix which are often represented as a bimatrix. The payoff bimatrix for the Battle of the Sexes is

$$\begin{array}{lcc} \text{Male}\backslash\text{Female} & \text{Coy} & \text{Fast} \\ \begin{array}{l}\text{Faithful}\\ \text{Philanderer}\end{array} & \multicolumn{2}{c}{\left(\begin{array}{cc} (B - \frac{C_R}{2} - C_C, B - \frac{C_R}{2} - C_C) & (B - \frac{C_R}{2}, B - \frac{C_R}{2}) \\ (0,0) & (B, B - C_R) \end{array} \right).} \end{array} \tag{23.10}$$

Here B is the fitness gained by having an offspring, C_R is the (potentially shared) cost of raising the offspring, and C_C is the cost of engaging in a courtship. All three of these terms are clearly positive. The above bimatrix is written in the form (A_1, A_2^T), where matrix A_1 is the payoff matrix for males (player 1) and matrix A_2 is the payoff matrix for females (player 2), respectively.

For such games, to define a two-species NE, we study the position of the two equal payoff lines, one for each sex. The equal payoff line for males (see the horizontal dashed line in Fig. 23.1) is defined to be those $(\mathbf{p}, \mathbf{q}) \in \Delta_2 \times \Delta_2$ for which the payoff when playing Faithful equals the payoff when playing the Philanderer strategy, i.e.,

$$(1,0)A_1\mathbf{q}^T = (0,1)A_1\mathbf{q}^T$$

which yields $q_1 = \frac{C_R}{2(B-C_C)}$. Similarly, along the equal payoff line for females (see the vertical dashed line in Fig. 23.1), the payoff when playing strategy Coy must equal the payoff when playing strategy Fast, i.e.,

$$(1,0)A_2\mathbf{p}^T = (0,1)A_2\mathbf{p}^T.$$

If the two equal payoff lines do not intersect in the unit square, no completely mixed strategy (both for males and females) is an NE (Fig. 23.1A, B). In fact, there is a unique ESS (Philanderer, Coy), i.e., with no mating (clearly not appropriate for a real population), for sufficiently small B (Fig. 23.1A), $B < \min(C_R/2 + C_C, C_R)$, a unique ESS (Philanderer, Fast) for sufficiently high B (Fig. 23.1B), when $B > C_R$. For intermediate B satisfying $C_R/2 + C_C < B < C_R$, there is a two-species weak ESS

$$\mathbf{p} = \left(\frac{C_R - B}{C_C + C_R - B}, \frac{C_C}{C_C + C_R - B}\right), \quad \mathbf{q} = \left(\frac{C_R}{2(B - C_C)}, 1 - \frac{C_R}{2(B - C_C)}\right),$$

where at least the fitness of one species increases toward this equilibrium, except when $p_1 = \frac{C_R - B}{C_C + C_R - B}$ or $q_1 = \frac{C_R}{2(B-C_C)}$ (Fig. 23.1C). In all three cases of

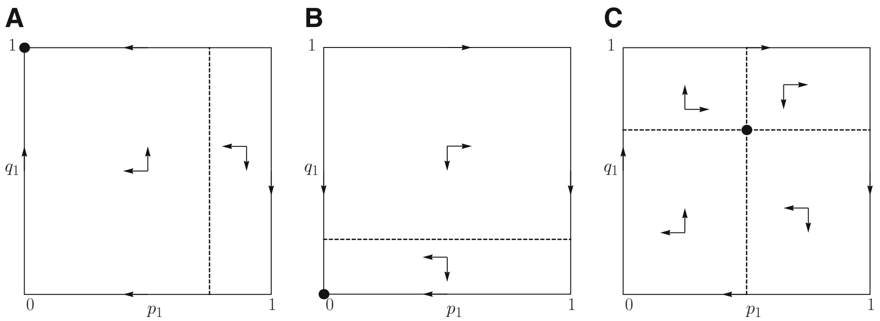


Fig. 23.1 The ESS for the Battle of the Sexes game. Panel A assumes small B and the only ESS is $(p_1, q_1) = (0, 1) =$ (Philanderer, Coy). Panel B assumes large B and the only ESS is $(p_1, q_1) = (0, 0) =$ (Philanderer, Fast). For intermediate values of B (panel C), there is an interior NE. The *dashed lines* are the two equal payoff lines for males (*horizontal line*) and females (*vertical line*). The direction in which the male and female payoffs increase are shown by *arrows* (e.g., a *horizontal arrow* to the right means the first strategy (Faithful) has the higher payoff for males, whereas a *downward arrow* means the second strategy (Fast) has the higher payoff for females). We observe that in panel C these arrows are such that at least the payoff of one player increases toward the Nash–Pareto pair, with the exception of the points that lie on the equal payoff lines. This qualifies the interior NE as a two-species weak ESS

Fig. 23.1, the NE is a Nash-Pareto pair, because it is impossible for both players to simultaneously deviate from the Nash equilibrium and increase their payoffs. In panels A and B, both arrows point in the direction of the NE. In panel C at least one arrow is pointing to the NE $(\mathbf{p}, \mathbf{q})$ if both players deviate from that equilibrium. However, this interior NE is not a two-species ESS since, when only one player (e.g., player one) deviates, no arrow points in the direction of $(\mathbf{p}, \mathbf{q})$. This happens on the equal payoff lines (dashed lines). For example, let us consider points on the vertical dashed line above the NE. Here vertical arrows are zero vectors and horizontal arrows point away from $\mathbf{p}$. Excluding points on the vertical and horizontal line from the definition of a two-species ESS leads to a two-species weak ESS.

6.2 The Owner–Intruder Game

The Owner–Intruder game is an extension of the Hawk–Dove game, where player 1 (the owner) and player 2 (the intruder) have distinct roles (i.e., they cannot be interchanged as is the case of symmetric games). In particular an individual can play either of Hawk or Dove in either of the two roles. This leads to the bimatrix representation of the Hawk–Dove game below, which cannot be collapsed down to the single 2×2 matrix from Sect. 2, because the strategy that an individual plays may be conditional upon the role that it occupies (in Sect. 2 there are no such distinct roles). The bimatrix of the game is

$$\begin{array}{lcc} \text{Owner}\backslash\text{Intruder} & \text{Hawk} & \text{Dove} \\ \text{Hawk} & \multirow{2}{*}{} \\ \text{Dove} \end{array} \begin{pmatrix} (\frac{V-C}{2}, \frac{V-C}{2}) & (V, 0) \\ (0, V) & (V/2, V/2) \end{pmatrix}.$$

Provided we assume that each individual has the same chance to be an owner or an intruder, the game can be symmetrized with the payoffs to the symmetrized game given in the following payoff matrix,

$$\begin{array}{l} \\ \text{Hawk} \\ \text{Dove} \\ \text{Bourgeois} \\ \text{Anti-Bourgeois} \end{array} \begin{array}{c} \begin{array}{cccc} \text{Hawk} & \text{Dove} & \text{Bourgeois} & \text{Anti-Bourgeois} \end{array} \\ \begin{pmatrix} (V-C)/2 & V & (3V-C)/4 & (3V-C)/4 \\ 0 & V/2 & V/4 & V/4 \\ (V-C)/4 & 3V/4 & V/2 & (2V-C)/4 \\ (V-C)/4 & 3V/4 & (2V-C)/4 & V/2 \end{pmatrix} \end{array}$$

where

Hawk	–	play Hawk when both owner and intruder,
Dove	–	play Dove when both owner and intruder,
Bourgeois	–	play Hawk when owner and Dove when intruder,
Anti-Bourgeois	–	play Dove when owner and Hawk when intruder.

It is straightforward to show that if $V \geq C$, then Hawk is the unique ESS (Fig. 23.2A), and that if $V < C$, then Bourgeois and Anti-Bourgeois (alternatively

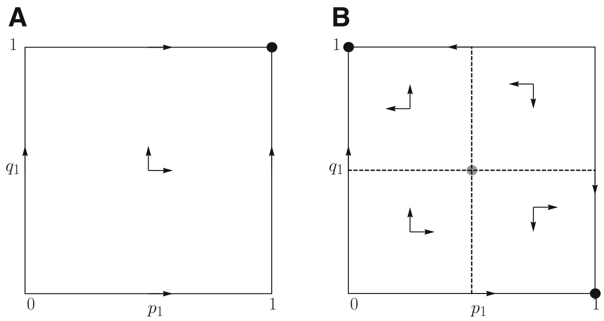


Fig. 23.2 The ESS for the Owner–Intruder game. Panel A assumes $V > C$ and the only ESS is to be Hawk at both roles (i.e., $(p_1, q_1) = (1, 1) = (\text{Hawk}, \text{Hawk})$). If $V < C$ (Panel B), there are two boundary ESSs (*black dots*) corresponding to the Bourgeois ($(p_1, q_1) = (1, 0) = (\text{Hawk}, \text{Dove})$) and Anti-Bourgeois ($(p_1, q_1) = (0, 1) = (\text{Dove}, \text{Hawk})$) strategies. The directions in which the owner and intruder payoffs increase are shown by *arrows* (e.g., a *horizontal arrow* to the right means the Hawk strategy has the higher payoff for owner, whereas a *downward arrow* means the Dove strategy has the higher payoff for intruder). The interior NE (the *light gray dot* at the intersection of the two equal payoff lines) is not a two-species ESS as there are regions (the *upper-left* and *lower-right corners*) where both arrows point in directions away from this point

called Marauder) are the only ESSs (Fig. 23.2B). As we see, there are only pure strategy solutions, as opposed to the case of the Hawk–Dove game, which had a mixed ESS V/C for $V < C$, because the interior NE in Fig. 23.2B is not a Nash–Pareto pair as in the upper-left and lower-right regions both arrows are pointing away from the NE. Thus, if both players simultaneously deviate from the Nash equilibrium, their payoffs increase. Consequently, this NE is not a two-species (weak) ESS.

Important recent work on this model and its ramifications for the part that respecting ownership plays has been carried out by Mesterton Gibbons, Sherratt and coworkers (see Mesterton-Gibbons and Sherratt 2014; Sherratt and Mesterton-Gibbons 2015). In particular, why in real populations is the Bourgeois respect for ownership strategy so common and the Anti-Bourgeois strategy so rare? One explanation offered by Maynard Smith (1982) was "infinite regress." In this argument, immediately after a contest, the winner becomes the owner of the territory, and the loser becomes a potential intruder which could immediately rechallenge the individual that has just displaced it. In an Anti-Bourgeois population, this would result in the new owner conceding and the new intruder (the previous owner) once again being the owner, but then the displaced owner could immediately rechallenge, and the process would continue indefinitely. It is shown in Mesterton-Gibbons and Sherratt (2014) that under certain circumstances, but not always, this allows Bourgeois to be the unique ESS. Sherratt and Mesterton-Gibbons (2015) discuss many issues, such as uncertainty of ownership, asymmetry of resource value, continuous contest investment (as in the war of attrition), and potential signaling of intentions (what they call "secret handshakes," similar to some of the signals we

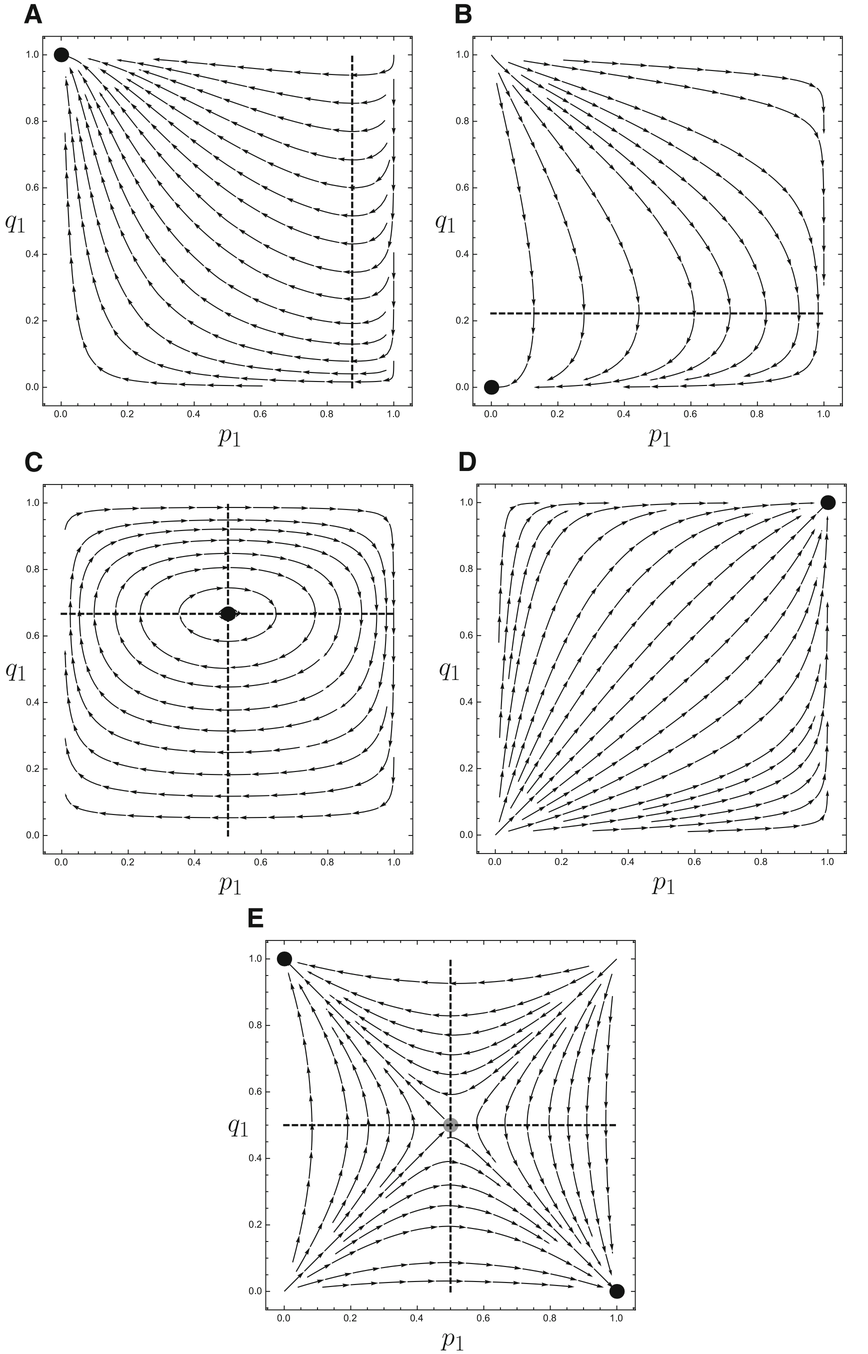


Fig. 23.3 (Continued)

discuss in Sect. 10) in detail. There are many reasons that can make evolution of Anti-Bourgeois unlikely, and it is probably a combination of these that make it so rare.

6.3 Bimatrix Replicator Dynamics

The single-species replicator dynamics such as those for the Hawk–Dove game (Sect. 2.1) can be extended to two roles as follows (Hofbauer and Sigmund 1998). Note that here this is interpreted as two completely separate populations, i.e., any individual can only ever occupy one of the roles, and its offspring occupy that same role. If $A = (a_{ij})_{\substack{i=1,\ldots,n \\ j=1,\ldots,m}}$ and $B = (b_{ij})_{\substack{i=1,\ldots,m \\ j=1,\ldots,n}}$ are the payoff matrices to an individual in role 1 and role 2, respectively, the corresponding replicator dynamics are

$$\tfrac{d}{dt} p_{1i}(t) = p_{1i}\big((A\mathbf{p_2}^T)_i - \mathbf{p_1}A\mathbf{p_2}^T\big) \quad i = 1, \ldots, n;$$

$$\tfrac{d}{dt} p_{2j}(t) = p_{2j}\big((B\mathbf{p_1}^T)_j - \mathbf{p_2}B\mathbf{p_1}^T\big) \quad j = 1, \ldots, m;$$

where $\mathbf{p_1} \in \Delta_n$ and $\mathbf{p_2} \in \Delta_m$ are the population mixtures of individuals in role 1 and 2, respectively. For example, for the two-role, two-strategy game, where without loss of generality we can set $a_{11} = a_{22} = b_{11} = b_{22} = 0$ (since as for matrix games, adding a constant to all of the payoffs an individual gets against a given strategy does not affect the NEs/ ESSs), we obtain

$$\begin{aligned} \frac{dx}{dt} &= x(1-x)(a_{12} - (a_{12} + a_{21})y), \\ \frac{dy}{dt} &= y(1-y)(b_{12} - (b_{12} + b_{21})x), \end{aligned} \tag{23.11}$$

where x is the frequency of the first strategy players in the role 1 population and y is the corresponding frequency for role 2. Hofbauer and Sigmund (1998) show that orbits converge to the boundary in all cases except if $a_{12}a_{21} > 0$, $b_{12}b_{21} > 0$, and $a_{12}b_{12} < 0$, which yield closed periodic orbits around the internal equilibrium. Replicator dynamics for the Battle of the Sexes and the Owner–Intruder game are shown in Fig. 23.3.

Fig. 23.3 Bimatrix replicator dynamics (23.11) for the Battle of the Sexes game (A–C) and the Owner–Intruder game (D, E), respectively. Panels A–C correspond to panels given in Fig. 23.1, and panels D and E correspond to those of Fig. 23.2. This figure shows that trajectories of the bimatrix replicator dynamics converge to a two-species ESS as defined in Sects. 6.1 and 6.2. In particular, the interior NE in panel E is not a two-species ESS, and it is an unstable equilibrium for the bimatrix replicator dynamics. In panel C the interior NE is two-species weak ESS, and it is (neutrally) stable for the bimatrix replicator dynamics

Note there are some problems with the interpretation of the dynamics of two populations in this way, related to the assumption of exponential growth of populations, since the above dynamics effectively assume that the relative size of the two populations remains constant (Argasinski 2006).

7 The Habitat Selection Game

Fretwell and Lucas (1969) introduced the Ideal Free Distribution (IFD) to describe a distribution of animals in a heterogeneous environment consisting of discrete patches $i = 1, \ldots, n$. The IFD assumes that animals are free to move between several patches, travel is cost-free, each individual knows perfectly the quality of all patches, and all individuals have the same competitive abilities. Assuming that these patches differ in their basic quality B_i (i.e., their quality when unoccupied), the IFD model predicts that the best patch will always be occupied.

Let us assume that patches are arranged in descending order ($B_1 > \cdots > B_n > 0$) and m_i is the animal abundance in patch i. Let $p_i = m_i/(m_1 + \cdots + m_n)$ be the proportion of animals in patch i, so that $\mathbf{p} = (p_1, \ldots, p_n)$ describes the spatial distribution of the population. For a monomorphic population, p_i also specifies the individual strategy as the proportion of the lifetime an average animal spends in patch i. We assume that the payoff in each patch, $V_i(p_i)$, is a decreasing function of animal abundance in that patch, i.e., the patch payoffs are negatively density dependent. Then, fitness of a mutant with strategy $\tilde{\mathbf{p}} = (\tilde{p}_1, \ldots, \tilde{p}_n)$ in the resident monomorphic population with distribution $\mathbf{p} = (p_1, \ldots, p_n)$ is

$$E(\tilde{\mathbf{p}}, \mathbf{p}) = \sum_{i=1}^{n} \tilde{p}_i V_i(p_i).$$

However, we do not need to make the assumption that the population is monomorphic, because what really matters in calculating $E(\tilde{\mathbf{p}}, \mathbf{p})$ above is the animal distribution $\mathbf{p}$. If the population is not monomorphic, this distribution can be different from strategies animals use, and we call it the population mean strategy. Thus, in the Habitat Selection game, individuals do not enter pairwise conflicts, but they play against the population mean strategy (referred to as a "playing the field" or "population" game).

Fretwell and Lucas (1969) introduced the concept of the Ideal Free Distribution which is a population distribution $\mathbf{p} = (p_1, \ldots, p_n)$ that satisfies two conditions:

1. There exists a number $1 \leq k \leq n$ such that $p_1 > 0, \ldots, p_k > 0$ and $p_{k+1} = \cdots = p_n = 0$
2. $V_1(p_1) = \cdots = V_k(p_k) = V^*$ and $V^* \geq V_i(p_i)$ for $i = k + 1, \ldots, n$.

They proved that provided patch payoffs are negatively density dependent (i.e., decreasing functions of the number of individuals in a patch), then there exists a unique IFD which Cressman and Křivan (2006) later showed is an ESS. In the next section, we will discuss two commonly used types of patch payoff functions.

7.1 Parker's Matching Principle

Parker (1978) considered the case where resource input rates r_i, $i = 1, \dots, n$ are constant and resources are consumed immediately when they enter the patch and so there is no standing crop. This leads to a particularly simple definition of animal patch payoffs as the ratio of the resource input rate divided by the number of individuals there, i.e.,

$$V_i = \frac{r_i}{m_i} = \frac{r_i}{p_i M} \tag{23.12}$$

where $M = m_1 + \cdots + m_n$ is the overall population abundance. The matching principle then says that animals distribute themselves so that their abundance in each patch is proportional to the rate with which resources arrive into the patch, $p_i/p_j = r_i/r_j$. This is nothing other than the IFD for payoff functions (23.12). It is interesting to notice that all patches will be occupied independently of the total population abundance. Indeed, as the consumer density in the $i-$th patch decreases, payoff $r_i/(p_i M)$ increases, which attracts some animals, and there cannot be unoccupied patches. There is an important difference between this (nonlinear) payoff function (23.12) and the linear payoff function that we consider in the following Eq. (23.13), because as the local population abundance in a patch decreases, then (23.12) tends to infinity, but (23.13) tends to r_i. This means that in the first case there cannot be unoccupied patches (irrespective of their basic patch quality r_i) because the payoffs in occupied patches are finite, but the payoff in unoccupied patches would be infinite (provided all $r_i > 0$). This argument does not apply in the case of the logistic payoff (23.13). This concept successfully predicts the distribution of house flies that arrive at a cow pat where they immediately mate (Blanckenhorn et al. 2000; Parker 1978, 1984) or of fish that are fed at two feeders in a tank (Berec et al. 2006; Milinski 1979, 1988).

In the next section, we consider the situation where resources are not consumed immediately upon entering the system.

7.2 Patch Payoffs are Linear

Here we consider two patches only, and we assume that the payoff in habitat $i (= 1, 2)$ is a linearly decreasing function of population abundance:

$$V_i = r_i\left(1 - \frac{m_i}{K_i}\right) = r_i\left(1 - \frac{p_i M}{K_i}\right) \tag{23.13}$$

where m_i is the population density in habitat i, r_i is the intrinsic per capita population growth rate in habitat i, and K_i is its carrying capacity. The total population size in the two-habitat environment is denoted by $M(= m_1 + m_2)$, and the proportion of the population in habitat i is $p_i = m_i/M$. Payoff (23.13) is often used in population dynamics where it describes the logistic population growth.

Let us consider an individual which spends proportion $\tilde{p}_1$ of its lifetime in habitat 1 and $\tilde{p}_2$ in habitat 2. Provided total population density is fixed at M, then its fitness in the population with mean strategy $\mathbf{p} = (p_1, p_2)$ is

$$E(\tilde{\mathbf{p}}, \mathbf{p}) = \tilde{p}_1 V_1(p_1) + \tilde{p}_2 V_2(p_2) = \tilde{\mathbf{p}}\, U \mathbf{p}^T,$$

where

$$U = \begin{pmatrix} r_1(1 - \frac{M}{K_1}) & r_1 \\ r_2 & r_2(1 - \frac{M}{K_2}) \end{pmatrix}$$

is the payoff matrix with two strategies, where strategy i represents staying in patch i $(i = 1, 2)$. This shows that the Habitat Selection game with a linear payoff can be written for a fixed population size as a matrix game. If the per capita intrinsic population growth rate in habitat 1 is higher than that in habitat 2 ($r_1 > r_2$), the IFD is (Křivan and Sirot 2002)

$$p_1 = \begin{cases} 1 & \text{if } M < K_1 \frac{r_1 - r_2}{r_1} \\ \dfrac{r_2 K_1}{r_2 K_1 + r_1 K_2} + \dfrac{K_1 K_2 (r_1 - r_2)}{(r_2 K_1 + r_1 K_2) M} & \text{otherwise.} \end{cases} \tag{23.14}$$

When the total population abundance is low, the payoff in habitat 1 is higher than the payoff in habitat 2 for all possible population distributions because the competition in patch 1 is low due to low population densities. For higher population abundances, neither of the two habitats is always better than the other, and under the IFD payoffs in both habitats must be the same ($V_1(p_1) = V_2(p_2)$). Once again, it is important to emphasize here that the IFD concept is different from maximization of the mean animal fitness

$$\overline{W}(\mathbf{p}, \mathbf{p}) = p_1 V_1(p_1) + p_2 V_2(p_2)$$

which would lead to

$$p_1 = \begin{cases} 1 & \text{if } M < K_1 \dfrac{r_1 - r_2}{2 r_1} \\ \dfrac{r_2 K_1}{r_1 K_2 + r_2 K_1} + \dfrac{K_1 K_2 (r_1 - r_2)}{2(r_1 K_2 + r_2 K_1) M} & \text{otherwise.} \end{cases} \tag{23.15}$$

The two expressions (23.14) and (23.15) are the same if and only if $r_1 = r_2$. Interestingly, by comparing (23.14) and (23.15), we see that maximizing mean fitness leads to fewer animals than the IFD in the patch with higher basic quality r_i (i.e., in patch 1).

7.3 Some Extensions of the Habitat Selection Game

The Habitat Selection game as described makes several assumptions that were relaxed in the literature. One assumption is that patch payoffs are decreasing functions of population abundance. This assumption is important because it guarantees that a unique IFD exists. However, patch payoffs can also be increasing functions of population abundance. In particular, at low population densities, payoffs can increase as more individuals enter a patch, and competition is initially weak. For example, more individuals in a patch can increase the probability of finding a mate. This is called the Allee effect. The IFD for the Allee effect has been studied in the literature (Cressman and Tran 2015; Fretwell and Lucas 1969; Křivan 2014; Morris 2002). It has been shown that for hump-shaped patch payoffs, up to three IFDs can exist for a given overall population abundance. At very low overall population abundances, only the most profitable patch will be occupied. At intermediate population densities, there are two IFDs corresponding to pure strategies where all individuals occupy patch 1 only, or patch 2 only. As population abundance increases, competition becomes more severe, and an interior IFD appears exactly as in the case of negative density-dependent payoff functions. At high overall population abundances, only the interior IFD exists due to strong competition among individuals. It is interesting to note that as the population numbers change, there can be sudden (discontinuous) changes in the population distribution. Such erratic changes in the distribution of deer mice were observed and analyzed by Morris (2002).

Another complication that leads to multiple IFDs is the cost of dispersal. Let us consider a positive migration cost c between two patches. An individual currently in patch 1 will migrate to patch 2 only if the payoff there is such that $V_2(p_2) - c \geq V_1(p_1)$. Similarly, an individual currently in patch 2 will migrate to patch 1 only if its payoff does not decrease by doing so, i.e., $V_1(p_1) - c \geq V_2(p_2)$. Thus, all distributions (p_1, p_2) that satisfy these two inequalities form the set of IFDs (Mariani et al. 2016).

The Habitat Selection game was also extended to situations where individuals perceive space as a continuum (e.g., Cantrell et al. 2007, 2012; Cosner 2005). The movement by diffusion is then combined, or replaced, by a movement along the gradient of animal fitness.

7.4 Habitat Selection for Two Species

Instead of a single species, we now consider two species with population densities M and N dispersing between two patches. We assume that individuals of these species compete in each patch both intra- and inter-specifically. Following our single-species Habitat Selection game, we assume that individual payoffs are linear functions of species distribution (Křivan and Sirot 2002; Křivan et al. 2008)

$$V_i(\mathbf{p},\mathbf{q}) = r_i\left(1 - \frac{p_i M}{K_i} - \frac{\alpha_i q_i N}{K_i}\right),$$
$$W_i(\mathbf{p},\mathbf{q}) = s_i\left(1 - \frac{q_i N}{L_i} - \frac{\beta_i p_i M}{L_i}\right),$$

where $\mathbf{p} = (p_1, p_2)$ denotes the distribution of species one and $\mathbf{q} = (q_1, q_2)$ the distribution of species two. Here, positive parameters α_i (respectively β_i) are interspecific competition coefficients, r_i (respectively s_i) are the intrinsic per capita population growth rates, and K_i (respectively L_i) are the environmental carrying capacities. The two-species Habitat Selection game cannot be represented in a bimatrix form (to represent it in a matrix form, we would need four matrices), because the payoff in patch i for a given species depends not only on the distribution (strategy) of its competitors but also on the distribution of its own conspecifics. The equal payoff line for species one (two) are those $(\mathbf{p},\mathbf{q}) \in \Delta_2 \times \Delta_2$ for which $V_1(\mathbf{p},\mathbf{q}) = V_2(\mathbf{p},\mathbf{q})$ ($W_1(\mathbf{p},\mathbf{q}) = W_2(\mathbf{p},\mathbf{q})$). Since payoffs are linear functions, these are lines in the coordinates p_1 and q_1, but as opposed to the case of bimatrix games in Sects. 6.1 and 6.2, they are neither horizontal nor vertical. If they do not intersect in the unit square, the two species cannot coexist in both patches at an NE. The most interesting case is when the two equal payoff lines intersect inside the unit square. Křivan et al. (2008) showed that the interior intersection is the two-species ESS provided

$$\begin{aligned} &r_1 s_1 K_2 L_2(1-\alpha_1\beta_1) + r_1 s_2 K_2 L_1(1-\alpha_1\beta_2) + \\ &\quad r_2 s_1 K_1 L_2(1-\alpha_2\beta_1) + r_2 s_2 K_1 L_1(1-\alpha_2\beta_2) > 0. \end{aligned}$$

Geometrically, this condition states that the equal payoff line for species one has a more negative slope than that for species two. This allows us to extend the concept of the single-species Habitat Selection game to two species that compete in two patches. In this case the two-species IFD is defined as a two-species ESS. We remark that the best response dynamics do converge to such two-species IFD (Křivan et al. 2008).

One of the predictions of the Habitat Selection game for two species is that as competition gets stronger, the two species will spatially segregate (e.g., Křivan and Sirot 2002; Morris 1999). Such spatial segregation was observed in experiments with two bacterial strains in a microhabitat system with nutrient-poor and nutrient-rich patches (Lambert et al. 2011).

8 Dispersal and Evolution of Dispersal

Organisms often move from one habitat to another, which is referred to as dispersal. We focus here on dispersal and its relation to the IFD discussed in the previous section.

Here we consider n habitat patches and a population of individuals that disperse between them. In what follows we will assume that the patches are either adjacent (in particular when there are just two patches) or that the travel time between them is negligible when compared to the time individuals spend in these patches. There are two basic questions:

1. When is dispersal an adaptive strategy, i.e., when does individual fitness increase for dispersing animals compared to those who are sedentary?
2. Where should individuals disperse?

To describe changes in population densities, we will consider demographic population growth in each patch and dispersal between patches. Dispersal is described by the propensity of individuals to disperse ($\delta \geq 0$) and by a dispersal matrix D. The entries of this matrix (D_{ij}) describe the transition probabilities that an individual currently in patch j moves to patch i per unit of time. We remark that D_{ii} is the probability of staying in patch i. Per capita population growth rates in patches are given by f_i (e.g., f_i can be the logistic growth rate V_i (23.13) in Sect. 7.2). The changes in population numbers are then described by population–dispersal dynamics

$$\frac{dm_i}{dt} = m_i f_i(m_i) + \delta \sum_{j=1}^{n} \left(D_{ij}(\mathbf{m})m_j - D_{ji}(\mathbf{m})m_i\right) \text{ for } i = 1, \ldots, n \tag{23.16}$$

where $\mathbf{m} = (m_1, \cdots, m_n)$ is the vector of population densities in n patches. Thus, the first term in the above summation describes immigration to patch i from other patches, and the second term describes emigration from patch i to other patches. In addition, we assume that D is irreducible, i.e., there are no isolated patches.

The case that corresponds to the passive diffusion between patches assumes that entries of the dispersal matrix are constant and the matrix is symmetric. It was shown (Takeuchi 1996) that when functions f_i are decreasing with $f_i(0) > 0$ and $f_i(K_i) = 0$ for some $K_i > 0$, then model (23.16) has an interior equilibrium which is globally asymptotically stable. However this does not answer the question of whether such an equilibrium is evolutionarily stable, i.e., whether it is resistant to invasion of mutants with the same traits (parameters) as the resident population, but different propensity to disperse δ. The answer to this question depends on the entries of the dispersal matrix. An interior population distribution $\mathbf{m}^* = (m_1^*, \ldots, m_n^*)$ will be the IFD provided patch payoffs in all patches are the same, i.e., $f_1(m_1^*) = \cdots = f_n(m_n^*)$. This implies that at the population equilibrium, there is no net dispersal, i.e.,

$$\delta \sum_{j=1}^{n} \left(D_{ij} m_j^* - D_{ji} m_i^*\right) = 0.$$

There are two possibilities. Either

$$\sum_{j=1}^{n}\left(D_{ij}m_j^* - D_{ji}m_i^*\right) = 0, \tag{23.17}$$

or $\delta = 0$. The pattern of equalized immigration and emigration satisfying (23.17) is called "balanced dispersal" (Doncaster et al. 1997; Holt and Barfield 2001; McPeek and Holt 1992). Under balanced dispersal, there is an inverse relation between local population size and its dispersal rate. In other words, individuals at good sites are less likely to disperse than those from poor sites. When dispersal is unbalanced, Hastings (1983) showed that mutants with lower propensity to disperse will outcompete the residents and no dispersal ($\delta = 0$) is the only evolutionarily stable strategy.

However, dispersal can be favored even when it is not balanced. Hamilton and May (1977) showed that unconditional and costly dispersal among very many patches can be promoted because it reduces competition between relatives. Their model was generalized by Comins et al. (1980) who assumed that because of stochastic effects, a proportion e of patches can become empty at any time step. A proportion p of migrants survives migration and re-distributes at random (assuming the Poisson distribution) among the patches. These authors derived analytically the evolutionarily stable dispersal strategy that is given by a complicated implicit formula (see formula (3) in Comins et al. 1980). As population abundance increases, the evolutionarily stable dispersal rate converges to a simpler formula

$$\delta = \frac{e}{1 - p(1-e)}.$$

Here the advantage of dispersal results from the possibility of colonizing an extinct patch.

Evolution of mobility in predator–prey systems was also studied by Xu et al. (2014). These authors showed how interaction strength between mobile vs. sessile prey and predators influences the evolution of dispersal.

9 Foraging Games

Foraging games describe interactions between prey, their predators, or both. These games assume that either predator or prey behave in order to maximize their fitness. Typically, the prey strategy is to avoid predators while predators try to track their prey. Several models that focus on various aspects of predator–prey interactions were developed in the literature (e.g., Brown and Vincent 1987; Brown et al. 1999, 2001; Vincent and Brown 2005).

An important component of predation is the functional response defined as the per predator rate of prey consumption (Holling 1959). It also serves as a basis for models of optimal foraging (Stephens and Krebs 1986) that aim to predict diet selection of predators in environments with multiple prey types. In this section we start with a model of optimal foraging, and we show how it can be derived using extensive form games (see Volume I, ▶ Chap. 6, "Evolutionary Game Theory"). As an example of a predator–prey game, we then discuss predator–prey distribution in a two-patch environment.

9.1 Optimal Foraging as an Agent Normal Form of an Extensive Game

Often it is assumed that a predator's fitness is proportional to its prey intake rate, and the functional response serves as a proxy of fitness. In the case of two or more prey types, the multi-prey functional response is the basis of the diet choice model (Charnov 1976; Stephens and Krebs 1986) that predicts the predator's optimal diet as a function of prey densities in the environment. Here we show how functional responses can be derived using decision trees of games given in extensive form (Cressman 2003; Cressman et al. 2014; see also Broom et al. 2004, for an example of where this methodology was used in a model of food stealing). Let us consider a decision tree in Fig. 23.4 describing a single predator feeding on two prey types. This decision tree assumes that a searching predator meets prey type 1 with probability p_1 and prey type 2 with probability p_2 during the search time τ_s. For simplicity we will assume that $p_1 + p_2 = 1$. Upon an encounter with a prey individual, the predator decides whether to attack the prey (prey type 1 with

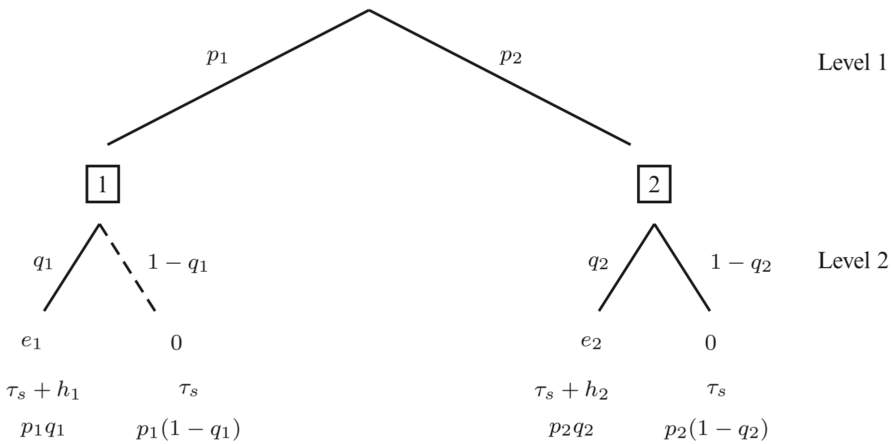


Fig. 23.4 The decision tree for two prey types. The first level gives the prey encounter distribution. The second level gives the predator activity distribution. The final row of the diagram gives the probability of each predator activity event and so sums to 1. If prey 1 is the more profitable type, the edge in the decision tree corresponding to not attacking this type of prey is never followed at optimal foraging (indicated by the *dashed edge* in the tree)

probability q_1 and prey type 2 with probability q_2) or not. When a prey individual is captured, the energy that the predator receives is denoted by e_1 or e_2. The predator's cost is measured by the time lost. This time consists of the search time τ_s and the time needed to handle the prey (h_1 for prey type 1 and h_2 for prey type 2).

Calculation of functional responses is based on renewal theory which proves that the long-term intake rate of a given prey type can be calculated as the mean energy intake during one renewal cycle divided by the mean duration of the renewal cycle (Houston and McNamara 1999; Stephens and Krebs 1986). A single renewal cycle is given by a predator passing through the decision tree in Fig. 23.4. Since type i prey are only killed when the path denoted by p_i and then q_i is followed, the functional response to prey $i\,(= 1, 2)$ is

$$f_i(q_1, q_2) = \frac{p_i q_i}{p_1\,(q_1(\tau_s + h_1) + (1 - q_1)\tau_s) + p_2\,(q_2(\tau_s + h_2) + (1 - q_2)\tau_s)}$$
$$= \frac{p_i q_i}{\tau_s + p_1 q_1 h_1 + p_2 q_2 h_2}.$$

When x_i denotes density of prey type i in the environment and the predator meets prey at random, $p_i = x_i/x$, where $x = x_1 + x_2$. Setting $\lambda = 1/(\tau_s x)$ leads to

$$f_i(q_1, q_2) = \frac{\lambda x_i q_i}{1 + \lambda x_1 q_1 h_1 + \lambda x_2 q_2 h_2}.$$

These are the functional responses assumed in standard two prey type models. The predator's rate of energy gain is given by

$$E(q_1, q_2) = e_1 f_1(q_1, q_2) + e_2 f_2(q_1, q_2) = \frac{e_1 p_1 q_1 + e_2 p_2 q_2}{\tau_s + p_1 q_1 h_1 + p_2 q_2 h_2}. \tag{23.18}$$

This is the proxy of the predator's fitness which is maximized over the predator's diet (q_1, q_2), $(0 \le q_i \le 1, i = 1, 2)$.

Here, using the agent normal form of extensive form game theory (Cressman 2003), we show an alternative, game theoretical approach to find the optimal foraging strategy. This method assigns a separate player (called an agent) to each decision node (here 1 or 2). The possible decisions at this node become the agent's strategies, and its payoff is given by the total energy intake rate of the predator it represents. Thus, all of the virtual agents have the same common payoff. The optimal foraging strategy of the single predator is then a solution to this game. In our example, player 1 corresponds to decision node 1 with strategy set $\Delta_1 = \{q_1 \mid 0 \le q_1 \le 1\}$ and player 2 to node 2 with strategy set $\Delta_2 = \{q_2 \mid 0 \le q_2 \le 1\}$. Their common payoff $E(q_1, q_2)$ is given by (23.18), and we seek the NE of the two-player game. Assuming that prey type 1 is the more profitable for the predator, as its energy content per unit handling time is higher than the profitability of the second prey type (i.e., $e_1/h_1 > e_2/h_2$) we get $E(1, q_2) > E(q_1, q_2)$ for all $0 \le q_1 < 1$ and $0 \le q_2 \le 1$. Thus, at any NE, player 1 must play $q_1 = 1$. The NE strategy of player 2 is then any best response to $q_1 = 1$ (i.e., any q_2 that satisfies $E(1, q_2') \le E(1, q_2)$

for all $0 \le q_2' \le 1$) which yields

$$q_2 = \begin{cases} 0 & \text{if } p_1 > p_1^* \\ 1 & \text{if } p_1 < p_1^* \\ [0,1] & \text{if } p_1 = p_1^*, \end{cases} \tag{23.19}$$

where

$$p_1^* = \frac{e_2 \tau_s}{e_1 h_2 - e_2 h_1}. \tag{23.20}$$

This NE coincides with the optimal strategy derived by maximization of (23.18). It makes quite striking predictions. While the more profitable prey type is always included in the predator's diet, inclusion of the less profitable prey type is independent of its own density and depends on the more profitable prey type density only. This prediction was experimentally tested with great tits (e.g., Berec et al. 2003; Krebs et al. 1977). That the Nash equilibrium coincides with the optimal foraging strategy (i.e., with the maximum of E) in this model is not a coincidence. Cressman et al. (2014) proved that this is so for all foraging games with a 2-level decision tree. For decision tress with more levels, they showed that the optimal foraging strategy is always an NE of the corresponding agent normal form game and that other, nonoptimal, NE may also appear.

9.2 A Predator-Prey Foraging Game

As an example we consider here a predator–prey foraging game between prey and predators in a two-patch environment. If x_i denotes the abundance of prey in patch $i (= 1, 2)$, the total abundance of prey is $x = x_1 + x_2$ and, similarly, the total abundance of predators is $y = y_1 + y_2$. Let $\mathbf{u} = (u_1, u_2)$ be the distribution of prey and $\mathbf{v} = (v_1, v_2)$ be the distribution of predators. We neglect the travel time between patches so that $u_1 + u_2 = v_1 + v_2 = 1$ (i.e., each animal is either in patch 1 or patch 2). We assume that the prey population grows exponentially at each patch with the per capita population growth rate r_i and it is consumed by predators. The killing rate is given by the functional response. For simplicity we neglect the handling time so that the functional response in patch i is $f_i = \lambda_i x_i$, i.e., the per prey per predator killing rate is λ_i. The prey payoff in patch i is given by the per capita prey population growth rate in that patch, i.e., $r_i - \lambda_i v_i y$ as there are $v_i y$ predators in patch i. The fitness of a prey individual is

$$V(\mathbf{u}, \mathbf{v}) = (r_1 - \lambda_1 v_1 y)u_1 + (r_2 - \lambda_2 v_2 y)u_2. \tag{23.21}$$

The predator payoff in patch i is given by the per capita predator population growth rate $e_i u_i x - m_i$, where e_i is a coefficient by which the energy gained by feeding on prey is transformed into new predators and m_i is the per capita predator mortality rate in patch i. The fitness of a predator with strategy $\mathbf{v} = (v_1, v_2)$ when the prey use strategy $\mathbf{u} = (u_1, u_2)$ is

$$W(\mathbf{v}, \mathbf{u}) = (e_1\lambda_1 u_1 x - m_1)v_1 + (e_2\lambda_2 u_2 x - m_2)v_2. \tag{23.22}$$

This predator–prey game can be represented by the following payoff bimatrix

$$\begin{array}{lcc} \text{Prey}\backslash\text{Predator} & \text{Patch 1} & \text{Patch 2} \\ \begin{array}{l}\text{Patch 1}\\ \text{Patch 2}\end{array} & \multicolumn{2}{c}{\left(\begin{array}{cc} (r_1 - \lambda_1 y, e_1\lambda_1 x - m_1) & (r_1, -m_2) \\ (r_2, -m_1) & (r_2 - \lambda_2 y, e_2\lambda_2 x - m_2) \end{array}\right).} \end{array}$$

That is, the rows in this bimatrix correspond to the prey strategy (the first row means the prey are in patch 1; the second row means the prey are in patch 2), and similarly columns represent the predator strategy. The first of the two expressions in the entries of the bimatrix is the payoff for the prey, and the second is the payoff for the predators.

For example, we will assume that for prey patch 1 has a higher basic patch quality when compared to patch 2 (i.e., $r_1 \geq r_2$) while for predators patch 1 has a higher mortality rate ($m_1 > m_2$). The corresponding NE is (Křivan 1997)

(a) (u_1^*, v_1^*) if $x > \dfrac{m_1 - m_2}{e_1\lambda_1}, \quad y > \dfrac{r_1 - r_2}{\lambda_1}$,

(b) $(1, 1)$ if $x > \dfrac{m_1 - m_2}{e_1\lambda_1}, \quad y < \dfrac{r_1 - r_2}{\lambda_1}$,

(c) $(1, 0)$ if $x < \dfrac{m_1 - m_2}{e_1\lambda_1}$,

where

$$(u_1^*, v_1^*) = \left(\frac{m_1 - m_2 + e_2\lambda_2 x}{(e_1\lambda_1 + e_2\lambda_2)x}, \frac{r_1 - r_2 + \lambda_2 y}{(\lambda_1 + \lambda_2)y}\right).$$

If prey abundance is low (case (c)), all prey will be in patch 1, while predators will stay in patch 2. Because the mortality rate for predators in patch 1 is higher than in patch 2 and prey abundance is low, patch 2 is a refuge for predators. If predator abundance is low and prey abundance is high (case (b)), both predators and prey will aggregate in patch 1. When the NE is strict (cases (b) and (c) above), it is also the ESS because there is no alternative strategy with the same payoff. However, when the NE is mixed (case (a)), there exist alternative best replies to it. This mixed NE is the two-species weak ESS. It is globally asymptotically stable for the continuous-time best response dynamics (Křivan et al. 2008) that model dispersal behavior whereby individuals move to the patch with the higher payoff. We remark that for some population densities ($x = \frac{m_1 - m_2}{e_1\lambda_1}$ and $y = \frac{r_1 - r_2}{\lambda_1}$), the NE is not uniquely defined, which is a general property of matrix games (the game is nongeneric in this case).

10 Signaling Games

Signaling between animals occurs in a number of contexts. This can be signals, often but not necessarily between conspecifics, warning of approaching predators. This situation can be game theoretic, as the signaler runs a potentially higher risk of being targeted by the predator. There are also cases of false signals being given when no predator is approaching to force food to be abandoned which can then be consumed by the signaler (Flower 2011). Alternatively within a group of animals, each individual may need to decide how to divide their time between vigilance and feeding, where each individual benefits from the vigilance of others as well as itself, and this has been modeled game-theoretically (e.g., Brown 1999; McNamara and Houston 1992; Sirot 2012).

Another situation occurs between relatives over items of food, for example, a parent bird feeding its offspring. Young birds beg aggressively for food, and the parent must decide which to feed, if any (it can instead consume the item itself). The most well-known model of this situation is the Sir Philip Sidney game (Maynard Smith 1991) and is a model of cost-free signaling (Bergstrom and Lachmann 1998).

The classic example of a signaling game is between potential mates. Males of differing quality advertise this quality to females, often in a way that is costly, and the females choose who to mate with based upon the strength of the signal. Examples are the tail of the peacock or the elaborate bowers created by bowerbirds. There is obviously a large incentive to cheat, and so how are such signals kept honest? A signal that is not at least partly correlated to quality would be meaningless, and so would eventually be ignored. The solution as developed by Zahavi (1975, 1977), the *handicap principle*, is that these costly signals are easier to bear by higher-quality mates and that evolution leads to a completely honest signal, where each quality level has a unique signal.

10.1 Grafen's Signaling Game

The following signaling model is due to (Grafen 1990a,b). Consider a population with a continuum of male quality types q and a single type of female. Assume that a male of quality q gives a signal $a = A(q)$ of this quality, where higher values of a are more costly. It is assumed that there is both a minimum quality level $q_0 > 0$ (there may or may not be a maximum quality level) and a minimum signal level $a_0 \geq 0$ (which can be thought of as giving no signal). When a female receives a signal, she allocates a quality level to the signal $P(a)$. We have a nonlinear asymmetric game with sequential decisions; in particular the nonlinearity makes this game considerably more complicated than asymmetric games such as the Battle of the Sexes of Sect. 6.1. The female pays a cost for misassessing a male of quality q as being of quality p of $D(q, p)$, which is positive for $p \neq q$, with $D(q, q) = 0$. Assuming that the probability density of males of quality q is $g(q)$, the payoff to the female, which is simply minus the expected cost, is

$$-\int_{q_0}^{\infty} D(q,p)g(q)dq.$$

An honest signaling system with strategies A^* and P^* occurs if and only if $P^*(A^*(q)) = q$, for all q. We note that here the female never misassesses a male and so pays zero cost. Clearly any alternative female assessment strategy would do worse. But how can we obtain stability against alternative (cheating) male strategies?

The fitness of a male of quality q, $W(a,p,q)$, depends upon his true quality, the quality assigned to him by the female and the cost of his signal. $W(a,p,q)$ will be increasing in p and decreasing in a. For stability of the honest signal, we need that the incremental advantage of a higher level of signaling is greater for a high-quality male than for a low-quality one, so that

$$-\frac{\frac{\partial}{\partial a}W(a,p,q)}{\frac{\partial}{\partial p}W(a,p,q)} \tag{23.23}$$

is strictly decreasing in q (note that the ratio is negative, so minus this ratio is positive), i.e., the higher quality the male, the lower the ratio of the marginal cost to the marginal benefit for an increase in the level of advertising. This ensures that completely honest signaling cannot be invaded by cheating, since costs to cheats to copy the signals of better quality males would be explicitly higher than for the better quality males, who could always thus achieve a cost they were willing to pay that the lower quality cheats would not.

The following example male fitness function is given in Grafen (1990a) (the precise fitness function to the female does not affect the solution provided that correct assessment yields 0, and any misassessment yields a negative payoff)

$$W(a,p,q) = p^r q^a, \tag{23.24}$$

with qualities in the range $q_0 \leq q < 1$ and signals of strength $a \geq a_0$, for some $r > 0$.

We can see that the function from (23.24) satisfies the above conditions on $W(a,p,q)$. In particular consider the condition from expression (23.23)

$$-\frac{\partial}{\partial a}W(a,p,q) = -p^r q^a \ln q, \qquad \frac{\partial}{\partial p}W(a,p,q) = rp^{r-1}q^a$$

which are the increase in cost per unit increase in the signal level and the increase in the payoff per unit increase in the female's perception (which in turn is directly caused by increases in signal level), respectively. The ratio from (23.23), which is proportional to the increase in cost per unit of benefit that this would yield, becomes $-p \ln q/r$, takes a larger value for lower values of q. Thus there is an honest signaling solution. This is shown in Grafen (1990a) to be given by

$$A(q) = a_0 - r \ln\left(\frac{\ln(q)}{\ln(q_0)}\right), \quad P(a) = q_0^{\exp(-(a-a_0)/r)}.$$

11 Conclusion

In this chapter we have covered some of the important evolutionary game models applied to biological situations. We should note that we have left out a number of important theoretical topics as well as areas of application. We briefly touch on a number of those below.

All of the games that we have considered involved either pairwise games, or playing the field games, where individuals effectively play against the whole population. In reality contests will sometimes involve groups of individuals. Such models were developed in Broom et al. (1997), for a recent review see Gokhale and Traulsen (2014). In addition the populations were all both effectively infinite and *well-mixed* in the sense that for any direct contest involving individuals, each pair was equally likely to meet. In reality populations are finite and have (e.g., spatial) structure. The modeling of evolution in finite populations often uses the Moran process (Moran 1958), but more recently games in finite populations have received significant attention (Nowak 2006). These models have been extended to include population structure by considering evolution on graphs (Lieberman et al. 2005), and there has been an explosion of such model applications, especially to consider the evolution of cooperation. Another feature of realistic populations that we have ignored is the state of the individual. A hungry individual may behave differently to one that has recently eaten, and nesting behavior may be different at the start of the breeding season to later on. A theory of state-based models has been developed in Houston and McNamara (1999).

In terms of applications, we have focused on classical biological problems, but game theory has also been applied to medical scenarios more recently. This includes the modeling of epidemics, especially with the intention of developing defense strategies. One important class of models (see, e.g., Nowak and May 1994) considers the evolution of the virulence of a disease as the epidemic spreads. An exciting new line of research has recently been developed which considers the development of cancer as an evolutionary game, where the population of cancer cells evolves in the environment of the individual person or animal (Gatenby et al. 2010). A survey of alternative approaches is considered in Durrett (2014).

12 Cross-References

▶ Biology and Evolutionary Games

Acknowledgements This project has received funding from the European Union Horizon 2020 research and innovation programe under the Marie Sklodowska-Curie grant agreement No 690817. VK acknowledges support provided by the Institute of Entomology (RVO:60077344).

References

Argasinski K (2006) Dynamic multipopulation and density dependent evolutionary games related to replicator dynamics. A metasimplex concept. Math Biosci 202:88–114
Axelrod R (1981) The emergence of cooperation among egoists. Am Political Sci Rev 75:306–318
Axelrod R (1984) The evolution of cooperation. Basic Books, New York
Axelrod R, Hamilton WD (1981) The Evolution of cooperation. Science 211:1390–1396
Ball MA, Parker GA (2007) Sperm competition games: the risk model can generate higher sperm allocation to virgin females. J Evol Biol 20:767–779
Bendorf J, Swistak P (1995) Types of evolutionary stability and the problem of cooperation. Proc Natl Acad Sci USA 92:3596–3600
Berec M, Křivan V, Berec L (2003) Are great tits parus major really optimal foragers? Can J Zool 81:780–788
Berec M, Křivan V, Berec L (2006) Asymmetric competition, body size and foraging tactics: testing an ideal free distribution in two competing fish species. Evol Ecol Res 8:929–942
Bergstrom C, Lachmann M (1998) Signaling among relatives. III. Talk is cheap. Proc Natl Acad Sci USA 95:5100–5105
Bishop DT, Cannings C (1976) Models of animal conflict. Adv appl probab 8:616–621
Blanckenhorn WU, Morf C, Reuter M (2000) Are dung flies ideal-free distributed at their oviposition and mating site? Behaviour 137:233–248
Broom M, Rychtar J (2013) Game-theoretical models in biology. CRC Press/Taylor & Francis Group, Boca Raton
Broom M, Cannings C, Vickers G (1997) Multi-player matrix games. Bull Austral Math Soc 59:931–952
Broom M, Luther RM, Ruxton GD (2004) Resistance is useless? – extensions to the game theory of kleptoparasitism. Bull Math Biol 66:1645–1658
Brown JS (1999) Vigilance, patch use and habitat selection: foraging under predation risk. Evol Ecol Res 1:49–71
Brown JS, Vincent TL (1987) Predator-prey coevolution as an evolutionary game. Lect Notes Biomath 73:83–101
Brown JS, Laundré JW, Gurung M (1999) The ecology of fear: optimal foraging, game theory, and trophic interactions. J Mammal 80:385–399
Brown JS, Kotler BP, Bouskila A (2001) Ecology of fear: Foraging games between predators and prey with pulsed resources. Ann Zool Fennici 38:71–87
Cantrell RS, Cosner C, DeAngelis DL, Padron V (2007) The ideal free distribution as an evolutionarily stable strategy. J Biol Dyn 1:249–271
Cantrell RS, Cosner C, Lou Y (2012) Evolutionary stability of ideal free dispersal strategies in patchy environments. J Math Biol 65:943–965. doi:10.1007/s00285-011-0486-5
Charnov EL (1976) Optimal foraging: attack strategy of a mantid. Am Nat 110:141–151
Comins H, Hamilton W, May R (1980) Evolutionarily stable dispersal strategies. J Theor Biol 82:205–230
Cosner C (2005) A dynamic model for the ideal-free distribution as a partial differential equation. Theor Popul Biol 67:101–108

Cressman R (2003) Evolutionary dynamics and extensive form games. The MIT Press, Cambridge, MA
Cressman R, Křivan V (2006) Migration dynamics for the ideal free distribution. Am Nat 168:384–397
Cressman R, Tran T (2015) The ideal free distribution and evolutionary stability in habitat selection games with linear fitness and Allee effect. In: Cojocaru MG (ed) Interdisciplinary topics in applied mathematics, modeling and computational science. Springer proceedings in mathematics & statistics, vol 117. Springer, Cham, pp 457–464
Cressman R, Křivan V, Brown JS, Gáray J (2014) Game-theoretical methods for functional response and optimal foraging behavior. PLOS ONE 9:e88,773
Darwin C (1871) The descent of man and selection in relation to sex. John Murray, London
Dawkins R (1976) The selfish gene. Oxford University Press, Oxford
Doncaster CP, Clobert J, Doligez B, Gustafsson L, Danchin E (1997) Balanced dispersal between spatially varying local populations: an alternative to the source-sink model. Am Nat 150:425–445
Dugatkin LA, Reeve HK (1998) Game theory & animal behavior. Oxford University Press, New York
Durrett R (2014) Spatial evolutionary games with small selection coefficients. Electron J Probab 19:1–64
Fehr E, Gachter S (2002) Altruistic punishment in humans. Nature 415:137–140
Flood MM (1952) Some experimental games. Technical report RM-789-1, The RAND corporation, Santa Monica
Flower T (2011) Fork-tailed drongos use deceptive mimicked alarm calls to steal food. Proc R Soc B 278:1548–1555
Fretwell DS, Lucas HL (1969) On territorial behavior and other factors influencing habitat distribution in birds. Acta Biotheor 19:16–32
Gatenby RA, Gillies R, Brown J (2010) The evolutionary dynamics of cancer prevention. Nat Rev Cancer 10:526–527
Gilpin ME (1975) Group selection in predator-prey communities. Princeton University Press, Princeton
Gokhale CS, Traulsen A (2014) Evolutionary multiplayer games. Dyn Games Appl 4:468–488
Grafen A (1990a) Biological signals as handicaps. J Theor Biol 144:517–546
Grafen A (1990b) Do animals really recognize kin? Anim Behav 39:42–54
Haigh J (1975) Game theory and evolution. Adv Appl Probab 7:8–11
Hamilton WD (1964) The genetical evolution of social behavior. J Theor Biol 7:1–52
Hamilton WD (1967) Extraordinary sex ratios. Science 156:477–488
Hamilton WD, May RM (1977) Dispersal in stable environments. Nature 269:578–581
Hammerstein P, Parker GA (1982) The asymmetric war of attrition. J Theor Biol 96:647–682
Hardin G (1968) The tragedy of the commons. Science 162:1243–1248
Hastings A (1983) Can spatial variation alone lead to selection for dispersal? Theor Popul Biol 24:244–251
Hofbauer J, Sigmund K (1998) Evolutionary games and population dynamics. Cambridge University Press, Cambridge, UK
Holling CS (1959) Some characteristics of simple types of predation and parasitism. Can Entomol 91:385–398
Holt RD, Barfield M (2001) On the relationship between the ideal-free distribution and the evolution of dispersal. In: Danchin JCE, Dhondt A, Nichols J (eds) Dispersal. Oxford University Press, Oxford/New York, pp 83–95
Houston AI, McNamara JM (1999) Models of adaptive behaviour. Cambridge University Press, Cambridge, UK
Kerr B, Riley M, Feldman M, Bohannan B (2002) Local dispersal promotes biodiversity in a real-life game of rock-paper-scissors. Nature 418:171–174
Kokko H (2007) Modelling for field biologists and other interesting people. Cambridge University Press, Cambridge, UK

Komorita SS, Sheposh JP, Braver SL (1968) Power, use of power, and cooperative choice in a two-person game. J Personal Soc Psychol 8:134–142
Krebs JR, Erichsen JT, Webber MI, Charnov EL (1977) Optimal prey selection in the great tit (*Parus major*). Anim Behav 25:30–38
Křivan V (1997) Dynamic ideal free distribution: effects of optimal patch choice on predator-prey dynamics. Am Nat 149:164–178
Křivan V (2014) Competition in di- and tri-trophic food web modules. J Theor Biol 343:127–137
Křivan V, Sirot E (2002) Habitat selection by two competing species in a two-habitat environment. Am Nat 160:214–234
Křivan V, Cressman R, Schneider C (2008) The ideal free distribution: a review and synthesis of the game theoretic perspective. Theor Popul Biol 73:403–425
Lambert G, Liao D, Vyawahare S, Austin RH (2011) Anomalous spatial redistribution of competing bacteria under starvation conditions. J Bacteriol 193:1878–1883
Lieberman E, Hauert C, Nowak MA (2005) Evolutionary dynamics on graphs. Nature 433:312–316
Lorenz K (1963) Das sogenannte Böse zur Naturgeschichte der Aggression. Verlag Dr. G Borotha-Schoeler
Mariani P, Křivan V, MacKenzie BR, Mullon C (2016) The migration game in habitat network: the case of tuna. Theor Ecol 9:219–232
Maynard Smith J (1974) The theory of games and the evolution of animal conflicts. J Theor Biol 47:209–221
Maynard Smith J (1982) Evolution and the theory of games. Cambridge University Press, Cambridge, UK
Maynard Smith J (1991) Honest signalling: the Philip Sidney Game. Anim Behav 42:1034–1035
Maynard Smith J, Parker GA (1976) The logic of asymmetric contests. Anim Behav 24:159–175
Maynard Smith J, Price GR (1973) The logic of animal conflict. Nature 246:15–18
McNamara JM, Houston AI (1992) Risk-sensitive foraging: a review of the theory. Bull Math Biol 54:355–378
McPeek MA, Holt RD (1992) The evolution of dispersal in spatially and temporally varying environments. Am Nat 140:1010–1027
Mesterton-Gibbons M, Sherratt TN (2014) Bourgeois versus anti-Bourgeois: a model of infinite regress. Anim Behav 89:171–183
Milinski M (1979) An evolutionarily stable feeding strategy in sticklebacks. Zeitschrift für Tierpsychologie 51:36–40
Milinski M (1988) Games fish play: making decisions as a social forager. Trends Ecol Evol 3:325–330
Moran P (1958) Random processes in genetics. In: Mathematical proceedings of the Cambridge philosophical society, vol 54. Cambridge University Press, Cambridge, UK, pp 60–71
Morris DW (1999) Has the ghost of competition passed? Evol Ecol Res 1:3–20
Morris DW (2002) Measuring the Allee effect: positive density dependence in small mammals. Ecology 83:14–20
Nowak MA (2006) Five rules for the evolution of cooperation. Science 314:1560–1563
Nowak MA, May RM (1994) Superinfection and the evolution of parasite virulence. Proc R Soc Lond B 255:81–89
Parker GA (1978) Searching for mates. In: Krebs JR, Davies NB (eds) Behavioural ecology: an evolutionary approach. Blackwell, Oxford, pp 214–244
Parker GA (1984) Evolutionarily stable strategies. In: Krebs JR, Davies NB (eds) Behavioural ecology: an evolutionary approach. Blackwell, Oxford, pp 30–61
Parker GA, Thompson EA (1980) Dung fly struggles: a test of the war of attrition. Behav Ecol Sociobiol 7:37–44
Poundstone W (1992) Prisoner's Dilemma. Oxford University Press, New York
Selten R (1980) A note on evolutionarily stable strategies in asymmetrical animal conflicts. J Theor Biol 84:93–101

Sherratt TN, Mesterton-Gibbons M (2015) The evolution of respect for property. J Evol Biol 28:1185–1202

Sigmund K (2007) Punish or perish? Retaliation and collaboration among humans. Trends Ecol Evol 22:593–600

Sinervo B, Lively CM (1996) The rock-paper-scissors game and the evolution of alternative male strategies. Nature 380:240–243

Sirot E (2012) Negotiation may lead selfish individuals to cooperate: the example of the collective vigilance game. Proc R Soc B 279:2862–2867

Spencer H (1864) The Principles of biology. Williams and Norgate, London

Stephens DW, Krebs JR (1986) Foraging theory. Princeton University Press, Princeton

Takeuchi Y (1996) Global dynamical properties of Lotka-Volterra systems. World Scientific Publishing Company, Singapore

Taylor PD, Jonker LB (1978) Evolutionary stable strategies and game dynamics. Math Biosci 40:145–156

Vincent TL, Brown JS (2005) Evolutionary game theory, natural selection, and Darwinian dynamics. Cambridge University Press, Cambridge, UK

Webb JN, Houston AI, McNamara JM, Székely T (1999) Multiple patterns of parental care. Anim Behav 58:983–993

Wynne-Edwards VC (1962) Animal dispersion in relation to social behaviour. Oliver & Boyd, Edinburgh

Xu F, Cressman R, Křivan V (2014) Evolution of mobility in predator-prey systems. Discret Contin Dyn Syst Ser B 19:3397–3432

Zahavi A (1975) Mate selection–selection for a handicap. J Theor Biol 53:205–214

Zahavi A (1977) Cost of honesty–further remarks on handicap principle. J Theor Biol 67:603–605

Games in Aerospace: Homing Missile Guidance

24

Joseph Z. Ben-Asher and Jason L. Speyer

Contents

J. Z. Ben-Asher (✉)
Aerospace Engineering, Technion – Israel Institute of Technology, Haifa, Israel
e-mail: yossiba@tx.technion.ac.il

J. L. Speyer
Mechanical and Aerospace Engineering, University of California, Los Angeles, CA, USA
e-mail: speyer@seas.ucla.edu

T. Başar, G. Zaccour (eds.), *Handbook of Dynamic Game Theory*,
https://doi.org/10.1007/978-3-319-44374-4_25

Abstract

The development of a homing missile guidance law against an intelligent adversary requires the solution to a differential game. First, we formulate the deterministic homing guidance problem as a linear dynamic system with an indefinite quadratic performance criterion (LQ). This formulation allows the navigation ratio to be greater than three, which is obtained by the one-sided linear-quadratic regulator and appears to be more realistic. However, this formulation does not allow for saturation in the actuators. A deterministic game allowing saturation is formulated and shown to be superior to the LQ guidance law, even though there is no control penalty. To improve the performance of the quadratic differential game solution in the presence of saturation, trajectory-shaping feature is added. Finally, if there are uncertainties in the measurements and process noise, a disturbance attenuation function is formulated that is converted into a differential game. Since only the terminal state enters the cost criterion, the resulting estimator is a Kalman filter, but the guidance gains are a function of the assumed system variances.

Keywords

Pursuit-evasion games · Homing missile guidance · Disturbance attenuation

1 Motivation and Objectives

Homing missiles are characterized by high requirements in terms of small miss distances against fast moving and possibly maneuvering targets. Modern defense systems are typically equipped with highly sophisticated subsystems, both on board the intercepting missiles and as part of the supporting systems (e.g., radar systems), enabling good estimates of the target's flying parameters. Guidance methods which exploit such real-time measurements are more likely to become candidates for homing missiles, while simple proportional-navigation-based guidance methods will usually fall short with respect to satisfying the tough requirements. Advanced methods based on one-sided optimization problems and on differential game problems are explored in Ben-Asher and Yaesh (1998), Zarchan (1997), Bryson and Ho (1975), and Nesline and Zarchan (1981). In the present chapter, we will present this approach and its applicability to homing missile applications. This will be accomplished by introducing a representative guidance scheme that may or may not use the target's acceleration estimates in order to improve the end-game performance against possible target's maneuvers. It should be emphasized that we elaborate here several candidates from the set of possible end-game guidance schemes and this choice is by no means claimed to be the best. In practice, one should design a guidance system in accordance with its own specific requirements. However, the lessons learned here with respect to design considerations, as well as the optimization and simulation approaches, are of a general type and may serve as guidelines for improved methodology.

This chapter begins with a short history of homing guidance in Sect. 2. In Sect. 3, some of the assumptions and models of the missile and target dynamics are presented. In Sect. 4, a missile guidance law is developed based on the solution to a linear-quadratic, differential game. Given the importance of actuator saturation on homing missile performance, a differential game which includes saturation is presented in Sect. 5. To improve the performance of the quadratic differential game solution in the presence of saturation, trajectory-shaping feature is added in Sect. 6. In Sect. 7, measurement uncertainty is included in the linear-quadratic, differential game formulation. Conclusions are drawn in Sect. 8

2 Historical Background

Since the first successful test of the Lark missile (Zarchan 1997) in December 1950, proportional navigation (PN in the sequel) has come to be widely employed by homing missiles. Under this scheme, the missile is governed by

$$n_c = N' V_c \dot{\sigma} \tag{24.1}$$

where n_c is the missile's normal acceleration, $\dot{\sigma}$ is the change rate of the line of sight (the dot represents time derivative), V_c is the closing velocity, and N' is the so-called navigation constant or navigation ratio ($N' > 2$).

In the mid-1960s, it was realized that PN with $N' = 3$ is, in fact, an optimal strategy for the linearized problem, when the cost J is the control effort, as follows:

$$J = \int_0^{t_f} n_c^2(t)dt \tag{24.2}$$

subject to $y(t_f) = 0$, where y is the relative interceptor-target separation and t_f is the collision time (the elapsed time from the beginning of the end game till interception). For this case, $\dot{\sigma}$ may be expressed as (the derivation is given in the next section Eq. (24.18))

$$\dot{\sigma} = \frac{y + \dot{y} t_{go}}{V_c (t_{go})^2}, t_{go} = t_f - t \tag{24.3}$$

and hence

$$n_c = \frac{3(y + \dot{y} t_{go})}{t_{go}^2} \tag{24.4}$$

In cases where target maneuvers are significant, extensions of the PN law have been developed such as augmented proportional navigation (APN, see Nesline and Zarchan 1981) where the commanded interceptor's acceleration n_c depends on the target acceleration n_T, so that:

$$n_c = N' V_c \dot{\sigma} + \frac{N'}{2} n_T \tag{24.5}$$

It was also realized that when the transfer function relating the commanded and actual accelerations n_c and n_L, respectively, has a significant time lag (which is typical for aerodynamically controlled missiles at high altitudes), the augmented proportional navigation law can lead to a significant miss distance. To overcome this difficulty, the optimal guidance law (OGL, see Zarchan 1997) was proposed whereby J (Eq. (24.2)) is minimized, subject to state equation constraints which include the missile's dynamics of the form

$$\frac{n_L}{n_c} = \frac{1}{1 + Ts} \tag{24.6}$$

where s is the Laplace transformed differentiation operator and T – the missile's time lag (which depends on the flight conditions at the end game). The resulting guidance law easily overcomes the large time-lag problem, but is strongly dependent on the time constant and the time-to-go, namely:

$$n_c = \frac{N'(y + \dot{y}t_{go})}{t_{go}^2} + \frac{N'}{2}n_T - KN'n_L = N'V_c\dot{\sigma} + \frac{N'}{2}n_T - KN'n_L \tag{24.7}$$

The gains N' and K are defined as follows:

$$N' = \frac{6h^2\left(e^{-h} - 1 + h\right)}{2h^3 + 3 + 6h - 6h^2 - 12he^{-h} - 3e^{-2h}} \tag{24.8}$$

and

$$K = \frac{1}{h^2}\left(e^{-h} + h - 1\right) \tag{24.9}$$

where $h = t_{go}/T$. Improved guidance schemes can be obtained with an appropriate selection of a cost function that replaces J of (24.2). To this end, we next present the preliminary assumptions that render this derivation feasible and realizable.

3 Mathematical Modeling

We shall make the following assumptions:

1. The end game is two-dimensional and gravity is compensated independently.
2. The speeds of the pursuer (the missile) P and the evader (the target) E are constant during the end game (approximately true for short end games).
3. The trajectories of P and E can be linearized around their collision course.
4. The pursuer is more maneuverable than the evader.
5. The pursuer and the evader can measure their own accelerations in addition to the line-of-sight rate and have an estimate of the time-to-go.

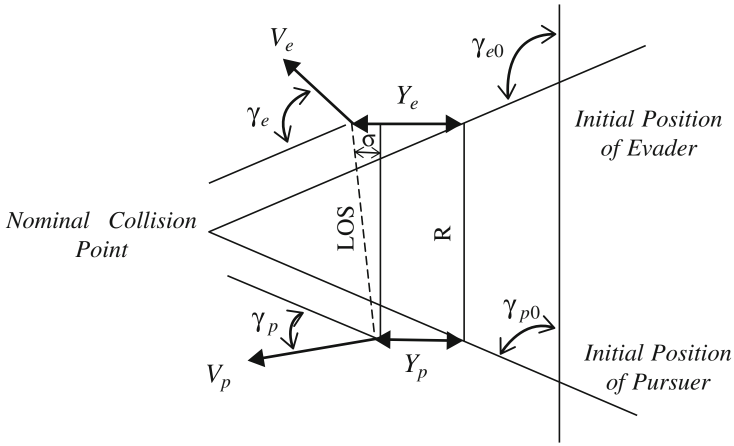


Fig. 24.1 Geometry for homing missile guidance

6. Maneuvering dynamics of the pursuer is approximated by first-order transfer functions with the respective time constants T.

We assume that the collision condition is satisfied (Fig. 24.1), namely,

$$V_p \sin(\gamma_{p0}) - V_e \sin(\gamma_{e0}) = 0 \tag{24.10}$$

where $(V_p,\ V_e)$ and $(\gamma_{p0},\ \gamma_{e0})$ are the pursuer's and evader's velocities and nominal heading angles, respectively. In this case, the nominal closing velocity V_c is given by

$$V_c = -\dot{R} = V_p \cos\left(\gamma_{p0}\right) - V_e \cos(\gamma_{e0}) \tag{24.11}$$

and the (nominal) terminal time is given by

$$t_f = \frac{R}{V_c} \tag{24.12}$$

where R is the nominal length of the line of sight.

If we allow Y_e and Y_p to be the separation (see Fig. 24.1) of the pursuer and the evader, respectively, from the nominal line of sight, and let y be the relative separation (i.e., $y = Y_e - Y_p$), we obtain the following dynamic equation:

$$\dot{y} = \dot{Y}_e - \dot{Y}_p = V_e \sin(\gamma_{e0} + \gamma_e) - V_p \sin(\gamma_{p0} + \gamma_p) \tag{24.13}$$

where γ_p, γ_e are the deviations from the base line of the pursuer's and evader's headings, respectively, as a result of control actions applied. If these deviations are small enough, we may use small angles approximation to obtain

$$\sin(\gamma_{p0} + \gamma_p) \approx \sin(\gamma_{p0}) + \cos(\gamma_{p0})\gamma_p \tag{24.14}$$

$$\sin(\gamma_{e0} + \gamma_e) \approx \sin(\gamma_{e0}) + \cos(\gamma_{e0})\gamma_e \tag{24.15}$$

Substituting the results into (24.10) and (24.13), we find (using (24.10)) that $\dot{y}$ becomes

$$\dot{y} = \dot{Y}_e - \dot{Y}_p = V_e \cos(\gamma_{e0})\gamma_e - V_p \cos(\gamma_{p0})\gamma_p \tag{24.16}$$

We can also find an expression for the line-of-sight (LOS) angle and its rate of change. Recall that σ is the line-of-sight angle, and, without loss of generality, let $\sigma(0) = 0$. We observe that $\sigma(t)$ is

$$\sigma(t) = \frac{y}{R} \tag{24.17}$$

hence

$$\dot{\sigma(t)} = \frac{d}{dt}\left(\frac{y}{R}\right) = \frac{y}{V_c(t_f - t)^2} + \frac{\dot{y}}{V_c(t_f - t)} \tag{24.18}$$

and we have thus obtained Eq. (24.3).

4 Linear-Quadratic Differential Game

4.1 Problem Formulation

Define:

$$x_1 = y,\ x_2 = \frac{dy}{dt},\ x_3 = -V_p \cos(\gamma_{p0})\frac{d\gamma_p}{dt} \tag{24.19}$$

$$u = -V_p \cos(\gamma_{p0})\frac{d\gamma_{pc}}{dt},\ w = V_e \cos(\gamma_{e0})\frac{d\gamma_e}{dt} \tag{24.20}$$

where x_3 is the pursuer's actual acceleration and is the corresponding command. The problem then has the following state space representation

$$\dot{x} = Ax + Bu + Dw \tag{24.21}$$

$$A = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & -\frac{1}{T} \end{bmatrix} \quad B = \begin{bmatrix} 0 \\ 0 \\ \frac{1}{T} \end{bmatrix} \quad D = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}$$

where u and w are the normal accelerations of the pursuer and the evader, respectively, which play as the adversaries in the following minimax problem:

$$\min_u \max_w J = \frac{a}{2}x_1^2(t_f) + \frac{b}{2}x_2^2(t_f) + \frac{c}{2}x_3^2(t_f) + \frac{1}{2}\int_0^{t_f} \left[u^2(t) - \gamma^2 w^2(t)\right] dt \tag{24.22}$$

The coefficient a weights the miss distance and is always positive. The coefficients b and c weight the terminal velocity and acceleration, respectively, and are nonnegative. The former is used to obtain the desired sensitivity reduction to time-to-go errors, while the latter can be used to control the terminal acceleration. The constant γ penalizes evasive maneuvers and therefore, by assumption 4 of Sect. 1, is required to be $\gamma > 1$.

Notice that other possible formulations are employed in the next section, namely, hard bounding the control actions rather than penalizing them in the cost. The basic ideas concerning the terminal cost components, however, are widely applicable.

4.2 Optimal Control Law

The theory of linear-quadratic differential games is covered in many textbooks (e.g., Bryson and Ho 1975, Chap. 10 and Ben-Asher and Yaesh 1998, Chap. 2) as well as elsewhere in this handbook; thus, we shall not cover it here. The solution to our problem can be obtained by solving the following Riccati equation

$$-\dot{P} = PA + A^T P - PBB^T P + \gamma^{-2} PDD^T P \tag{24.23}$$

where the terminal conditions are

$$P(t_f) = \begin{bmatrix} a & 0 & 0 \\ 0 & b & 0 \\ 0 & 0 & c \end{bmatrix} \tag{24.24}$$

and the optimal pursuer strategy includes state feedback as well as feedforward of the target maneuvers (Ben-Asher and Yaesh 1998, Chap. 2). It is of the form

$$u = -B^T P x \tag{24.25}$$

Let

$$S \equiv P^{-1} \tag{24.26}$$

and define

$$S = \begin{bmatrix} S1 & S2 & S3 \\ S2 & S4 & S5 \\ S3 & S5 & S6 \end{bmatrix} \tag{24.27}$$

Thus,

$$\dot{S} = AS + SA^T - BB^T + \gamma^{-2}DD^T \tag{24.28}$$

and

$$S(t_f) = \begin{bmatrix} a^{-1} & 0 & 0 \\ 0 & b^{-1} & 0 \\ 0 & 0 & c^{-1} \end{bmatrix} \tag{24.29}$$

Equation (24.28) can be rewritten explicitly as

$$\dot{S} = \begin{bmatrix} -2S2 & -S4 - S3 & -S5 + S3/T \\ & -2S5 - 1/\gamma^2 & -S6 + S5/T \\ & & 2S6/T + 1/T^2 \end{bmatrix} \tag{24.30}$$

Since the equation for $\dot{S}_6$ is independent of the others, the solution for S6 can be easily obtained

$$S6(t) = -\frac{1}{2T} + e^{2h}\left(\frac{1}{2T} + \frac{1}{c}\right) \tag{24.31}$$

where $h = \frac{\tau}{T}$ and $\tau = t_f - t$. We can now continue to solve sequentially for S5–S1. The following results have been obtained by Maple symbolic code.

$$S5(t) = -\frac{1}{2} - \frac{1}{2}e^{2h} - \frac{Te^{2h}}{c} + \frac{e^h(c+T)}{c} \tag{24.32}$$

$$S4(t) = \tau + \frac{T}{2}e^{2h} + \frac{T^2}{c}e^{2h} - 2Te^h - \frac{2T^2}{c}e^h - \frac{\tau}{\gamma^2} + \frac{3T}{2} + \frac{T^2}{c} + \frac{1}{b} \tag{24.33}$$

$$S3(t) = \frac{T}{2} + \frac{T}{2}e^{2h} + \frac{T^2e^{2h}}{c} - e^h\tau - \frac{e^hT\tau}{c} - \frac{T^2e^h}{c} \tag{24.34}$$

$$S2(t) = \frac{\tau^2}{2} - -\left(\frac{T^2}{2} + \frac{T^3}{c}\right)e^{2h} - \frac{T^2c + 2T^3}{2c} + \left(T^2 - \frac{2T^3}{c} + T\tau + \frac{T^2\tau}{c}\right)e^h$$
$$+ \frac{\tau^2}{2\gamma^2} - T\tau - \frac{T^2\tau}{c} - \frac{\tau}{b} \tag{24.35}$$

$$S1(t) = \frac{\tau^3}{3} + \left(\frac{T^3}{2} + \frac{T^4}{c}\right)e^{2h} - \left(2T^2\tau - \frac{2T^4}{c} + \frac{2T^3\tau}{c}\right)e^h - \frac{\tau^3}{3\gamma^2} + T\tau^2$$
$$+ T^2\tau + \frac{T^4 + 2T^3\tau}{c} + \frac{T^2\tau^2}{c} + \frac{\tau^2}{b} + \frac{1}{a} \tag{24.36}$$

The feedback gains can now be obtained by inverting S and by using Eqs. (24.26) and (24.25). Due to the complexity of the results, this stage is best performed numerically, rather than symbolically.

For the special case $b = 0, c = 0$, we obtain (see Ben-Asher and Yaesh 1998)

$$u = N' V_c \dot{\sigma} - K N' n_L \tag{24.37}$$

The gain N' is defined as follows:

$$N' = \frac{6h^2(e^{-h} - 1 + h)}{\frac{6}{aT^3} + 2(1 - \gamma^{-2})h^3 + 3 + 6h - 6h^2 - 12he^{-h} - 3e^{-2h}} \tag{24.38}$$

and K is as in (24.9)

$$K = \frac{1}{h^2}(e^{-h} + h - 1) \tag{24.39}$$

Notice that Eq. (24.8) is obtained from (24.38) for the case $a \to \infty, \gamma \to \infty$ – a case of non-maneuvering target. When $T \to 0$ (ideal pursuers' dynamics; see Bryson and Ho 1975) one readily obtains from (24.38):

$$N' = \frac{6h^3}{\frac{6}{aT^3} + 2(1 - \gamma^{-2})h^3} = \frac{3\tau^3}{\frac{3}{a} + (1 - \gamma^{-2})\tau^3}; K = 0 \tag{24.40}$$

For $a \to \infty, \gamma \to \infty$, we recover Eq. (24.4), i.e., proportional navigation. Notice that for (24.38) and (24.40), with a given weighting factor a, there is a minimum value $\gamma = \gamma_{cr}$. Below this value, there is a finite escape time for the controls and the solution ceases to exist.

4.3 Numerical Results

In this section, we consider a numerical example that illustrates the merits of the differential-game guidance law. The time-constant T of a hit-to-kill (HTK) missile is taken to be 0.23 s and its maximal acceleration is 10 g. The conflict is assumed to be a nearly head on with a closing velocity of 200 m/s. We analyze the effect on the miss distance of a 1.5 g constant acceleration target maneuver which takes place at an arbitrary time-to-go between 0 and 4T (with 0.1 s time increments). The effect of four guidance laws will be analyzed:

1. **Proportional navigation (PN)** – Eq. (24.4)
2. **Optimal control guidance law (OGL)** – Eq. (24.8)
3. **Linear-quadratic differential game guidance law based on ideal dynamics (LQDG0)** – Eq. (24.40). In the implementation, we will employ the following parameters (with the appropriate units): $a = 10^6; \gamma = 1.5$

4. **Linear-quadratic differential game guidance law based on realistic dynamics (LQDG1)** – Eq. (24.38). In the implementation, we will employ the following parameters (with the appropriate units): $a = 10^6; \gamma = 7$

The histograms for all four cases are depicted in Fig. 24.2a, b. Table 24.1 summarizes the expected miss and the median (also known as a CEP – circular error probability). In practice, the latter is the more common performance measure.

Clearly the guidance laws that consider the time delay (OGL and LQDG1) outperform the guidance laws which neglect it (PN and LQDG0). The differential game

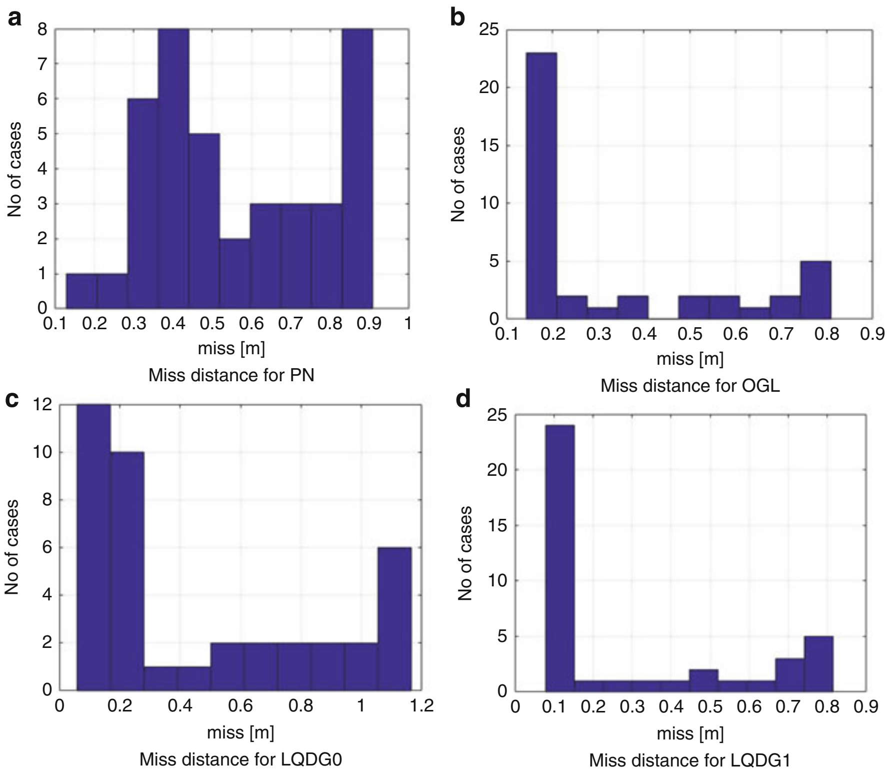


Fig. 24.2 Histograms of miss distance for four cases. (**a**) Miss distance for PN. (**b**) Miss distance for OGL. (**c**) Miss distance for LQDG0. (**d**) Miss distance for LQDG1

Table 24.1 Miss distance for various guidance law

Guidance laws	Expected miss	Median (CEP)
PN	0.564 m	0.513 m
OGL	0.339 m	0.177 m
LQDG0	0.477 m	0.226 m
LQDG1	0.290 m	0.102 m

approach further improves the performance under both assumptions, especially for the CEP performance measure.

5 Differential Games with Hard Bounds

This section is based on the works of Gutman and Shinar (Gutman 1979; Shinar 1981, 1989).

5.1 Problem Formulation

In this formulation, we make the following assumptions:

1. *The end game is two-dimensional and gravity is compensated independently.*
2. *The speeds of the pursuer (the missile) P and the evader (the target) E are constant during the end game (approximately true for short end games).*
3. *The trajectories of P and E can be linearized around their collision course.*
4. *Both missiles have and bounded lateral accelerations a_{Pm} and a_{Em}, respectively. The pursuer is more maneuverable than the evader ($a_{Pm} > a_{Em}$).*
5. *Maneuvering dynamics of both missiles are approximated by first-order transfer functions with the respective time constants T_E and T_P. The evader is more agile than the pursuer such that $T_E/a_{Em} < T_P/a_{Pm}$*
6. *The pursuer and the evader can measure their own accelerations in addition to the line-of-sight rate and have an estimate of the time-to-go.*

The problem then has the following state space representation

$$\dot{x} = Ax + Bu \cdot a_{P_m} + Dw \cdot a_{e_m} \tag{24.41}$$

$$A = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & -\frac{1}{T_p} & 0 \\ 0 & 0 & 0 & -\frac{1}{T_e} \end{bmatrix} \quad B = \begin{bmatrix} 0 \\ 0 \\ \frac{1}{T_p} \\ 0 \end{bmatrix} \quad D = \begin{bmatrix} 0 \\ 0 \\ 0 \\ \frac{1}{T_e} \end{bmatrix}$$

where x_3 is the pursuer's actual acceleration, x_4 is the evader's actual acceleration, and $u\,(|u| \le 1)$ and $w\,(|w| \le 1)$ are the commanded normal accelerations of the pursuer and the evader, respectively, which play as the adversaries in the following minimax problem:

$$\min_{|u|\le 1} \max_{|w|\le 1} J = |x_1(t_f)| = |E^T x(t_f)|$$

$$E^T = \begin{bmatrix} 1 & 0 & 0 & 0 \end{bmatrix} \tag{24.42}$$

The problem involves two nondimensional parameters of physical significance: the pursuer/evader maximum maneuverability ratio

$$\mu = \frac{a_{p_m}}{a_{e_m}} \tag{24.43}$$

and the ratio of evader/pursuer time constants

$$\varepsilon = \frac{T_e}{T_p} \tag{24.44}$$

Based on our assumptions, $\mu > 1$ and $\varepsilon\mu < 1$. The transformation

$$Z(t) = E^T \Phi(t_f, t) x(t) \tag{24.45}$$

where $\Phi(t_f, t)$ is the transition matrix of the original homogeneous system, reduces the vector equation (24.41) to a scalar one. It is easy to see from (24.42) and (24.45) that the new state variable is the *zero-effort miss distance*. We will use as an independent variable the normalized time-to-go

$$h = \frac{t_f - t}{T_p} \tag{24.46}$$

Define the normalized miss distance

$$Z(h) = \frac{Z(t)}{T_p^2 a_{E_m}} \tag{24.47}$$

We obtain

$$Z(h) = \left(x_1 + x_2 T_p h + x_3 T_E^2 \Psi(h/\varepsilon) - x_4 T_p^2 \Psi(h)\right) / T_p^2 a_{E_m} \tag{24.48}$$

where

$$\Psi(h) = \left(e^{-h} + h - 1\right) \tag{24.49}$$

Eq. (24.48) imbeds the assumption of perfect information, meaning that all the original state variables (x_1, x_2, as well as the lateral accelerations x_3 and x_4) are known to both players. Using the nondimensional variables, the normalized game dynamics become

$$\frac{dZ(h)}{dh} = \mu\Psi(h)u - \varepsilon\Psi(h/\varepsilon)w \tag{24.50}$$

The nondimensional payoff function of the game is the normalized miss distance

$$J = Z(0) \tag{24.51}$$

to be minimized by the pursuer and maximized by the evader.

5.2 Optimal Control Law

The Hamiltonian of the game is

$$H = \lambda_Z(\mu\Psi(h)u - \varepsilon\Psi(h/\varepsilon)w) \tag{24.52}$$

where λ_Z is the costate variable, satisfying the adjoint equation

$$\frac{d\lambda_Z}{dh} = -\frac{\partial H}{\partial Z} = 0 \tag{24.53}$$

and the transversality condition

$$\lambda_Z(0) = \left.\frac{\partial J}{\partial Z}\right|_{h=0} = \text{sign}(Z(0)) \tag{24.54}$$

The candidates for the optimal strategies are

$$u^* = -\text{sign}(\mu\Psi(h)Z(0))w^* = -\text{sign}(\varepsilon\Psi(h/\varepsilon)Z(0)) \tag{24.55}$$

Let us define

$$\Gamma(h) = -(\mu\Psi(h) - \varepsilon\Psi(h/\varepsilon)) \tag{24.56}$$

The state equation becomes:

$$\frac{dZ^*}{dh} = \Gamma(h)\text{sgn}(Z(0)) \tag{24.57}$$

$\Gamma(h)$ is a continuous function, and since $\mu > 1$ and $\varepsilon\mu < 1$, we get

$$\begin{aligned} \Gamma(h) &\approx (\mu - 1)h > 0 \ for\ h \gg 1 \\ \Gamma(h) &\approx (\mu\varepsilon - 1)\frac{h^2}{2\varepsilon} < 0 \ for\ h \ll 1 \end{aligned} \tag{24.58}$$

Thus, there is a solution to $\Gamma(h) = 0$ for sufficiently large h.

A family of regular optimal trajectories are generated by integrating (24.57) from $h = 0$. These trajectories do not cover the entire (Z, h). In fact, the game solution consists of decomposing the plane (Z, h) into two regions (see Fig. 24.3). In the regular region, D_1 the optimal strategies are of bang-bang type given by (24.55), and the value of the game is a unique function of the initial conditions. The boundaries of this region are the pair of optimal trajectories $Z^*_+(h)$ and $Z^*_-(h)$, reaching tangentially the h axis $(Z = 0)$, as shown in Fig. 24.3, at h_s which is the solution of the tangency equation $\Gamma(h) = 0$. In the other region D_0, enclosed by the boundaries $Z^*_+(h)$ and $Z^*_-(h)$, the optimal control strategies are arbitrary. All the trajectories

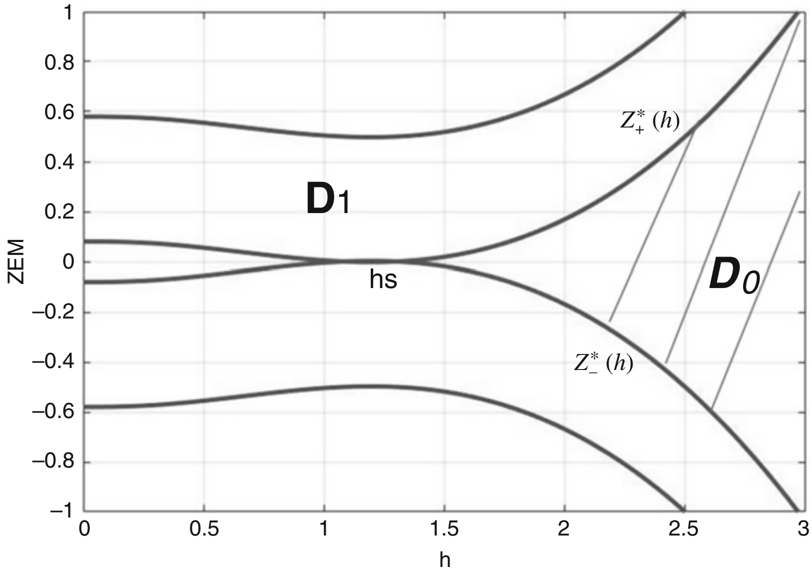


Fig. 24.3 Game space decomposition

starting in D_0 must go through the point $(Z = 0, h = h_s)$ where D_0 terminates. Therefore, the value of the game for the entire region of D_0 is the finite integral of $\Gamma(h)$ between $h = 0$ and $h = h_s$:

$$Miss = \left(\mu(1-\varepsilon)\Psi(h_s) - \frac{1}{2}(\mu - 1)h_s^2\right) \tag{24.59}$$

When a trajectory starting in D_0 reaches $(Z = 0,\ h = h_s)$, the evader must select the direction of its maximal maneuver (either to the right or to the left) and the pursuer has to follow it. Therefore, the optimal evasive maneuver that guarantees J^* is a maximal maneuver in a fixed direction for the duration of at least h_s.

The important case of ideal evader dynamics ($T_E = \varepsilon = 0$) was solved first by Gutman (1979). Note that as the pursuer in practice does not know T_E, it may be a worst-case assumption. More importantly, the evader's lateral acceleration is not required, because it becomes a control variable; hence, the term multiplying the actual lateral acceleration of the evader in Eq. (24.48) vanishes and the implementation of the guidance law becomes simpler. In this case by letting $\varepsilon \to 0$ in (24.56), we obtain

$$\Gamma(h) = -(\mu\Psi(h) - h) \tag{24.60}$$

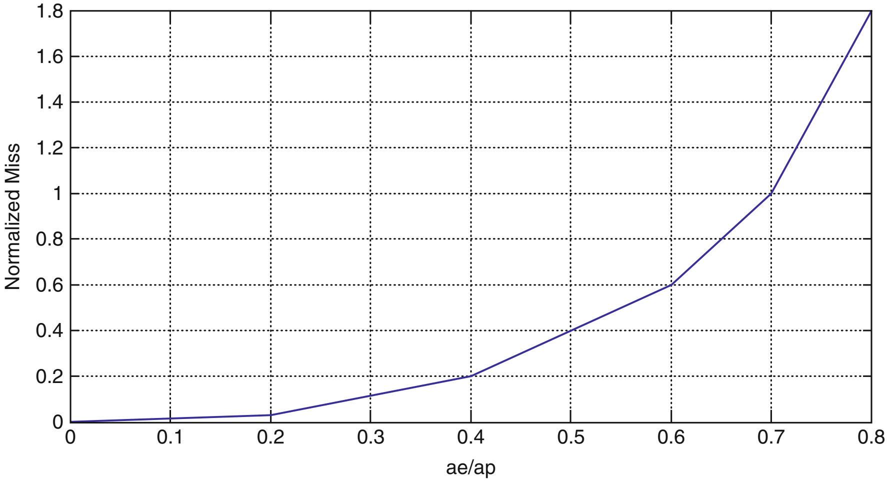


Fig. 24.4 Guaranteed miss distance

Thus solving $\Gamma(h) = 0$ leads to

$$h_s = \mu\Psi(h_s) \tag{24.61}$$

And from (24.58) and (24.61), the normalized miss is simply

$$Miss = \left(h_s - \frac{1}{2}(\mu - 1)h_s^2\right) \tag{24.62}$$

Using the numerical example of the last section ($\mu = 6.667, T_p = 0.23s, a_E = 15$), we solve (24.61) to obtain $h_s \approx 0.335$. With this value, we achieve from (24.62) a miss of about 1.4 cm, i.e., nearly perfect hit. The values of the guaranteed normalized miss distance for other values of the maneuverability ratio are depicted in Fig. 24.4.

6 Trajectory Shaping in Linear-Quadratic Games

6.1 Objectives

An advantage of the linear-quadratic differential game (LQDG) formulation is its flexibility, which enables it not only to include in the cost function additional weights on other terminal variables but also to introduce some "trajectory shaping" by augmenting the cost function with a running-cost (quadratic-integral) term on the state variables. This term affects the trajectory and minimizes its deviations from zero. In Ben-Asher et al. (2004), it was discovered that the trajectory-shaping term also leads to attenuation of the disturbances created by random maneuvering of the evader.

6.2 Problem Formulation

Using the model of Sect. 2, we formulate the following minimax problem:

$$\dot{x} = Ax + Bu + Dw$$

$$A = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & -\frac{1}{T} \end{bmatrix} B = \begin{bmatrix} 0 \\ 0 \\ \frac{1}{T} \end{bmatrix} D = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} \tag{24.63}$$

$$\min_u \max_w J = \frac{b}{2} x_1^2(t_f) + \frac{1}{2} \int_0^{t_f} \left[x^T(t) Q x(t) + u^2(t) - \gamma^2 w^2(t) \right] dt \tag{24.64}$$

Q is a semi-positive-definite trajectory-shaping penalty matrix that penalizes the state deviation from the nominal collision path. For simplicity, we restrict this case to the following form:

$$Q = \begin{bmatrix} q & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \tag{24.65}$$

Other forms of Q may also be of interest but are beyond the scope of our treatment here.

6.3 Optimal Solution

The optimal pursuer strategy includes state feedback of the form (see Sect. 6 below)

$$u = -B^T P x \tag{24.66}$$

where

$$-\dot{P} = A^T P + PA - PB^T BP + \gamma^{-2} PD^T DP + Q \tag{24.67}$$

$$P(t_f) = \begin{bmatrix} b & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \tag{24.68}$$

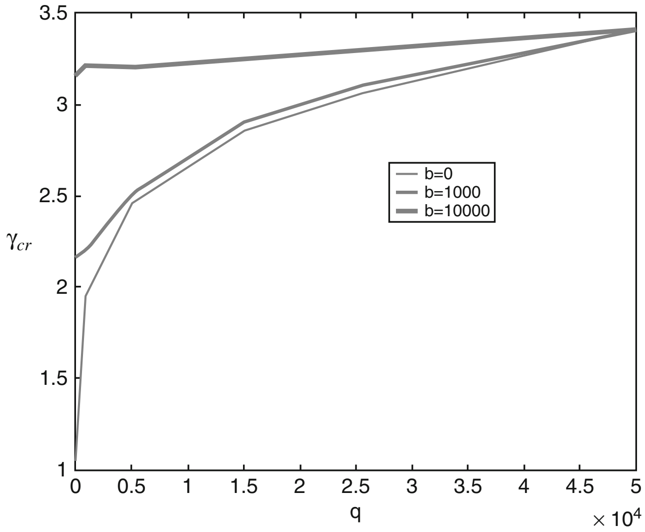


Fig. 24.5 Critical γ

Notice that for a given interval $[0, t_f]$, there exists a (minimal) γ_{cr} such that for $\gamma \le \gamma_{cr}$ the solution does exist (finite escape time). The search for this critical value is the subject of the next subsection.

6.4 Asymptotic Values of γ_{cr}

For the sequel, we will consider an example with $T = 0.25$ s, $t_f = 4$ s. Figure 24.5 depicts γ_{cr} as a function of q for three values of b, namely, 0, 1000, 10,000 as obtained by searching the values of γ where the solution for the DRE (24.67) ceases to exist (finite escape times). There are two obvious limit cases for our problem, namely, $q = 0$ and $b = 0$. The first case, $q = 0$, has the closed-form solution given in Sect. 2. Thus, we require that the positive roots of the function

$$\frac{6}{bT^3} + 2(1 - \gamma^{-2})h^3 + 3 + 6h - 6h^2 - 12he^{-h} - 3e^{-2h} \tag{24.69}$$

should not lie in $[0, t_f/T]$ (recall that h is the normalized time-to-go). Adding the trajectory-shaping term with positive q to the cost J would require a higher γ_{cr}. The second limit case results from Theorem 4.8 of Basar and Bernhard (1995) which states that the problem (24.63) and (24.64) with b = 0 has a solution if and only if the corresponding algebraic Riccati equation (ARE):

$$0 = A^T P + PA - PB^T BP + \gamma^{-2} PD^T DP + Q \tag{24.70}$$

has a nonnegative definite solution. Thus, the intersection in Fig. 24.5 with the γ_{cr} axis (i.e., $q = 0$) is also the critical result of Eq. (24.69) when its positive roots lie in the game duration. On the other hand, the values for $b = 0$ (lower curve) coincide with the γ_{cr} values of the ARE (Eq. (24.70)), the minimal values with nonnegative solutions. As expected, increasing b and/or q (positive parameters in the cost) results with monotone increase of γ_{cr}. Notice the interesting asymptotic phenomenon that for very high q, the γ_{cr} values approach the AREs. The above observations can help us to estimate lower and upper bounds for γ_{cr}. For a given problem – formulated with a given pair (b, q) – we can first solve Eq. (24.38) to find the critical values for the case $q = 0$. This will provide a lower bound for the problem. We then can solve the ARE with very high q ($q \gg \max(q, b)$). The critical value of the ARE problem provides an estimate for the upper value of γ_{cr}. Having obtained lower and upper estimates, one should search in the relevant segment and solve the DRE (Eq. (24.67)) with varying γ until a finite escape time occurs within the game duration.

6.5 Numerical Results

In this section, we consider a numerical example that illustrates the merits of the differential-game guidance law with those of trajectory shaping. The time-constant

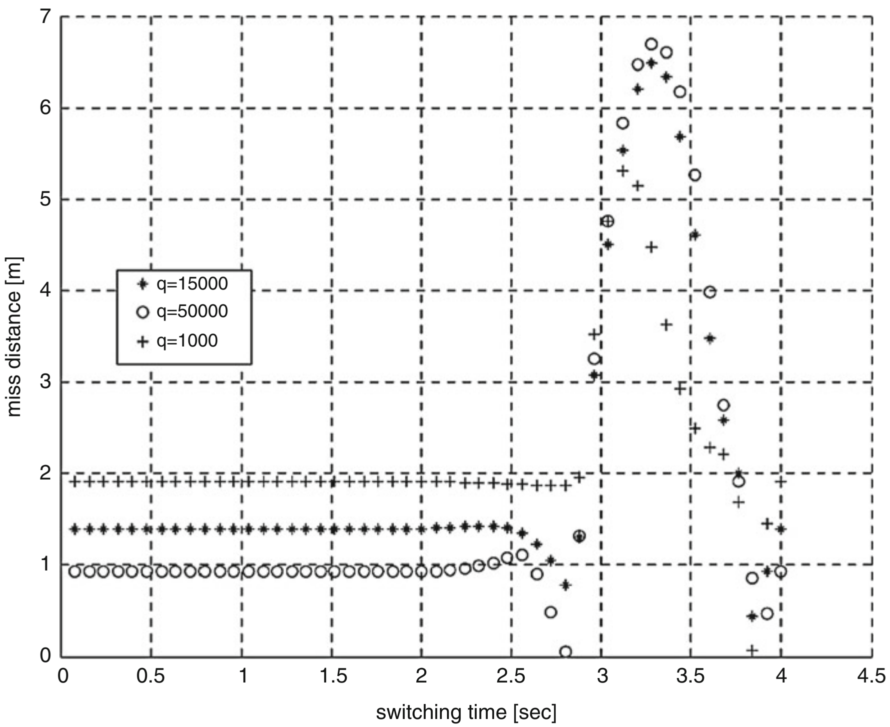


Fig. 24.6 Miss distance for LQDG with trajectory shaping

T of the missile is taken to be 0.25 s and its maximal acceleration is 15 g. The conflict is assumed to be a nearly head on and to last 4 s, with a closing velocity of 1500 m/s. We analyze the effect on the miss distance of a 7.5 g constant acceleration target maneuver which takes place at an arbitrary time. Because it is not advisable to work at (or very close to) the true conjugate value of Fig. 24.5, we use $\gamma = \gamma_{cr} + 0.1$. The effect of the parameter q is shown in Fig. 24.6 which presents the miss distance as function of the target maneuvering time. As q increases, smaller miss distances are obtained, up to a certain value of q for which the miss distance approaches a minimal value almost independent of the switch point in the major (initial) part of the end game. Larger miss distances are obtained only for a limited interval of switch points $t_{go} \in [T, 3T]$. Figure 24.7 compares the miss distances of the linear-quadratic game (LQDG) with the bounded-control game (DGL) of Sect. 5. The pursuer control in DGL is obtained by using the discontinuous control:

$$u = u_{max}\text{sign}(Z(t_f)) \tag{24.71}$$

$Z(t_f)$ is

$$Z(h) = (x_1 + x_2 T_p h + x_3 T^2 \Psi(h)) \tag{24.72}$$

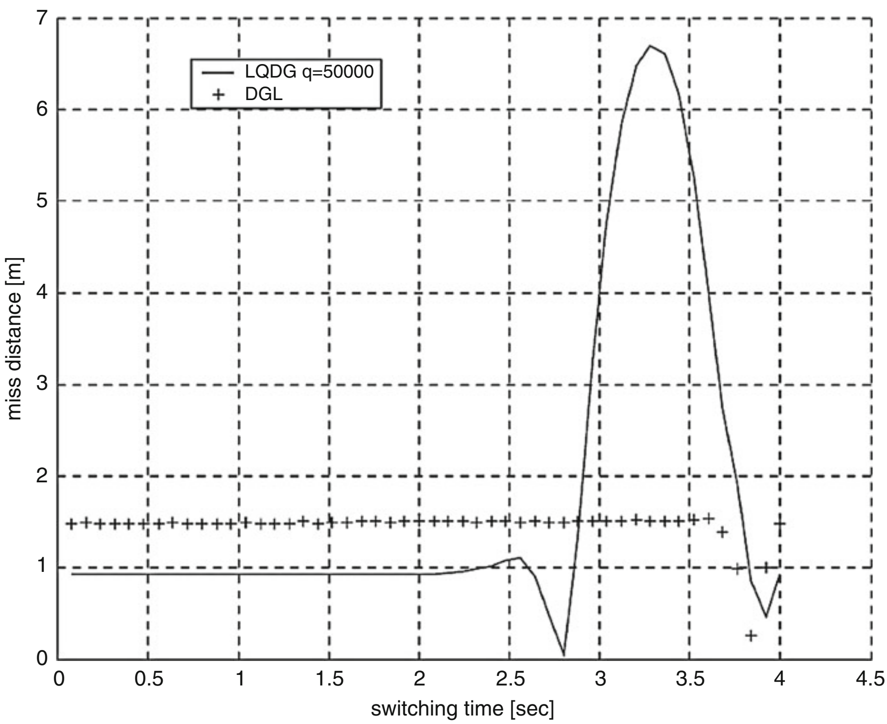


Fig. 24.7 Miss distance for LQDG and hard bounded DG

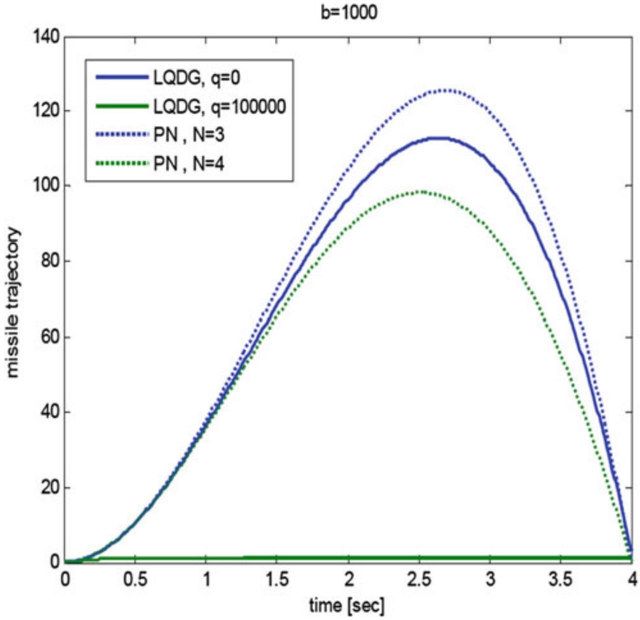


Fig. 24.8 Trajectories

where, as before

$$\Psi(h) = (e^{-h} + h - 1) \tag{24.73}$$

For most cases, the results of LQDG are slightly better. Only for evasive maneuvers that take place in the time frame $t_f - t = T$ to $t_f - t = 3T$ is the performance of the bounded-control game solution superior. Assuming a uniformly distributed t_s switch, between 0 and 4 s (as we did in Sect. 2), we get average miss distances of 1.4 and 1.6 m for LQDG and DGL, respectively.

A heuristic explanation for the contribution of the trajectory-shaping term is as follows. Figures 24.8 and 24.9 are representative results for a 9 g target at $t = 0$ (the missile has a 30 g maneuverability). PN guidance and the classical linear-quadratic game (q = 0; Sect. 3) avoid maneuvering at early stages because the integral of the control term is negatively affected and deferring the maneuver is profitable. It trades off early control effort for terminal miss. Adding the new term (with $q = 10^5$) forces the missile to react earlier to evasive maneuvers at the expense of a larger control effort, in order to remain closer to the collision course. This, in fact, is the underlying philosophy of the hard-bound differential-game approach that counteracts the instantaneous zero-effort miss.

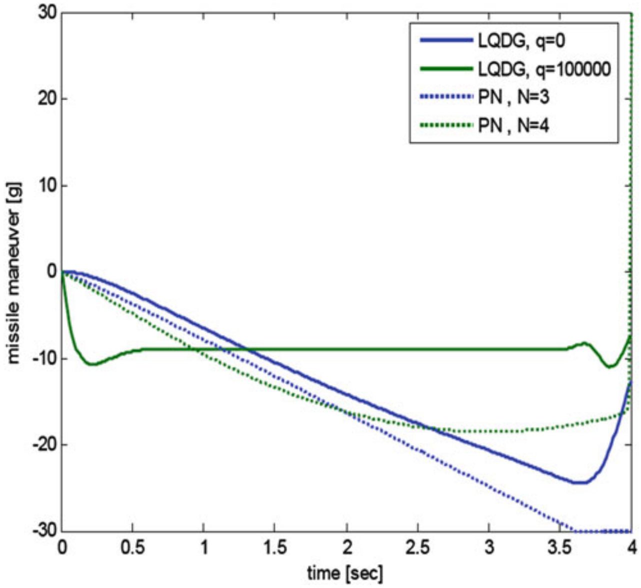


Fig. 24.9 Controls

7 Disturbance Attenuation Approach to Missile Guidance with Measurement Uncertainty

In this section, we extend the linear-quadratic-game problem to include uncertainty in the measurement. This is done by constructing a disturbance attenuation function, which is converted into an indefinite cost criterion to be minimized by the control and maximized by the measurement uncertainty, the process uncertainty, which act like an intelligent adversary, and the initial conditions (Basar and Bernhard 1995; Rhee and Speyer 1991; Speyer and Jacobson 2010). The original formulation was using the linear-quadratic-Gaussian problem Speyer (1976) and presented in a general form in Speyer and Chung (2008), whose solutions reduce to the same linear-quadratic-game problem that includes uncertainty in the measurement. This formulation generalizes the results of Sect. 3 and allows greater latitude in the guidance law design. We first present the general disturbance attenuation problem and then specialize it to the missile guidance problem.

7.1 The General Disturbance Attenuation Problem

The objective is to design a compensator based only on the measurement history, such that the transmission from the disturbances to the performance outputs are limited in some sense. To make these statements more explicit, consider the dynamical system

$$\dot{x} = Ax + Bu + \Gamma w, \quad x(t_0) = x_o, \tag{24.74}$$

$$z = Hx + v, \tag{24.75}$$

where $z \in \mathcal{R}^q$ is the measurement, w is the process disturbance error, v is the measurement disturbance error, and x_0 is an unknown initial condition. The performance outputs are measures of desired system performance, such as good tracking error or low actuation inputs to avoid saturation. The general measure can be written as

$$y = Cx + Du, \tag{24.76}$$

where $y \in \mathbb{R}^p$.

A general representation of the input-output relationship between disturbances (v, w, x_0) and output performance measure y is the disturbance attenuation function

$$D_a = \frac{\|y\|_2^2}{\|\tilde{w}\|_2^2}, \tag{24.77}$$

where

$$\|y\|_2^2 \triangleq \frac{1}{2}\left[x^T(t_f)Q_f x^T(t_f) + \int_{t_0}^{t_f} (x^T Qx + u^T Ru)dt\right], \tag{24.78}$$

$$\|y\|^2 \triangleq y^T y = x^T C^T Cx + u^T D^T Du, \;\; C^T C = Q, \;\; C^T D = 0, \;\; D^T D = R, \tag{24.79}$$

and $\tilde{w} = [w^T, v^T, x_0^T]^T$, with

$$\|\tilde{w}\|_2^2 = \frac{1}{2}\left[\int_{t_0}^{t_f} (w^T W^{-1} w + v^T V^{-1} v)dt + x_0^T P_0^{-1} x_0\right]. \tag{24.80}$$

In this formulation, the disturbances w and v are not dependent (thus, the cross terms in Eq. (24.80) are zero), and W and V are the associated weighings, respectively, that represent the spectral densities of the disturbances.

The disturbance attenuation problem is to find a controller $u = u(Z_t) \in \mathcal{U}$ where the measurement history $Z_t = \{z(s) : 0 \leq s \leq t\}$ so that the disturbance attenuation problem is bounded as

$$D_a \leq \gamma^2, \tag{24.81}$$

for all admissible sequences of w and v, $x_0 \epsilon \mathcal{R}^n$. The choice of γ^2 cannot be completely arbitrary. There exists a γ_c^2 where if $\gamma^2 \leq \gamma_c^2$, the solution to the problem, has a finite escape time. This problem is converted to a differential game problem with performance index obtained from manipulating Eqs. (24.77) and (24.81) as

$$J(u, \tilde{w} : t_0, t_f) = \|y\|_2^2 - \gamma^2 \|\tilde{w}\|_2^2. \tag{24.82}$$

For convenience, define a process for a function $\hat{w}(t)$ as

$$\hat{w}_a^b \triangleq \{\hat{w}(t) : a \leq t \leq b\}. \tag{24.83}$$

The differential game is then to find the minimax solution as

$$J^\circ(u^\circ, \tilde{w}^\circ : t_0, t_f) = \min_{u_{t_0}^{t_f}} \max_{w_{t_0}^{t_f}, v_{t_0}^{t_f}, x_0} J(u, \tilde{w} : t_0, t_f). \tag{24.84}$$

Assume that the min and max operations are interchangeable. Let t be the "current" time. This problem is solved by dividing the problem into a future part, $\tau > t$, and past part, $\tau < t$, and joining them together with a connection condition. Therefore, expand Eq. (24.84) as

$$J^\circ(u^\circ, \tilde{w}^\circ : t_0, t_f) = \min_{u_{t_0}^{t}} \max_{w_{t_0}^{t}, v_{t_0}^{t}, x_0} \left[J(u, \tilde{w} : t_0, t) + \min_{u_t^{t_f}} \max_{w_t^{t_f}, v_t^{t_f}} J(u, \tilde{w} : t, t_f) \right]. \tag{24.85}$$

Note that for the future time interval no measurements are available. Therefore, minimizing with respect to $v_t^{t_f}$, given the form of the performance index (24.82), produces the worst future process for $v_t^{t_f}$ as

$$v(\tau) = 0, \qquad t < \tau \leq t_f. \tag{24.86}$$

Therefore, the game problem associated with the future reduces to a game between only u and w. The controller of Eq. (24.25) is

$$u^\circ(t) = -R^{-1} B^T S_G(t_f, t; Q_f) x(t), \tag{24.87}$$

where x is not known and $S_G(t_f, t; Q_f)$ is propagated in (24.90) and is also given in (24.23). The objective is to determine x as a function of the measurement history by solving the problem associated with the past, where t is the current time and $t = t_0$ is the initial time.

The optimal controller $u = u(Z_t)$ is now written as

$$u^\circ = -R^{-1} B^T S_G(t_f, t; Q_f)(I - \gamma^{-2} P(t_0, t; P_0) S_G(t_f, t; Q_f))^{-1} \hat{x}(t) = \Lambda(t)\hat{x}(t). \tag{24.88}$$

where

$$\dot{\hat{x}}(t) = A(t)\hat{x}(t) + B(t)u(t) + \gamma^{-2} P(t_0, t; P_0) Q(t) \hat{x}(t)$$

$$+ P(t_0, t; P_0)H^T(t)V^{-1}(t)(z(t) - H(t)\hat{x}(t)), \quad \hat{x}(t_0) = 0, \qquad (24.89)$$

$S_G(t_f, t; Q_f)$ is propagated backward by the matrix Riccati differential equation (RDE)

$$\begin{aligned} -\dot{S}_G(t_f, t; Q_f) &= Q(t) + S_G(t_f, t; Q_f)A(t) + A^T(t)S_G(t_f, t; Q_f) \\ &-S_G(t_f, t; Q_f)(B(t)R^{-1}(t)B^T(t) - \gamma^{-2}\Gamma(t)W(t)\Gamma^T(t))S_G(t_f, t; Q_f), \\ S_G(t_f, t_f; Q_f) &= Q_f, \end{aligned} \qquad (24.90)$$

and $P(t_0, t; P_0)$ is propagated forward

$$\begin{aligned} \dot{P}(t_0, t; P_0) = {} & A(t)P(t_0, t; P_0) + P(t_0, t; P_0)A^T(t) + \Gamma(t)W(t)\Gamma^T(t) \\ & -P(t_0, t; P_0)(H^T(t)V^{-1}(t)H(t) - \gamma^{-2}Q(t))P(t_0, t; P_0), \\ & P(t_0, t_0; P_0) = P_0, \end{aligned} \qquad (24.91)$$

The controller (24.88) is optimal if and only if

1. There exists a solution $P(t_0, t; P_0)$ to the RDE (24.91) over the interval $[t_0, t_f]$.
2. There exists a solution $S_G(t_f, t; Q_f)$ to the RDE (24.90) over the interval $[t_0, t_f]$.
3. $P^{-1}(t_0, t; P_0) - \gamma^{-2}S_G(t_f, t; Q_f) > 0$ over the interval $[t_0, t_f]$.

The third condition is known as the spectral radius condition.

7.2 Disturbance Attenuation Controller Specialized to Missile Guidance

As considered in Sect. 4.1 in the cost criterion (24.22), there is no weighting on the state except at the terminal time, i.e., $Q(t) = 0$. As shown in Speyer (1976), the guidance gains can be determined in closed form, since the RDE (24.90), which is solved backward in time, can be determined in closed form. Therefore, the guidance gains are similar to those of (24.25) where $S_G(t_f, t; Q_f)$, which is propagated by (24.90) and is similar to (24.23), and the filter in (24.89) reduces to the Kalman filter, where $P(t_0, t; P_0)$ in (24.91) is the variance of the error in the state estimate. However, the worst state, $(I - \gamma^{-2}P(t_0, t; P_0)S_G(t_f, t; Q_f))^{-1}\hat{x}(t)$, rather than just the state estimate, $\hat{x}(t)$, operating on the guidance gains (24.88), can have a significant effect, depending on the value chosen for γ^{-2}.

It has been found useful in designing homing guidance systems to idealize the equations of motion of a symmetric missile to lie in a plane. In this space using very simple linear dynamics, the linear quadratic problem produces a controller equivalent to the proportional navigation guidance law of Sect. 4, used in most homing guidance systems. The geometry for this homing problem is given in Fig. 24.1 where the missile pursuer is to intercept the target.

The guidance system includes an active radar and gyro instrumentation for measuring the line-of-sight (LOS) angle σ, a digital computer, and autopilot instrumentation such as accelerometers, gyros, etc. The lateral motion of the missile and target perpendicular to the initial LOS is idealized by the following four state models using (24.19) and (24.20):

$$\begin{bmatrix} \dot{x}_1 \\ \dot{x}_2 \\ \dot{n}_T \\ \dot{A}_M \end{bmatrix} = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & -1 \\ 0 & 0 & -2a_T & 0 \\ 0 & 0 & 0 & -W_1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ n_T \\ A_M \end{bmatrix} + \begin{bmatrix} 0 \\ W_1/W_z \\ 0 \\ W_1(1 + W_1/W_z) \end{bmatrix} u + \begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \end{bmatrix} w \quad (24.92)$$

where the state $x = \begin{bmatrix} x_1 & x_2 & n_T & A_M \end{bmatrix}^T$, with x_1 is the lateral position perpendicular to the initial LOS, x_2 is the relative lateral velocity, n_T is the target acceleration normal to the initial LOS, A_M is the missile acceleration normal to the initial LOS, u is the missile acceleration command normal to the initial LOS, W_1 is the dominant pole location of the autopilot, W_z is the non-minimal phase zero location of the autopilot, $2a_T$ is the assumed target lag, and w is the adversary command weighted in the cost criterion by W. The missile autopilot and target acceleration are approximately included in the guidance formulation in order to include their effect in the guidance gains. Therefore, (24.92) is represented by (24.75). Note that (24.92) includes a zero and a lag, somewhat extending (24.21), which has only the lag.

The measurement sequence, for small line-of-sight angles, is approximately

$$y = \sigma + v \approx x_1/V_c\tau_g + v \quad (24.93)$$

where V_c is the closing velocity, $\tau_g = t_f - t$ is the time-to-go, and v is measurement noise with weighting V. The weighting V may not be known a priori and may be estimated online. v is the noise into the radar and is composed of scintillation, glint, receiver, and jamming noise. $V_c\tau_g$ is estimated online by using an active radar. Therefore, (24.92) and (24.93) are represented by (24.75).

A useful gain to calculate is the navigation ratio (NR) which operates on the estimated line-of-sight rate $\hat{\dot{\sigma}}$. The gains are the row vector Λ in (24.88) where the scalar R is renamed $\bar{\gamma}$. This navigation ratio is calculated directly from the gain Λ_{x_1} operating on $\hat{x}_1$

$$NR = \tau_g^2 \Lambda_{x_1}. \quad (24.94)$$

Note that $\Lambda_{x_1} = \Lambda_{x_2}/\tau_g$ where Λ_{x_2} is the gain operating on $\hat{\dot{x}}_1 = \hat{x}_2$ and $\hat{\dot{\sigma}} = -\hat{x}_1/V_c\tau_g^2 + \hat{x}_2/V_c\tau_g$. The guidance law, sometimes called "biased proportional navigation," is

$$u = NRV_c\hat{\dot{\sigma}} + \Lambda_{n_T}\hat{n} + \Lambda_{A_M} A_M. \quad (24.95)$$

where $\hat{\dot{\sigma}}$ is the estimated line-of-sight rate, Λ_{n_T} is the gain on target acceleration, and Λ_{A_M} is the gain on missile acceleration. The missile acceleration A_M is not obtained from the Kalman filter but directly from the accelerometer on the missile and is assumed to be very accurate. Note that the guidance law (24.95) for the system with disturbances generalizes the deterministic law given in (24.37).

7.3 Numerical Results

The results reported here were first reported in Speyer (1976) for the guidance law determined by solving the linear exponential Gaussian problem. The guidance law given by (24.95) was programmed into a high-fidelity digital simulation. This nonlinear digital simulation models in three dimensions a realistic engagement between missile and target. The parameter γ^2 is chosen so that the gains of the disturbance attenuation guidance scheme are close to those of the baseline guidance gains designed on LQG theory when there is no jamming noise entering the active radar of the missile. This cannot be made precise because of the presence of scintillation, fading, and radar receiver noise. Intuitively, the advantage of the adaptive guidance is that during engagements when the missile has a large energy advantage over the target, the disturbance attenuation formulation is used ($\gamma^2 > 0$). This allows the gains to increase over the baseline gain when jamming noise is present since the measurement variance will increase. This variance is estimated online Speyer (1976). With increased gains, a trade-off is made between control effort and decreased terminal variance of the miss. The objective of the guidance scheme is to arrive at a small miss distance. Since this is tested on a nonlinear simulation, the added control effort is seen to bring in the tails of the distribution of the terminal miss distance as shown by Monte Carlo analysis.

All shots were against a standard target maneuver of about 1.25 gs at an altitude (H) of about 24.0 km and at 15 km down range (R). In Table 24.2, a comparison is made between relative miss distances of the baseline (LQG) and the disturbance attenuation guidance with the guidance based on the disturbance attenuation in the form of a ratio with the LQG miss distances in the numerator. The relative misses between the LQG and the disturbance attenuation guidance laws for 40, 66, and 90 % of the runs are found to lie below 1.0 as well as some of the relative largest misses.

That is, with a large jamming power of 50,000 W and a Monte Carlo average of 50 runs, it is seen that the baseline performs slightly better than the new controller. However, for a few runs, large misses were experienced by the baseline guidance (cf. Table 24.2); the worst having a relative miss of 4.13. Placed in a standard jamming power environment of 2500 W, over 25 runs the new controller and the baseline are almost identical although the relative maximum miss of the baseline to the new controller is 0.58. This relative miss by the disturbance attenuation controller is smaller than the largest miss experienced by the baseline guidance which falls within the 90th percentile when large jamming power occurs. The important

Table 24.2 Ratio of miss distances between baseline LQG controller and disturbance attenuation controller

	High jamming power	Standard jamming power
Percentile	(50,000 W)	(2500 W)
40th	0.88	0.98
66th	0.80	1.01
90th	0.99	0.92
Largest misses	4.13, 1.63, 1.13, 1.06	0.58

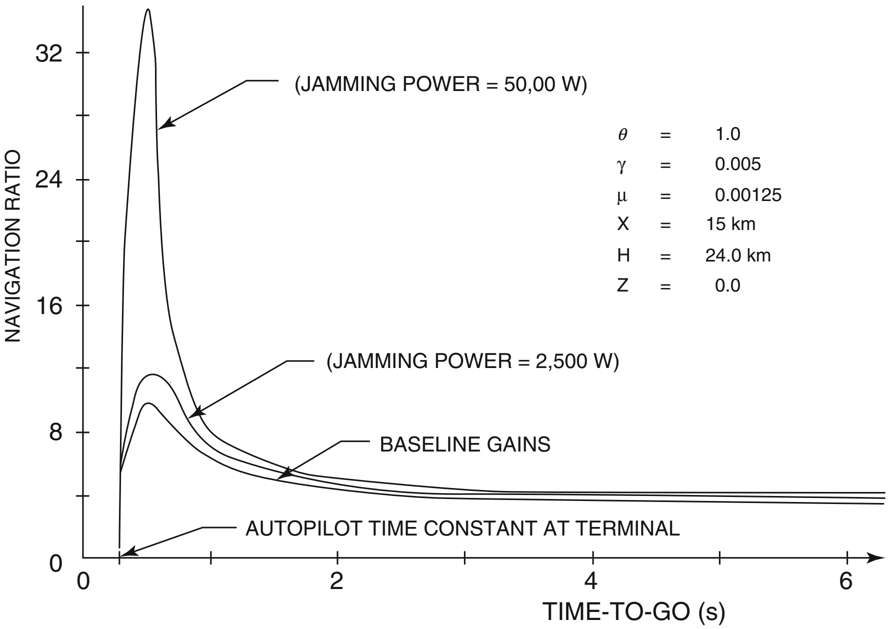


Fig. 24.10 Navigation ratio using disturbance attenuation gains

contribution of the disturbance attenuation controller is in its ability to pull in the tails of the miss distance distribution especially for high jamming while achieving similar performance as the baseline under standard jamming. The variations in the navigation ratio (NR) due to different jamming environments is plotted for a typical run in Fig. 24.10 against time-to-go (τ_g). This gain increases proportionally to the jamming noise environment which is estimated online. The LQG gain forms a lower bound on the disturbance attenuation gain. Note that the navigation ratio increases even for large τ_g as the jamming power increases. The baseline gains are essentially the adaptive gains when no jamming noise is present. Fine tuning the disturbance attenuation gain is quite critical since negative gain can occur if P is too large because the spectral radius condition, $P^{-1}(t_0, t; P_0) - \gamma^{-2} S_G(t_f, t; Q_f) > 0$, can be violated. Bounds on the largest acceptable P should be imposed. Note that the gains below the terminal time constant of the autopilot (see Fig. 24.10) are not effective.

8 Conclusions

Missile guidance appears to be the premier example in aerospace where differential games has been an enabling technology. Presented here are algorithms for the real-time application of differential games with important physical constraints. The homing missile problem is only approximately solved. Time-to-go estimates are heuristic, and system observability is dependent on the trajectory. In essence, the homing missile guidance problem is a dual-control differential game. Approximately solving this problem will allow much new innovation.

Acknowledgements We thank Dr. Maya Dobrovinsky from Elbit Systems for converting part of this document from Word to LaTeX.

References

Basar T, Bernhard P (1995) H_∞ optimal control and related minimax design problems: a dynamic game approach, 1st edn. Birkhäuser, Boston

Ben-Asher JZ, Yaesh I (1998) Advances in missile guidance theory, vol 180. Progress in astronautics and aeronautics. AIAA, Washington, DC

Ben-Asher JZ, Shinar Y, Wiess H, Levinson S (2004) Trajectory shaping in linear-quadratic pursuit-evasion games. AIAA J. Guidance, Control and Dynamics 27(6):1102–1105

Bryson AE, Ho YC (1975) Applied optimal control. Hemisphere, New York

Gutman S (1979) On optimal guidance for homing missiles. AIAA J. Guidance, Control and Dynamics 3(4):296–300

Nesline WN, Zarchan P (1981) A new look at classical vs. modern homing missile guidance. AIAA J. Guidance, Control and Dynamics 4(1):878–880

Rhee I, Speyer JL (1991) Game theoretic approach to a finite-time disturbance attenuation problem. IEEE Trans Autom Control 36(9):1021–1032

Shinar J (1981) In: Leondes CT (ed) Automatic control systems, vol 17, 8th edn. Academic Press, New York

Shinar J (1989) On new concepts effecting guidance law synthesis of future interceptor missiles. In: AIAA CP 89-3589, guidance, navigation and control conference, Boston. AIAA

Speyer JL (1976) An adaptive terminal guidance scheme based on an exponential cost criterion with application to homing missile guidance. IEEE Trans Autom Control 48:371–375

Speyer JL, Chung WH (2008) Stochastic processes, estimation, and control, 1st edn. SIAM, Philadelphia

Speyer JL, Jacobson DH (2010) Primer on optimal control theory, 1st edn. SIAM, Philadelphia

Zarchan P (1997) Tactical and strategic missile guidance, progress in astronautics and aeronautics. AIAA, Washington, DC

Stackelberg Routing on Parallel Transportation Networks

25

Walid Krichene, Jack D. Reilly, Saurabh Amin, and Alexandre M. Bayen

Contents

W. Krichene (✉) · A. Bayen
University of California at Berkeley, Berkeley, CA, USA
e-mail: walid@eecs.berkeley.edu; bayen@berkeley.edu

J. Reilly
Google Inc. Mountain View, CA, USA
e-mail: jackdreilly@gmail.com

S. Amin
MIT, Cambridge, MA, USA
e-mail: amins@mit.edu

T. Başar, G. Zaccour (eds.), *Handbook of Dynamic Game Theory*,
https://doi.org/10.1007/978-3-319-44374-4_26

Abstract

This chapter presents a game theoretic framework for studying Stackelberg routing games on parallel transportation networks. A new class of latency functions is introduced to model congestion due to the formation of physical queues, inspired from the fundamental diagram of traffic. For this new class, some results from the classical congestion games literature (in which latency is assumed to be a nondecreasing function of the flow) do not hold. A characterization of Nash equilibria is given, and it is shown, in particular, that there may exist multiple equilibria that have different total costs. A simple polynomial-time algorithm is provided, for computing the *best Nash equilibrium*, i.e., the one which achieves minimal total cost. In the Stackelberg routing game, a central authority (leader) is assumed to have control over a fraction of the flow on the network (*compliant flow*), and the remaining flow responds selfishly. The leader seeks to route the compliant flow in order to minimize the total cost. A simple Stackelberg strategy, the non-compliant first (NCF) strategy, is introduced, which can be computed in polynomial time, and it is shown to be optimal for this new class of latency on parallel networks. This work is applied to modeling and simulating congestion mitigation on transportation networks, in which a coordinator (traffic management agency) can choose to route a fraction of compliant drivers, while the rest of the drivers choose their routes selfishly.

Keywords

Transportation networks · Non-atomic routing game · Stackelberg routing game · Nash equilibrium · Fundamental diagram of traffic · Price of stability

1 Introduction

1.1 Motivation and Related Work

Routing games model the interaction between players on a network, where the cost for each player on an edge depends on the total congestion of that edge. Extensive work has been dedicated to the study of Nash equilibria for routing games (or Wardrop equilibria in the transportation literature, Wardrop 1952), in which players selfishly choose the routes that minimize their individual costs (latencies) (Beckmann et al. 1956; Dafermos 1980; Dafermos and Sparrow 1969). In general, Nash equilibria are inefficient compared to a system optimal assignment that minimizes the total cost on the network (Koutsoupias and Papadimitriou 1999). This inefficiency has been characterized for different classes of latency functions and network topologies (Roughgarden and Tardos 2004; Swamy 2007). This helps understand the inefficiencies caused by congestion in communication networks and transportation networks. In order to reduce the inefficiencies due to selfish routing, many instruments have been studied, including congestion pricing (Farokhi and Johansson 2015; Ozdaglar and Srikant 2007), capacity allocation (Korilis et al.

1997b), and Stackelberg routing (Aswani and Tomlin 2011; Korilis et al. 1997a; Roughgarden 2001; Swamy 2007).

1.1.1 Online Learning and Decision Dynamics in the Routing Game

The Nash equilibrium concept gives a characterization of the state of a network at equilibrium but does not specify how players arrive to the equilibrium. The study of decision dynamics provides an answer to this question and has been a fundamental topic in economics (Blume 1993), game theory (Shamma 2015; Weibull 1997), and online learning theory (Cesa-Bianchi and Lugosi 2006). These models usually assume that the game is played repeatedly (as opposed to a one-shot game) and that each player faces a sequential decision problem: At each iteration, the player takes an action and observes an outcome (which is also affected by the decisions of other players). The player can then use the outcome to update her decision on the next iteration. One of the natural questions that can be studied is whether the joint player decisions converge to an invariant set (typically, the Nash equilibrium of the one-shot game, or some other equilibrium concept). This question has a long history that dates back to Hannan (1957) who defined the regret and Blackwell (1956) who defined approachability, which became essential tools in the modeling and analysis of repeated games and convergence of player dynamics.

Decision dynamics have since been studied for several classes of games, such as potential games (Monderer and Shapley 1996), and many results provide convergence guarantees under different classes of decision dynamics (Benaïm 2015; Fox and Shamma 2013; Hofbauer and Sandholm 2009; Sandholm 2001). Although we do not study decision dynamics in this chapter, we review some of the work most relevant to routing games.

Routing games are a special case of potential games (Sandholm 2010), and decisions dynamics have been studied in the context of routing games: Blum et al. (2006) study general no-regret dynamics, Kleinberg et al. (2009) and Krichene et al. (2015a,b) study other classes of dynamics for which they give stronger convergence guarantees, and Fischer et al. (2010) studies a similar, sampling-based model. Several of these results relate the discrete algorithm to a continuous-time limit known as the replicator ODE, which is well studied in evolutionary game theory in general (Weibull 1997), and in routing games in particular (Drighès et al. 2014; Fischer and Vöcking 2004). Several studies build on these models of decision dynamics, to pose and solve estimation and control problems, such as estimating the latency functions on the network (Thai et al. 2015), estimating the learning rates of the dynamics (Lam et al. 2016), and solving optimal routing under selfish response (Krichene et al. 2016).

1.1.2 Stackelberg Routing Games

In the Stackelberg routing game, a subset of the players, corresponding to a fraction of the total flow, hereafter called the compliant flow, is centrally assigned by a coordinator (leader), then the remaining players (followers) choose their routes selfishly. The objective of the leader is to assign the compliant flow in a manner that minimizes a system-wide cost function, while anticipating the followers' selfish response. This setting is relevant in the planning and operation of transportation and communication networks. In transportation networks, advances in traveler information systems have made it possible to interact with individual drivers and exchange information through GPS-enabled smartphone applications or vehicular navigation systems (Work et al. 2010). These devices can be used by a traffic control center to provide routing advice that can improve the overall efficiency of the network. Naturally, the question arises on how the traffic control center should coordinate with the compliant drivers while accounting for the selfish response of other drivers, hence the importance of the Stackelberg routing framework. One might argue that the drivers who are offered routing advice are not guaranteed to follow the suggested routes, especially when these routes do not have minimal latency (in order to improve the system-wide efficiency, some drivers will be assigned routes that are suboptimal in the Nash sense). However, in some cases, it can be reasonably assumed that a fraction of the drivers will choose the routes suggested by the coordinator, despite immediate fairness concerns. For example, some drivers may have sufficient external incentives to be compliant with the coordinator. In addition, the compliant flow may also include drivers who care about the system-wide efficiency.

Stackelberg routing on parallel networks has been studied for the class of nondecreasing latency functions, and it is known that computing the optimal Stackelberg strategy is NP-hard (Roughgarden 2001). This led to the design of polynomial time approximate strategies such as *largest latency first* (Roughgarden 2001; Swamy 2007). While this class of latency functions provides a good model of congestion for a broad range of networks such as communication networks, it does not fully capture congestion phenomena in transportation. The main difference is that in transportation networks, the queuing of traffic results in an increase in density of vehicles (Daganzo 1994; Lebacque 1996; Lighthill and Whitham 1955; Richards 1956; Work et al. 2010), which in turn affects the latency. This phenomenon is sometimes referred to as horizontal queueing, since the queuing of vehicles takes physical space, as opposed to vertical queuing, such as queuing of packets in a communication link, which does not take physical space, and the notion of density is absent. Several authors have proposed different models of congestion to capture congestion phenomena specific to horizontal queuing and characterized the Nash equilibria under these models (Boulogne et al. 2001; Friesz and Mookherjee 2006; Lo and Szeto 2002; Wang et al. 2001). We introduce a new class of latency functions for congestion with horizontal queuing and study Nash and Stackelberg equilibria under this class. We restrict our study to parallel networks. Although simple, the parallel topology can be of practical importance in several situations, such as traffic planning and analysis. Even though transportation networks are rarely parallel, they can be approximated by a parallel network, for example, by only considering highways that connect two highly populated areas (Caltrans 2010). Figure 25.9 shows one such network that connects San Francisco to San Jose. We consider this network in Sect. 6.

1.2 Congestion on Horizontal Queues

The classical model for vertical queues assumes that the latency $\ell_n(x_n)$ on a link n is a nondecreasing function of the flow x_n on that link (Babaioff et al. 2009; Beckmann et al. 1956; Dafermos and Sparrow 1969; Roughgarden and Tardos 2002; Swamy 2007). However, for networks with horizontal queues (Lebacque 1996; Lighthill and Whitham 1955; Richards 1956), the latency not only depends on the flow but also on the density. For example, on a transportation network, the latency depends on the density of cars on the road (e.g., in cars per meter), and not only on the flow (e.g., in cars per second), since for a fixed value of flow, a lower density means higher velocity, hence lower latency. In order to capture this dependence on density, we introduce and discuss a simplified model of congestion that takes into account both flow and density. Let ρ_n be the density on link n, assumed to be uniform, for simplicity, and let the flow x_n be given by a continuous, concave function of the density:

$$x_n^\rho : [0, \rho_n^{\max}] \to [0, x_n^{\max}]$$
$$\rho_n \mapsto x_n = x_n^\rho(\rho_n)$$

Here, $x_n^{\max} > 0$ is the maximum flow or *capacity* of the link, and $\rho_n^{\max}$ is the maximum density that the link can hold. The function x_n^ρ is determined by the physical properties of the link. It is termed the *flux function* in conservation law theory (Evans 1998; LeVeque 2007) and the *fundamental diagram* in traffic flow theory (Daganzo 1994; Greenshields 1935; Papageorgiou et al. 1989). In general, it is a non-injective function. We make the following assumptions:

- There exists a unique density $\rho_n^{\text{crit}} \in (0, \rho_n^{\max})$ such that $x_n^\rho(\rho_n^{\text{crit}}) = x_n^{\max}$, called critical density. When $\rho_n \in [0, \rho_n^{\text{crit}}]$, the link is said to be in *free-flow*, and when $\rho_n \in (\rho_n^{\text{crit}}, \rho_n^{\max})$, it is said to be *congested*.
- In the congested regime, x_n^ρ is continuous decreasing from $(\rho_n^{\text{crit}}, \rho_n^{\max})$ onto $(0, x_n^{\max})$. In particular, $\lim_{\rho_n \to \rho_n^{\max}} x_n^\rho(\rho_n) = 0$ (the flow reduces to zero when the density approaches the maximum density).

These are standard assumptions on the flux function, following traffic flow theory (Daganzo 1994; Greenshields 1935; Papageorgiou et al. 1989). Additionally, we assume that in the free-flow regime, x_n^ρ is linearly increasing in ρ_n, and since $x_n^\rho(\rho_n^{\text{crit}}) = x_n^{\max}$, we have in the free-flow regime $x_n^\rho(\rho_n) = x_n^{\max}\rho_n/\rho_n^{\text{crit}}$. The assumption of linearity in free-flow is the only restrictive assumption, and it is essential in deriving the results on optimal Stackelberg strategies. Although somewhat restrictive, this assumption is common, and the resulting flux model is widely used in modeling transportation networks, such as in (Daganzo 1994; Papageorgiou et al. 1990). Figure 25.1 shows examples of such flux functions.

Since the density ρ_n and the flow x_n are assumed to be uniform on the link, the velocity v_n of vehicles on the link is given by $v_n = x_n/\rho_n$, and the latency is simply $L_n/v_n = L_n\rho_n/x_n$ where L_n is the length of link n. Thus to a given value

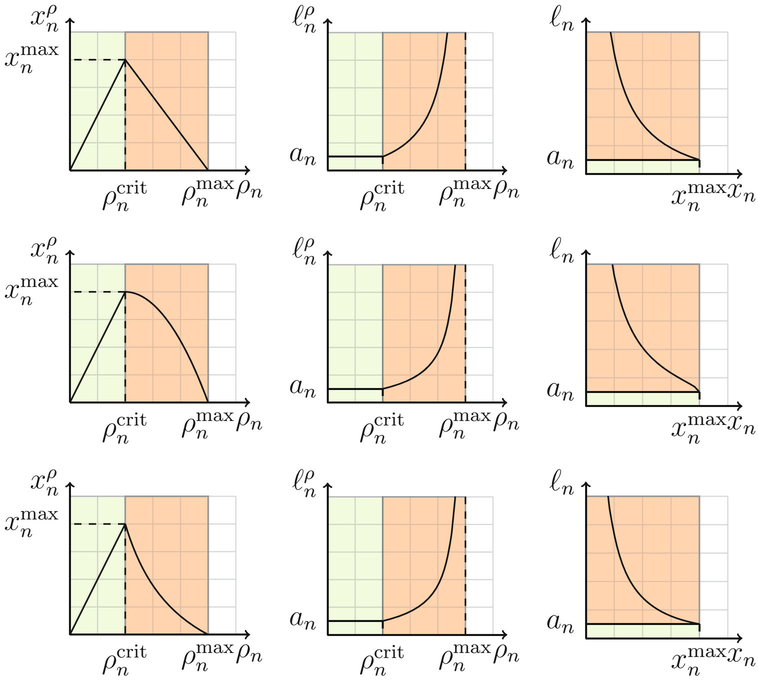


Fig. 25.1 Examples of flux functions for horizontal queues (*left*) and corresponding latency as a function of the density $\ell_n^\rho(\rho_n)$ (*middle*) and as a function of the flow and the congestion state $\ell_n(x_n, m_n)$ (*right*). The free-flow (respectively congested) regime is shaded in *green* (respectively *red*)

of the flow, there may correspond more than one value of the latency, since the flux function is non-injective in general. In other words, a given value x_n of flow of cars on a road segment can correspond to:

- Either a large concentration of cars moving slowly (high density, the road is *congested*), in which case the latency is large
- Or few cars moving fast (low density, the road is in *free-flow*), in which case the latency is small

1.3 Latency Function for Horizontal Queues

Given a flux function x_n^ρ, the latency can be easily expressed as a nondecreasing function of the density:

$$\begin{aligned} \ell_n^\rho &: [0, \rho_n^{\max}] \to \mathbb{R}_+ \\ \rho_n &\mapsto \ell_n^\rho(\rho_n) = \frac{L_n \rho_n}{x_n^\rho(\rho_n)} \end{aligned} \tag{25.1}$$

From the assumptions on the flux function, we have:

- In the free-flow regime, the flux function is linearly increasing, $x_n(\rho_n) = \frac{x_n^{\max}}{\rho_n^{\text{crit}}}\rho_n$. Thus the latency is constant in free-flow, $\ell_n^\rho(\rho_n) = \frac{L_n \rho_n^{\text{crit}}}{x_n^{\max}}$. We will denote its value by $a_n \triangleq \frac{L_n \rho_n^{\text{crit}}}{x_n^{\max}}$, called henceforth the *free-flow latency*.
- In the congested regime, x_n^ρ is bijective from $(\rho_n^{\text{crit}}, \rho_n^{\max})$ to $(0, x_n^{\max})$. Let

$$\begin{aligned} \rho_n^{\text{cong}} : (0, x_n^{\max}) &\to (\rho_n^{\text{crit}}, \rho_n^{\max}) \\ x_n &\mapsto \rho_n^{\text{cong}}(x_n) \end{aligned}$$

  be its inverse. It maps the flow x_n to the unique congestion density that corresponds to that flow. Thus in the congested regime, latency can be expressed as a function of the flow, $x_n \mapsto \ell_n^\rho(\rho_n^{\text{cong}}(x_n))$. This function is decreasing as the composition of the decreasing function ρ_n^{cong} and the increasing function ℓ_n^ρ.

We can therefore express the latency as a function of the flow in each of the separate regimes: free-flow (low density) and congested (high density). This leads to the following definition of HQSF latencies (horizontal queues, single-valued in free-flow). We introduce a binary variable $m_n \in \{0, 1\}$ which specifies whether the link is in the free-flow or the congested regime.

Definition 1 (HQSF latency class) A function

$$\begin{aligned} \ell_n : D_n &\to \mathbb{R}_+ \\ (x_n, m_n) &\mapsto \ell_n(x_n, m_n) \end{aligned} \tag{25.2}$$

defined on the domain[1]

$$D_n = [0, x_n^{\max}] \times \{0\} \cup (0, x_n^{\max}) \times \{1\}$$

is a HQSF latency function if it satisfies the following properties:

(A1) In the free-flow regime, the latency $\ell_n(\cdot, 0)$ is single valued (i.e., constant).
(A2) In the congested regime, the latency $x_n \mapsto \ell_n(x_n, 1)$ is decreasing on $(0, x_n^{\max})$.
(A3) $\lim_{x_n \to x_n^{\max}} \ell_n(x_n, 1) = a_n = \ell_n(x_n^{\max}, 0)$.

[1]The latency in congestion $\ell_n(\cdot, 1)$ is defined on the open interval $(0, x_n^{\max})$. In particular, if $x_n = 0$ or $x_n = x_n^{\max}$ then the link is always considered to be in free-flow. When the link is empty ($x_n = 0$), it is naturally in free-flow. When it is at maximum capacity ($x_n = x_n^{\max}$) it is in fact on the boundary of the free-flow and congestion regions, and we say by convention that the link is in free-flow.

Property (A1) is equivalent to the assumption that the flux function is linear in free-flow. Property (A2) results from the expression of the latency as the composition $\ell_n^\rho(\rho_n^{\text{cong}}(x_n))$, where ℓ_n^ρ is increasing and ρ_n^{cong} is decreasing. Property (A3) is equivalent to the continuity of the underlying flux function x_n^ρ.

Although it may be more natural to think of the latency as a nondecreasing function of the density, the above representation in terms of flow x_n and congestion state m_n will be useful in deriving properties of the Nash equilibria of the routing game. Finally, we observe, as an immediate consequence of these properties, that the latency in congestion is always greater than the free-flow latency: $\forall x_n \in (0, x_n^{\max})$, $\ell_n(x_n, 1) > a_n$. Some examples of HQSF latency functions (and the underlying flux functions) are illustrated in Fig. 25.1. We now give a more detailed derivation of a latency function from a macroscopic fundamental diagram of traffic.

1.4 A HQSF Latency Function from a Triangular Fundamental Diagram of Traffic

In this section we derive one example of an HQSF latency function ℓ_n from the fundamental diagram of traffic, corresponding to the top row in Fig. 25.1. We consider a triangular fundamental diagram, used to model traffic flow, for example, in (Daganzo 1994, 1995), i.e., a piecewise affine flux function x_n^ρ, given by

$$x_n^\rho(\rho_n) = \begin{cases} v_n^f \rho_n & \text{if } \rho_n \in [0, \rho_n^{\text{crit}}] \\ x_n^{\max} \frac{\rho_n - \rho_n^{\max}}{\rho_n^{\text{crit}} - \rho_n^{\max}} & \text{if } \rho_n \in (\rho_n^{\text{crit}}, \rho_n^{\max}] \end{cases}$$

The flux function is linear in free-flow with positive slope v_n^f called free-flow speed, affine in congestion with negative slope $v_n^c \triangleq x_n^{\max}/(\rho_n^{\text{crit}} - \rho_n^{\max})$, and continuous (thus $v_n^f \rho_n^{\text{crit}} = x_n^{\max}$). By definition, it satisfies the assumptions in Sect. 1.2. The latency is given by $L_n \rho_n / x_n^\rho(\rho_n)$ where L_n is the length of link n. It is then a simple function of the density

$$\ell_n^\rho(\rho_n) = \begin{cases} \frac{L_n}{v_n^f} & \rho_n \in [0, \rho_n^{\text{crit}}] \\ \frac{L_n \rho_n}{v_n^c(\rho_n - \rho_n^{\max})} & \rho_n \in (\rho_n^{\text{crit}}, \rho_n^{\max}] \end{cases}$$

which can be expressed as two functions of flow: a constant function $\ell_n(\cdot, 0)$ when the link is in free-flow and a decreasing function $\ell_n(\cdot, 1)$ when the link is congested

$$\ell_n(x_n, 0) = \frac{L_n}{v_n^f}$$

$$\ell_n(x_n, 1) = L_n \left(\frac{\rho_n^{\max}}{x_n} + \frac{1}{v_n^c} \right)$$

This defines a function ℓ_n that satisfies the assumptions of Definition 1 and thus belongs to the HQSF latency class. Figure 25.1 shows one example of a triangular fundamental diagram (top left) and the corresponding latency function ℓ_n (top right).

2 Game Model and Main Results

2.1 The Routing Game

We consider a non-atomic routing game on a parallel network, shown in Fig. 25.2. Here non-atomic means that the game involves a continuum of players, where each player corresponds to an infinitesimal (non-atomic) amount of flow (Roughgarden and Tardos 2004; Schmeidler 1973). The network has a single source and a single sink. Connecting the source and sink are N parallel links indexed by $n \in \{1, \dots, N\}$. We assume, without loss of generality, that the links are ordered by increasing free-flow latencies. To simplify the discussion, we further assume that free-flow latencies are distinct. Therefore we have $a_1 < a_2 < \cdots < a_N$. The network is subject to a constant positive flow demand r at the source. We will denote by (N, r) an instance of the routing game played on a network with N parallel links subject to demand r. The state of the network is given by a feasible flow assignment vector $\boldsymbol{x} \in \mathbb{R}_+^N$ such that $\sum_{n=1}^{N} x_n = r$ where x_n is the flow on link n and a congestion state vector $\boldsymbol{m} \in \{0, 1\}^N$ where $m_n = 0$ if the link is in free-flow and $m_n = 1$ if the link is congested, as defined above. All physical quantities (density and flow) are assumed to be static and uniform on the link.

Every non-atomic player chooses a route in order to minimize his/her individual latency (Roughgarden and Tardos 2002). If a player chooses link n, his/her latency is given by $\ell_n(x_n, m_n)$, where ℓ_n is a HQSF latency function. We assume that players know the latency functions.

Pure Nash equilibria of the game (which we will simply refer to as Nash equilibria) are assignments $(\boldsymbol{x}, \boldsymbol{m})$ such that every player cannot improve his/her latency by switching to a different link.

Definition 2 (Nash Equilibrium) A feasible assignment $(\boldsymbol{x}, \boldsymbol{m}) \in \mathbb{R}_+^N \times \{0, 1\}^N$ is a Nash equilibrium of the routing game instance (N, r) if $\forall n \in \operatorname{supp}(\boldsymbol{x})$, $\forall k \in \{1, \dots, N\}$, $\ell_n(x_n, m_n) \leq \ell_k(x_k, m_k)$.

Here $\operatorname{supp}(\boldsymbol{x}) = \left\{ n \in \{1, \dots, N\} | x_n > 0 \right\}$ denotes the support of $\boldsymbol{x}$. As a consequence of this definition, all links in the support of $\boldsymbol{x}$ have the same latency

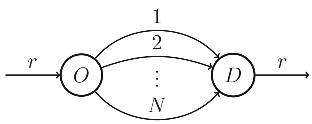


Fig. 25.2 Network with N parallel links under demand r

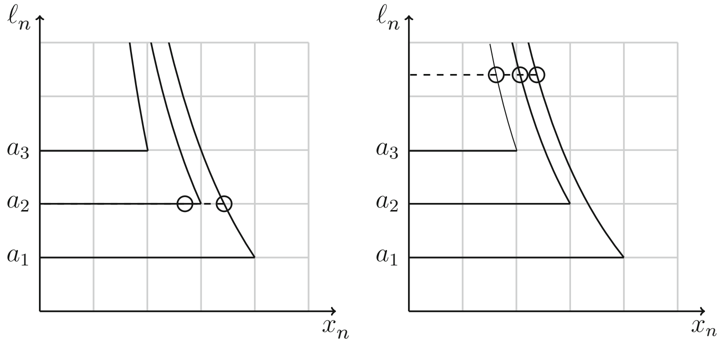


Fig. 25.3 Example of Nash equilibria for a three-link network. One equilibrium (*left*) has one link in free-flow and one congested link. A second equilibrium (*right*) has three congested links

ℓ_0, and links that are not in the support have latency greater than or equal to ℓ_0. We will denote by $\mathrm{NE}(N, r)$ the set of Nash equilibria of the instance (N, r). We note that a Nash equilibrium for the routing game is a *static* equilibrium; we do not model dynamics of density or flow. Figure 25.3 shows an example of a routing game instance and resulting Nash equilibria.

While a Nash equilibrium achieves minimal individual latencies, it does not minimize, in general, the *system cost* or *total cost* defined as follows:

Definition 3 The total cost of an assignment $(\boldsymbol{x}, \boldsymbol{m})$ is the total latency experienced by all players:

$$C(\boldsymbol{x}, \boldsymbol{m}) = \sum_{n=1}^{N} x_n \ell_n(x_n, m_n). \tag{25.3}$$

As detailed in Sect. 3, under the HQSF latency class, the routing game may have multiple Nash equilibria that have different total costs. We are interested, in particular, in Nash equilibria that have minimal cost, which are referred to as *best Nash equilibria* (BNE).

Definition 4 (Best Nash Equilibria) The set of best Nash equilibria is the set of equilibria that minimize the total cost, i.e.,

$$\mathrm{BNE}(N, r) = \underset{(\boldsymbol{x}, \boldsymbol{m}) \in \mathrm{NE}(N, r)}{\arg\min} \; C(\boldsymbol{x}, \boldsymbol{m}). \tag{25.4}$$

2.2 Stackelberg Routing Game

In the Stackelberg routing game, a coordinator (a central authority) is assumed to have control over a positive fraction α of the total flow demand r. We call α the *compliance rate*. The coordinator wants to route the *compliant flow* αr in a way that minimizes the system cost, while anticipating the response of the rest of the players, assumed to choose their routes selfishly after the strategy of the coordinator is revealed. We will refer to the flow of selfish players $(1-\alpha)r$ as the *non-compliant flow*. More precisely, the game is played as follows:

- First, the coordinator (the leader) chooses a *Stackelberg strategy*, i.e., an assignment $\boldsymbol{s} \in \mathbb{R}_+^N$ of the compliant flow (such that $\sum_{n=1}^N s_n = \alpha r$).
- Then, the Stackelberg strategy $\boldsymbol{s}$ of the leader is revealed, and the non-compliant players (followers) choose their routes selfishly and form a Nash equilibrium $(\boldsymbol{t}(\boldsymbol{s}), \boldsymbol{m}(\boldsymbol{s}))$, *induced*[2] by strategy $\boldsymbol{s}$. By definition, the induced equilibrium $(\boldsymbol{t}(\boldsymbol{s}), \boldsymbol{m}(\boldsymbol{s}))$ satisfies

$$\begin{aligned} &\forall n \in \operatorname{supp}(\boldsymbol{t}(\boldsymbol{s})),\ \forall k \in \{1, \ldots, N\}, \\ &\qquad \ell_n(s_n + t_n(\boldsymbol{s}), m_n(\boldsymbol{s})) \leq \ell_k(s_k + t_k(\boldsymbol{s}), m_k(\boldsymbol{s})) \end{aligned} \tag{25.5}$$

The total flow on the network is $\boldsymbol{s} + \boldsymbol{t}(\boldsymbol{s})$; thus the total cost is $C(\boldsymbol{s} + \boldsymbol{t}(\boldsymbol{s}), \boldsymbol{m}(\boldsymbol{s}))$. Note that a Stackelberg strategy $\boldsymbol{s}$ may induce multiple Nash equilibria in general. However, we define $(\boldsymbol{t}(\boldsymbol{s}), \boldsymbol{m}(\boldsymbol{s}))$ to be the best such equilibrium (the one with minimal total cost, which will be shown to be unique in Sect. 4).

We will use the following notation:

- (N, r, α) is an instance of the Stackelberg routing game played on a parallel network with N links under flow demand r with compliance rate α. Note that the routing game (N, r) is a special case of the Stackelberg routing game with $\alpha = 0$.
- $\mathrm{S}(N, r, \alpha) \subset \mathbb{R}_+^N$ is the set of Stackelberg strategies for the Stackelberg instance (N, r, α).
- $\mathrm{S}^\star(N, r, \alpha)$ is the *set of optimal Stackelberg strategies* defined as

$$\mathrm{S}^\star(N, r, \alpha) = \underset{\boldsymbol{s} \in \mathrm{S}(N,r,\alpha)}{\arg\min}\, C(\boldsymbol{s} + \boldsymbol{t}(\boldsymbol{s}), \boldsymbol{m}(\boldsymbol{s})). \tag{25.6}$$

2.3 Optimal Stackelberg Strategy

We now define a candidate Stackelberg strategy, which we call the *non-compliant first* strategy (NCF) and which we will prove to be optimal. The NCF strategy corresponds to first computing the best Nash equilibrium $(\bar{\boldsymbol{t}}, \bar{\boldsymbol{m}})$ of the non-

[2] We note that a feasible flow assignment $\boldsymbol{s}$ of compliant flow may fail to induce a Nash equilibrium $(\boldsymbol{t}, \boldsymbol{m})$ and therefore is not considered to be a valid Stackelberg strategy.

compliant flow for the routing game instance $(N, (1-\alpha)r)$, then finding a particular strategy $\boldsymbol{s}$ that induces $(\bar{\boldsymbol{t}}, \bar{\boldsymbol{m}})$.

Definition 5 (The non-compliant first (NCF) strategy) Consider the Stackelberg instance (N, r, α). Let $(\bar{\boldsymbol{t}}, \bar{\boldsymbol{m}})$ be the best Nash equilibrium of the non-compliant flow, $\{(\bar{\boldsymbol{t}}, \bar{\boldsymbol{m}})\} = \text{BNE}(N, (1-\alpha)r)$, and $\bar{k} = \max \text{supp}(\bar{\boldsymbol{t}})$ be the last link in its support. Then the non-compliant first strategy, denoted by $\text{NCF}(N, r, \alpha)$, is defined as follows:

$$\text{NCF}(N, r, \alpha) = \Big(0, \ldots, \overset{\bar{k}-1}{\sqcap\!\!\!\!0}, \overbrace{x_{\bar{k}}^{\max} - \bar{t}_{\bar{k}}}^{\bar{k}}, x_{\bar{k}+1}^{\max}, \ldots, x_{l-1}^{\max}, \alpha r - \Big(\sum_{n=\bar{k}}^{l-1} x_n^{\max} - \bar{t}_{\bar{k}}\Big), 0, \ldots, 0\Big) \tag{25.7}$$

where l is the maximal index in $\{\bar{k}+1, \ldots, N\}$ such that $\alpha r - \big(\sum_{n=\bar{k}}^{l-1} x_n^{\max} - \bar{t}_{\bar{k}}\big) \geq 0$.

In words, the NCF strategy saturates links one by one, by increasing index starting from link $\bar{k}$, the last link used by the non-compliant flow in the best Nash equilibrium of $(N, (1-\alpha)r)$. Thus it will assign $x_{\bar{k}}^{\max} - \bar{t}_{\bar{k}}$ to link $\bar{k}$, then $x_{\bar{k}+1}^{\max}$ to link $\bar{k}+1$, $x_{\bar{k}+2}^{\max}$ to link $\bar{k}+2$, and so on, until the compliant flow is assigned entirely (see Fig. 25.4). The following theorem states the main result.

Theorem 1 *Under the class of HQSF latency functions,* $\text{NCF}(N, r, \alpha)$ *is an optimal Stackelberg strategy for the Stackelberg instance* (N, r, α).

We give a proof of Theorem 1 in Sect. 4. We will also show that for the class of HQSF latency functions, the best Nash equilibria can be computed in polynomial time in the size N of the network, and as a consequence, the NCF strategy can also be computed in polynomial time. This stands in contrast to previous results under the class of nondecreasing latency functions, for which computing the optimal Stackelberg strategy is NP-hard (Roughgarden 2001).

3 Nash Equilibria

In this section, we study Nash equilibria of the routing game. We show that under the class of HQSF latency functions, there may exist multiple Nash equilibria that have different costs. Then we partition the set of equilibria into congested equilibria and single-link-free-flow equilibria. Finally, we characterize the best Nash equilibrium and show that it can be computed in quadratic time in the number of links.

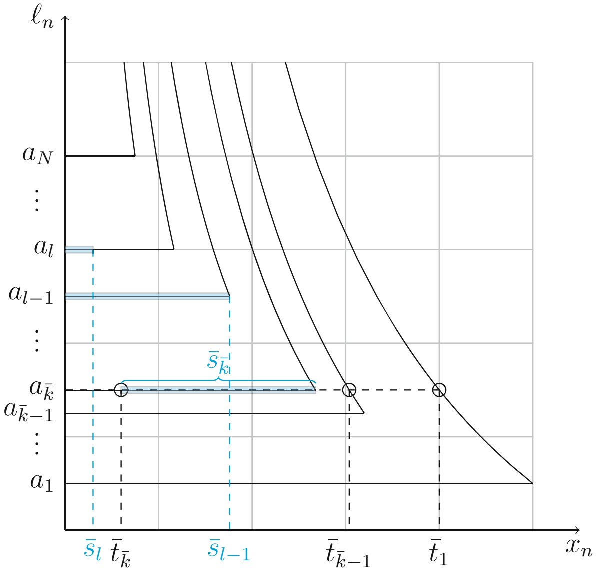


Fig. 25.4 Non-compliant first (NCF) strategy $\bar{s}$ and its induced equilibrium. *Circles* show the best Nash equilibrium $(\bar{t}, \bar{m})$ of the non-compliant flow $(1-\alpha)r$: link $\bar{k}$ is in free-flow, and links $\{1, \ldots, \bar{k}-1\}$ are congested. The Stackelberg strategy $\bar{s} = \text{NCF}(N, r, \alpha)$ is highlighted in *blue*

3.1 Structure and Properties of Nash Equilibria

We first give some properties of Nash equilibria.

Proposition 1 (Total cost of a Nash Equilibrium) *Let $(x, m) \in \text{NE}(N, r)$ be a Nash equilibrium for the instance (N, r). Then there exists $\ell_0 > 0$ such that $\forall n \in supp\,(x)$, $\ell_n(x_n, m_n) = \ell_0$, and $\forall n \notin supp\,(x)$, $\ell_n(0, 0) \geq \ell_0$. The total cost of the equilibrium is then $C(x, m) = r\ell_0$.*

Proposition 2 *Let $(x, m) \in \text{NE}(N, r)$ be a Nash equilibrium. Then $k \in supp\,(x) \Rightarrow \forall n < k$, link n is congested.*

Proof By contradiction, if $m_n = 0$, then $\ell_n(x_n, m_n) = a_n < a_k \leq \ell_k(x_k, m_k)$, which contradicts Definition 2 of a Nash equilibrium.

Corollary 1 (Support of a Nash equilibrium) *Let $(x, m) \in \text{NE}(N, r)$ be a Nash equilibrium and $k = \max supp\,(x)$ be the last link in the support of x (i.e., the one with the largest free-flow latency). Then we have $supp\,(x) = \{1, \ldots, k\}$.*

Proof Since $k \in \text{supp}(\boldsymbol{x})$, we have by Proposition 2 that $\forall n < k$, link n is congested, thus $n \in \text{supp}(\boldsymbol{x})$ (by definition, a congested link cannot be empty).

3.1.1 No Essential Uniqueness

For the HQSF latency class, the essential uniqueness property[3] does not hold, i.e., there may exist multiple Nash equilibria that have different costs; an example is given in Fig. 25.3.

3.1.2 Single-Link-Free-Flow Equilibria and Congested Equilibria

The example shows that in general, there may exist multiple Nash equilibria that have different costs, different congestion state vectors, and different supports. However, not every congestion state vector $\boldsymbol{m} \in \{0,1\}^N$ can be that of a Nash equilibrium: let $(\boldsymbol{x}, \boldsymbol{m}) \in \text{NE}(N, r)$ be a Nash equilibrium, and let $k = \max \text{supp}(\boldsymbol{x})$ be the index of the last link in the support of $\boldsymbol{x}$. Then by Proposition 2, we have that $\forall i < k, m_i = 1$, and $\forall i > k, m_i = 0$. Thus we have

- Either $\boldsymbol{m} = \left(1, \ldots, 1, \overset{k}{\sqcap}{0}, 0, \ldots, 0\right)$, i.e., the last link in the support is in free-flow, all other links in the support are congested. In this case we call $(\boldsymbol{x}, \boldsymbol{m})$ a *single-link-free-flow equilibrium* and denote the set of such equilibria by $\text{NE}_\text{f}(N, r)$.
- Or $\boldsymbol{m} = \left(1, \ldots, 1, \overset{k}{\sqcap}{1}, 0, \ldots, 0\right)$, i.e., all links in the support are congested. In this case we call $(\boldsymbol{x}, \boldsymbol{m})$ a *congested equilibrium* and denote the set of such equilibria by $\text{NE}_\text{c}(N, r)$.

3.2 Existence of Single-Link-Free-Flow Equilibria

Let $(\boldsymbol{x}, \boldsymbol{m})$ be a single-link-free-flow equilibrium, and let $k = \max \text{supp}(\boldsymbol{x})$. We have from Proposition 2 that links $\{1, \ldots, k-1\}$ are congested and link k is in free-flow. Therefore we must have $\forall n \in \{1, \ldots, k-1\}$, $\ell_n(x_n, 1) = \ell_k(x_k, 0) = a_k$. This uniquely determines the flow on the congested links:

Definition 6 (Congestion flow) Let $k \in \{2, \ldots, N\}$. Then $\forall n \in \{1, \ldots, k-1\}$, there exists a unique flow x_n such that $\ell_n(x_n, m_n) = a_k$. We denote this flow by $\hat{x}_n(k)$ and call it *k-congestion flow* on link n. It is given by

$$\hat{x}_n(k) = \ell_n(\cdot, 1)^{-1}(a_k). \tag{25.8}$$

We note that $\hat{x}_n(k)$ is decreasing in k, since $\ell_n(\cdot, 1)^{-1}$ is decreasing.

[3] The essential uniqueness property states that for the class of non-decreasing latency functions, all Nash equilibria have the same total cost. See for example (Beckmann et al. 1956; Dafermos and Sparrow 1969; Roughgarden and Tardos 2002).

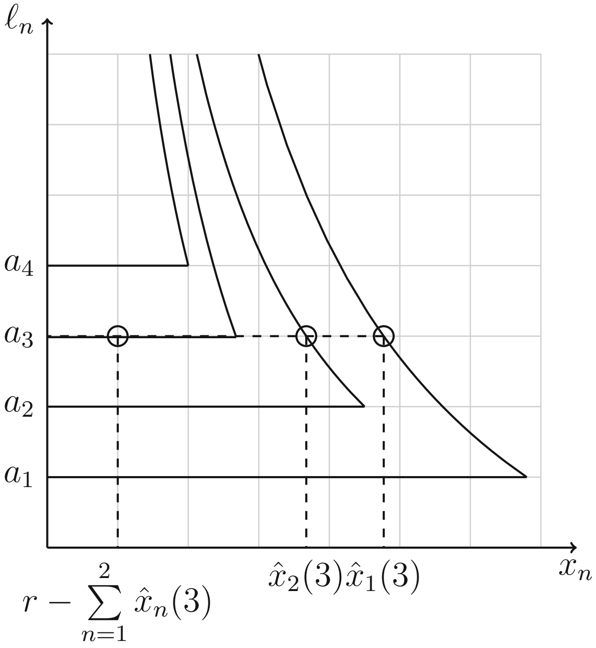


Fig. 25.5 Example of a single-link-free-flow equilibrium. Link 3 is in free-flow, and links 1 and 2 are congested. The common latency on all links in the support is a_3

Proposition 3 (Single-link-free-flow equilibria) $(\boldsymbol{x}, \boldsymbol{m})$ *is a single-link-free-flow equilibrium if and only if* $\exists k \in \{1, \ldots, N\}$ *such that* $0 < r - \sum_{n=1}^{k-1} \hat{x}_n(k) \leq x_k^{\max}$, *and*

$$\boldsymbol{x} \triangleq \left(\hat{x}_1(k), \ldots, \hat{x}_{k-1}(k), r - \sum_{n=1}^{k-1} \hat{x}_n(k), 0, \ldots, 0 \right) \tag{25.9}$$

$$\boldsymbol{m} \triangleq \left(1, \ldots, 1, \overset{\overset{k}{\sqcap}}{0}, \ldots, 0 \right) \tag{25.10}$$

Illustrations of Eqs. (25.10) and (25.9) are shown in Fig. 25.5.

Next, we give a necessary and sufficient condition for the existence of single-link-free-flow equilibria.

Lemma 1 Existence of single-link-free-flow equilibria
Let

$$r^{\mathrm{NE}}(N) \triangleq \max_{k \in \{1,\dots,N\}} \left\{ x_k^{\max} + \sum_{n=1}^{k-1} \hat{x}_n(k) \right\} \tag{25.11}$$

A single-link-free-flow equilibrium exists for the instance (N, r) *if and only if* $r \leq r^{\mathrm{NE}}(N)$.

Proof If a single-link-free-flow equilibrium exists, then by Proposition 3, it is of the form given by Eqs. (25.10) and (25.9) for some k. The flow on link k is then given by $r - \sum_{n=1}^{k-1} \hat{x}_n(k) \leq x_k^{\max}$. Therefore $r \leq x_k^{\max} + \sum_{n=1}^{k-1} \hat{x}_n(k) \leq r^{\mathrm{NE}}(N)$.

We prove the converse by induction on the size N of the network. Let $\mathtt{P}_N$ denote the property: $\forall r \in (0, r^{\mathrm{NE}}(N)]$, there exists a single-link-free-flow equilibrium for the instance (N, r).

For $N = 1$, it is clear that if $0 < r \leq x_1^{\max}$, there is a single-link-free-flow equilibrium simply given by $(x_1, m_1) = (r, 0)$.

Now let $N \geq 1$, assume $\mathtt{P}_N$ holds and let us show $\mathtt{P}_{N+1}$. Let $0 < r \leq r^{\mathrm{NE}}(N+1)$, and consider an instance $(N+1, r)$.

Case 1 *If* $r \leq r^{\mathrm{NE}}(N)$*, then by the induction hypothesis* $\mathtt{P}_N$*, there exists a single-link-free-flow equilibrium* $(\boldsymbol{x}, \boldsymbol{m})$ *for the instance* (N, r)*. Then* $(\boldsymbol{x}', \boldsymbol{m}')$ *defined as* $\boldsymbol{x}' = (x_1, \dots, x_N, 0)$ *and* $\boldsymbol{m}' = (m_1, \dots, m_N, 0)$ *is clearly a single-link-free-flow equilibrium for the instance* $(N+1, r)$.

Case 2 *If* $r^{\mathrm{NE}}(N) < r \leq r^{\mathrm{NE}}(N+1)$*, then by Proposition 3, an equilibrium exists if*

$$0 < r - \sum_{n=1}^{N} \hat{x}_n(N+1) \leq x_{N+1}^{\max}. \tag{25.12}$$

First, we note that since $r^{\mathrm{NE}}(N) < r^{\mathrm{NE}}(N+1)$, then

$$r^{\mathrm{NE}}(N+1) = x_{N+1}^{\max} + \sum_{n=1}^{N} \hat{x}_n(N+1),$$

thus

$$r \leq r^{\mathrm{NE}}(N+1) = x_{N+1}^{\max} + \sum_{n=1}^{N} \hat{x}_n(N+1),$$

which proves the second inequality in (25.12). To show the first inequality, we have

$$r > r^{\mathrm{NE}}(N) \geq x_N^{\max} + \sum_{n=1}^{N-1} \hat{x}_n(N)$$
$$\geq \hat{x}_N(N+1) + \sum_{n=1}^{N-1} \hat{x}_n(N+1),$$

where the last inequality results from the fact that $\hat{x}_n(N) \geq \hat{x}_n(N+1)$ and $x_N^{\max} \geq \hat{x}_N(N+1)$ by Definition 6 of congestion flow. This completes the induction.

Corollary 2 *The maximum demand r such that the set of Nash equilibria* $\mathrm{NE}(N, r)$ *is nonempty is $r^{\mathrm{NE}}(N)$.*

Proof By the previous Lemma, $r^{\mathrm{NE}}(N)$ is a lower bound on the maximum demand. To show that it is also an upper bound, suppose that $\mathrm{NE}(N, r)$ is nonempty and let $(\boldsymbol{x}, \boldsymbol{m}) \in \mathrm{NE}(N, r)$ and $k = \max \operatorname{supp}(\boldsymbol{x})$. Then we have $\operatorname{supp}(\boldsymbol{x}) = \{1, \ldots, k\}$ by Corollary 1, and by Definition 2 of a Nash equilibrium, $\forall n \leq k$, $\ell_n(x_n, m_n) = \ell_k(x_k, m_k) \geq a_k$, and therefore $x_n \leq \hat{x}_n(k)$. We also have $x_k \leq x_k^{\max}$. Combining the inequalities, we have

$$r = \sum_{n=1}^{k} x_n \leq x_k^{\max} + \sum_{n=1}^{k-1} \hat{x}_n(k) \leq r^{\mathrm{NE}}(N).$$

3.3 Number of Equilibria

Proposition 4 (An upper bound on the number of equilibria) *Consider a routing game instance (N, r). For any given $k \in \{1, \ldots, N\}$, there is at most one single-link-free-flow equilibrium and one congested equilibrium with support $\{1, \ldots, k\}$. As a consequence, by Corollary 1, the instance (N, r) has at most N single-link-free-flow equilibria and N congested equilibria.*

Proof We prove the result for single-link-free-flow equilibria, the proof for congested equilibria is similar. Let $k \in \{1, \ldots, N\}$, and assume $(\boldsymbol{x}, \boldsymbol{m})$ and $(\boldsymbol{x}', \boldsymbol{m}')$ are single-link-free-flow equilibria such that $\max \operatorname{supp}(\boldsymbol{x}) = \max \operatorname{supp}(\boldsymbol{x}') = k$. We first observe that by Corollary 1, $\boldsymbol{x}$ and $\boldsymbol{x}'$ have the same support $\{1, \ldots, k\}$, and by Proposition 2, $\boldsymbol{m} = \boldsymbol{m}'$. Since link k is in free-flow under both equilibria, we have $\ell_k(x_k, m_k) = \ell_k(x'_k, m'_k) = a_k$, and by Definition 2 of a Nash equilibrium, any link in the support of both equilibria has the same latency a_k, i.e., $\forall n < k$, $\ell_n(x_n, 1) = \ell_n(x'_n, 1) = a_k$. Since the latency in congestion is injective, we have $\forall n < k$, $x_n = x'_n$, therefore $\boldsymbol{x} = \boldsymbol{x}'$.

3.4 Best Nash Equilibrium

In order to study the inefficiency of Nash equilibria, and the improvement of performance that we can achieve using optimal Stackelberg routing, we focus our attention on best Nash equilibria and *price of stability* (Anshelevich et al. 2004) as a measure of their inefficiency.

Lemma 2 (Best Nash Equilibrium) *For a routing game instance* (N, r), $r \leq r^{\mathrm{NE}}(N)$, *the unique best Nash equilibrium is the single-link-free-flow equilibrium that has the smallest support*

$$\mathrm{BNE}(N, r) = \underset{(\boldsymbol{x},\boldsymbol{m}) \in \mathrm{NE_f}(N,r)}{\arg\min} \{\max \mathit{supp}\,(\boldsymbol{x})\}.$$

Proof We first show that a congested equilibrium cannot be a best Nash equilibrium. Let $(\boldsymbol{x}, \boldsymbol{m}) \in \mathrm{NE}(N, r)$ be a congested equilibrium, and let $k = \max \mathrm{supp}\,(x)$. By Proposition 1, the cost of $(\boldsymbol{x}, \boldsymbol{m})$ is $C(\boldsymbol{x}, \boldsymbol{m}) = \ell_k(x_k, 1)r > a_k r$. We observe that $(\boldsymbol{x}, \boldsymbol{m})$ restricted to $\{1, \ldots, k\}$ is an equilibrium for the instance (k, r); thus by Corollary 2, $r \leq r^{\mathrm{NE}}(k)$, and by Lemma 1, there exists a single-link-free-flow equilibrium $(\boldsymbol{x}', \boldsymbol{m}')$ for (k, r), with cost $C(\boldsymbol{x}', \boldsymbol{m}') \leq a_k r$. Clearly, $(\boldsymbol{x}'', \boldsymbol{m}'')$, defined as $\boldsymbol{x}'' = (x'_1, \ldots, x'_k, 0, \ldots, 0)$ and $\boldsymbol{m}'' = (m'_1, \ldots, m'_k, 0, \ldots, 0)$, is a single-link-free-flow equilibrium for the original instance (N, r), with cost $C(\boldsymbol{x}'', \boldsymbol{m}'') = C(\boldsymbol{x}', \boldsymbol{m}') \leq a_k r < C(\boldsymbol{x}, \boldsymbol{m})$, which proves that $(\boldsymbol{x}, \boldsymbol{m})$ is not a best Nash equilibrium. Therefore best Nash equilibria are single-link-free-flow equilibria. And since the cost of a single-link-free-flow equilibrium $(\boldsymbol{x}, \boldsymbol{m})$ is simply $C(\boldsymbol{x}, \boldsymbol{m}) = a_k r$ where $k = \max \mathrm{supp}\,(\boldsymbol{x})$, it is clear that the smaller the support, the lower the total cost. Uniqueness follows from Proposition 4.

3.4.1 Complexity of Computing the Best Nash Equilibrium

Lemma 2 gives a simple algorithm for computing the best Nash equilibrium for any instance (N, r): simply enumerate all single-link-free-flow equilibria (there are at most N such equilibria by Proposition 4), and select the one with the smallest support. This is detailed in Algorithm 3.

The congestion flow values $\{\hat{x}_n(k), 1 \leq n < k \leq N\}$ can be precomputed in $O(N^2)$. There are at most N calls to `freeFlowConfig`, which runs in $O(N)$ time; thus `bestNE` runs in $O(N^2)$ time. This shows that the best Nash equilibrium can be computed in quadratic time.

4 Optimal Stackelberg Strategies

In this section, we prove our main result that the NCF strategy is an optimal Stackelberg strategy (Theorem 1). Furthermore, we show that the entire set of optimal strategies $\mathrm{S}^{\star}(N, r, \alpha)$ can be computed in a simple way from the NCF strategy.

Algorithm 3 Best Nash equilibrium

```
procedure bestNE(N, r)
Inputs: Size of the network N, demand r
Outputs: Best Nash equilibrium (x, m)
for k ∈ {1, ..., N}
    let (x, m) = freeFlowConfig(k)
    if x_k ∈ [0, x_k^max]
        return (x, m)
return No-Solution

procedure freeFlowConfig(k)
Inputs: Free-flow link index k
Outputs: Assignment (x, m) = (x^{r,k}, m^k)
for n ∈ {1, ..., N}
    if n < k
        x_n = x̂_n(k), m_n = 1
    elseif n == k
        x_k = r − Σ_{n=1}^{k−1} x_n, m_k = 0
    else
        x_n = 0, m_n = 0
return (x, m)
```

Let $(\bar{\boldsymbol{t}}, \bar{\boldsymbol{m}})$ be the *best Nash equilibrium* for the instance $(N, (1-\alpha)r)$. It represents the best Nash equilibrium of the non-compliant flow $(1-\alpha)r$ when it is not sharing the network with the compliant flow. Let $\bar{k} = \max \operatorname{supp}\left(\bar{\boldsymbol{t}}\right)$ be the last link in the support of $\bar{\boldsymbol{t}}$. Let $\bar{\boldsymbol{s}}$ be the NCF strategy defined by Eq. (25.7). Then the total flow $\bar{\boldsymbol{x}} = \bar{\boldsymbol{s}} + \bar{\boldsymbol{t}}$ is given by

$$\bar{\boldsymbol{x}} = \left(\hat{x}_1(\bar{k}), \ldots, \hat{x}_{\bar{k}-1}(\bar{k}), x_{\bar{k}}^{\max}, x_{\bar{k}+1}^{\max}, \ldots, x_{l-1}^{\max}, r - \sum_{n=1}^{\bar{k}-1} \hat{x}_n(\bar{k}) - \sum_{n=\bar{k}}^{l-1} x_n^{\max}, 0, \ldots, 0\right), \tag{25.13}$$

and the corresponding latencies are

$$\left(a_{\bar{k}}, \ldots, \overset{\bar{k}}{\sqcap}{a_{\bar{k}}}, a_{\bar{k}+1}, \ldots, a_N\right). \tag{25.14}$$

Figure 25.4 shows the total flow $\bar{\boldsymbol{x}}_n = \bar{\boldsymbol{s}}_n + \bar{\boldsymbol{t}}_n$ on each link. Under $(\bar{\boldsymbol{x}}, \bar{\boldsymbol{m}})$, links $\left\{1, \ldots, \bar{k}-1\right\}$ are congested and have latency $a_{\bar{k}}$, links $\left\{\bar{k}, \ldots, l-1\right\}$ are in free-flow and at maximum capacity, and the remaining flow is assigned to link l.

We observe that for any Stackelberg strategy $\boldsymbol{s} \in \mathrm{S}(N, r, \alpha)$, the induced best Nash equilibrium $(\boldsymbol{t}(\boldsymbol{s}), \boldsymbol{m}(\boldsymbol{s}))$ is a single-link-free-flow equilibrium by Lemma 2, since $(\boldsymbol{t}(\boldsymbol{s}), \boldsymbol{m}(\boldsymbol{s}))$ is the best Nash equilibrium for the instance $(N, \alpha r)$ and latencies

$$\begin{aligned} \tilde{\ell}_n : \tilde{D}_n &\to \mathbb{R}_+ \\ (x_n, m_n) &\mapsto \ell_n(s_n + x_n, m_n) \end{aligned} \tag{25.15}$$

where $\tilde{D}_n \triangleq [0, \tilde{x}_n^{\max}] \times \{0\} \cup (0, \tilde{x}_n^{\max}) \times \{1\}$ and $\tilde{x}_n^{\max} \triangleq x_n^{\max} - s_n$.

4.1 Proof of Theorem 1: The NCF Strategy Is an Optimal Stackelberg Strategy

Let $\boldsymbol{s} \in S(N, r, \alpha)$ be any Stackelberg strategy and $(\boldsymbol{t}, \boldsymbol{m}) = (\boldsymbol{t}(\boldsymbol{s}), \boldsymbol{m}(\boldsymbol{s}))$ be the best Nash equilibrium of the non-compliant flow, induced by $\boldsymbol{s}$. To prove that the NCF startegy $\bar{s}$ is optimal, we will compare the costs induced by s and $\bar{s}$. Let $\boldsymbol{x} = \boldsymbol{s} + \boldsymbol{t}(\boldsymbol{s})$ and $\bar{\boldsymbol{x}} = \bar{\boldsymbol{s}} + \bar{\boldsymbol{t}}$ be the total flows induced by each strategy. To prove Theorem 1, we seek to show that $C(\boldsymbol{x}, \boldsymbol{m}) \geq C(\bar{\boldsymbol{x}}, \bar{\boldsymbol{m}})$.

The proof is organized as follows: we first compare the supports of the induced equilibria (Lemma 3) and then show that links $\{1, \ldots, l-1\}$ are more congested under $(\boldsymbol{x}, \boldsymbol{m})$ than under $(\bar{\boldsymbol{x}}, \bar{\boldsymbol{m}})$, in the following sense - they hold less flow and have greater latency (Lemma 4). Then we conclude by showing the desired inequality.

Lemma 3 *Let* $k = \max supp\,(\boldsymbol{t})$ *and* $\bar{k} = \max supp\,(\bar{\boldsymbol{t}})$. *Then* $k \geq \bar{k}$.

In other words, the last link in the support of $\boldsymbol{t}(\boldsymbol{s})$ has higher free-flow latency than the last link in the support of $\bar{\boldsymbol{t}}$.

Proof We first note that $(\boldsymbol{s} + \boldsymbol{t}(\boldsymbol{s}), \boldsymbol{m})$ restricted to supp $(\boldsymbol{t}(\boldsymbol{s}))$ is a Nash equilibrium. Then since link k is in free-flow, we have $\ell_k(s_k + t_k(\boldsymbol{s}), m_k) = a_k$, and since $k \in \text{supp}\,(\boldsymbol{t}(\boldsymbol{s}))$, we have by definition that any other link has greater or equal latency. In particular, $\forall n \in \{1, \ldots k-1\}$, $\ell_n(s_n + t_n(\boldsymbol{s}), m_n) \geq a_k$, thus $s_n + t_n(\boldsymbol{s}) \leq \hat{x}_n(k)$. Therefore we have $\sum_{n=1}^{k} s_n + t_n(\boldsymbol{s}) \leq \sum_{n=1}^{k-1} \hat{x}_n(k) + x_k^{\max}$. But $\sum_{n=1}^{k}(s_n + t_n(\boldsymbol{s})) \geq \sum_{n \in \text{supp}(\boldsymbol{t})} t_n(\boldsymbol{s}) = (1-\alpha)r$ since supp $(\boldsymbol{t}) \subseteq \{1, \ldots, k\}$. Therefore $(1-\alpha)r \leq \sum_{n=1}^{k-1} \hat{x}_n(k) + x_k^{\max}$. By Lemma 1, there exists a single-link-free-flow equilibrium for the instance $(N, (1-\alpha)r)$ supported on the first k links. Let $(\tilde{\boldsymbol{t}}, \tilde{\boldsymbol{m}})$ be such an equilibrium. The cost of this equilibrium is $(1-\alpha)r\ell_0$ where $\ell_0 \leq a_k$ is the free-flow latency of the last link in the support of $\tilde{\boldsymbol{t}}$. Thus $C(\tilde{\boldsymbol{t}}, \tilde{\boldsymbol{m}}) \leq (1-\alpha)ra_k$. Since by definition $(\bar{\boldsymbol{t}}, \bar{\boldsymbol{m}})$ is the *best Nash equilibrium* for the instance $(N, (1-\alpha)r)$ and has cost $(1-\alpha)ra_{\bar{k}}$, we must have $(1-\alpha)ra_{\bar{k}} \leq (1-\alpha)ra_k$, i.e., $a_{\bar{k}} \leq a_k$.

Lemma 4 *Under* $(\boldsymbol{x}, \boldsymbol{m})$*, the links* $\{1, \ldots, l-1\}$ *have greater (or equal) latency and hold less (or equal) flow than under* $(\bar{\boldsymbol{x}}, \bar{\boldsymbol{m}})$*, i.e.,* $\forall n \in \{1, \ldots, l-1\}$, $\ell_n(x_n, m_n) \geq \ell_n(\bar{x}_n, \bar{m}_n)$ *and* $x_n \leq \bar{x}_n$.

Proof Since $k \in \text{supp}\,(\boldsymbol{t})$, we have by definition of a Stackelberg strategy and its induced equilibrium that $\forall n \in \{1, \ldots, k-1\}$, $\ell_n(x_n, m_n) \geq \ell_k(x_k, m_k) \geq a_k$, see Eq. (25.5). We also have by definition of $(\bar{\boldsymbol{x}}, \bar{\boldsymbol{m}})$ and the resulting latencies given by Eq. (25.14), $\forall n \in \{1, \ldots, \bar{k}-1\}$, n is congested, and $\ell_n(x_n, m_n) = a_{\bar{k}}$. Thus using the fact that $k \geq \bar{k}$, we have $\forall n \in \{1, \ldots, \bar{k}-1\}$, $\ell_n(x_n, m_n) \geq a_k \geq a_{\bar{k}} = \ell_n(\bar{x}_n, \bar{m}_n)$, and $x_n \leq \hat{x}_n(k) \leq \hat{x}_n(\bar{k}) = \bar{x}_n$.

We have from Eq. (25.13) that $\forall n \in \{\bar{k}, \ldots, l-1\}$, n is in free-flow and at maximum capacity under $(\bar{\boldsymbol{x}}, \bar{\boldsymbol{m}})$ (i.e., $\bar{x}_n = x_n^{\max}$ and $\ell_n(\bar{x}_n) = a_n$). Thus $\forall n \in \{\bar{k}, \ldots, l-1\}$, $\ell_n(x_n, m_n) \geq a_n = \ell_n(\bar{x}_n, \bar{m}_n)$ and $x_n \leq x_n^{\max} = \bar{x}_n$. This completes the proof of the Lemma.

We can now show the desired inequality. We have

$$\begin{aligned} C(\boldsymbol{x}, \boldsymbol{m}) &= \sum_{n=1}^{N} x_n \ell_n(x_n, m_n) \\ &= \sum_{n=1}^{l-1} x_n \ell_n(x_n, m_n) + \sum_{n=l}^{N} x_n \ell_n(x_n, m_n) \\ &\geq \sum_{n=1}^{l-1} x_n \ell_n(\bar{x}_n, \bar{m}_n) + \sum_{n=l}^{N} x_n a_l \end{aligned} \tag{25.16}$$

where the last inequality is obtained using Lemma 4 and the fact that $\forall n \in \{l, \ldots, N\}$, $\ell_n(x_n, m_n) \geq a_n \geq a_l$. Then rearranging the terms, we have

$$C(\boldsymbol{x}, \boldsymbol{m}) \geq \sum_{n=1}^{l-1} (x_n - \bar{x}_n) \ell_n(\bar{x}_n, \bar{m}_n) + \sum_{n=1}^{l-1} \bar{x}_n \ell_n(\bar{x}_n, \bar{m}_n) + \sum_{n=l}^{N} x_n a_l.$$

Then we have $\forall n \in \{1, \ldots, l-1\}$,

$$(x_n - \bar{x}_n)(\ell_n(\bar{x}_n, \bar{m}_n) - a_l) \geq 0,$$

(by Lemma 4, $x_n - \bar{x}_n \leq 0$, and we have $\ell_n(\bar{x}_n, \bar{m}_n) \leq a_l$ by Eq. (25.14)). Thus

$$\sum_{n=1}^{l-1} (x_n - \bar{x}_n) \ell_n(\bar{x}_n, \bar{m}_n) \geq \sum_{n=1}^{l-1} (x_n - \bar{x}_n) a_l, \tag{25.17}$$

and we have

$$C(\boldsymbol{x}, \boldsymbol{m}) \geq \sum_{n=1}^{l-1} (x_n - \bar{x}_n) a_l + \sum_{n=1}^{l-1} \bar{x}_n \ell_n(\bar{x}_n, \bar{m}_n) + \sum_{n=l}^{N} x_n a_l$$
$$= a_l \left(\sum_{n=1}^{N} x_n - \sum_{n=1}^{l-1} \bar{x}_n \right) + \sum_{n=1}^{l-1} \bar{x}_n \ell_n(\bar{x}_n, \bar{m}_n)$$
$$= a_l \left(r - \sum_{n=1}^{l-1} \bar{x}_n \right) + \sum_{n=1}^{l-1} \bar{x}_n \ell_n(\bar{x}_n, \bar{m}_n).$$

But $a_l \left(r - \sum_{n=1}^{l-1} \bar{x}_n \right) = \bar{x}_l \ell_l(\bar{x}_l, \bar{m}_l)$ since $\operatorname{supp}(\bar{\boldsymbol{x}}) = \{1, \ldots, l\}$ and $\ell_l(\bar{x}_l, \bar{m}_l) = a_l$. Therefore

$$C(\boldsymbol{x}, \boldsymbol{m}) \geq \bar{x}_l \ell_l(\bar{x}_l, \bar{m}_l) + \sum_{n=1}^{l-1} \bar{x}_n \ell_n(\bar{x}_n, \bar{m}_n) = C(\bar{\boldsymbol{x}}, \bar{\boldsymbol{m}}).$$

This completes the proof of Theorem 1. □

Therefore the NCF strategy is an optimal Stackelberg strategy, and it can be computed in polynomial time since it is generated in linear time after computing the best Nash equilibrium BNE(N, $(1-\alpha)r$), which can be computed in $O(N^2)$.

The NCF strategy is, in general, not the unique optimal Stackelberg strategy. In the next section, we show that any optimal Stackelberg strategy can in fact be easily expressed in terms of the NCF strategy.

4.2 The Set of Optimal Stackelberg Strategies

In this section, we show that the set of optimal Stackelberg strategies $\mathrm{S}^\star(N, r, \alpha)$ can be generated from the NCF strategy. This shows in particular that the NCF strategy is robust, in a sense explained below.

Let $\bar{\boldsymbol{s}} = \mathrm{NCF}(N, r, \alpha)$ be the *non-compliant first* strategy, $\{(\bar{\boldsymbol{t}}, \bar{\boldsymbol{m}})\} = \mathrm{BNE}(N, (1-\alpha)r)$ be the Nash equilibrium induced by $\bar{\boldsymbol{s}}$, and $\bar{k} = \max \operatorname{supp}(\bar{\boldsymbol{t}})$ the last link in the support of the induced equilibrium, as defined above. By definition, the NCF strategy $\bar{\boldsymbol{s}}$ assigns zero compliant flow to links $\{1, \ldots, \bar{k} - 1\}$ and saturates links one by one, starting from $\bar{k}$ (see Eq. (25.7) and Fig. 25.4).

To give an example of an optimal Stackelberg strategy other than the NCF strategy, consider a strategy $\boldsymbol{s}$ defined by $\boldsymbol{s} = \bar{\boldsymbol{s}} + \boldsymbol{\varepsilon}$, where

$$\boldsymbol{\varepsilon} = \left(\varepsilon_1, 0, \ldots, 0, \overbrace{-\varepsilon_1}^{\bar{k}}, 0, \ldots, 0 \right)$$

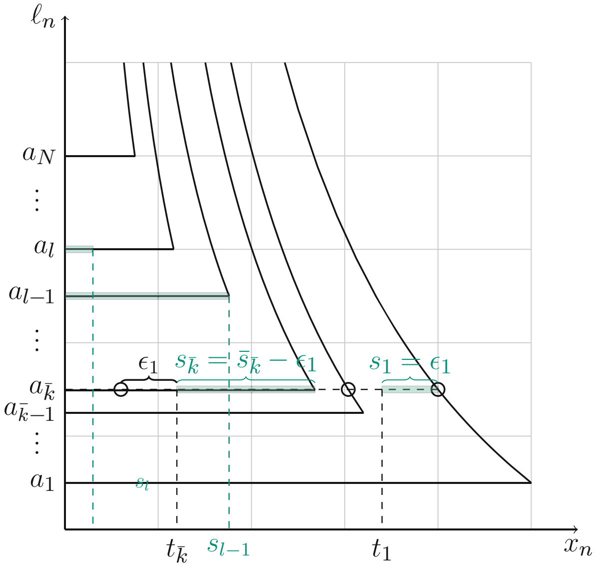


Fig. 25.6 Example of an optimal Stackelberg strategy $\boldsymbol{s} = \bar{\boldsymbol{s}} - \boldsymbol{\varepsilon}$. The *circles* show the best Nash equilibrium $(\bar{\boldsymbol{t}}, \bar{\boldsymbol{m}})$. The strategy $\boldsymbol{s}$ is highlighted in *green*

and is such that $s_1 = \varepsilon_1 \in [0, \hat{x}_1(\bar{k})]$ and $s_{\bar{k}} = \bar{s}_{\bar{k}} - \varepsilon_1 \geq 0$ (See Fig. 25.6). Strategy $\boldsymbol{s}$ will induce $\boldsymbol{t}(\boldsymbol{s}) = \bar{\boldsymbol{t}} - \boldsymbol{\varepsilon}$, and the resulting total cost is minimal since $C(\boldsymbol{s} + \boldsymbol{t}(\boldsymbol{s})) = C(\bar{\boldsymbol{s}} + \boldsymbol{\varepsilon} + \bar{\boldsymbol{t}} - \boldsymbol{\varepsilon}) = C(\bar{\boldsymbol{s}} + \bar{\boldsymbol{t}})$. This shows that $\boldsymbol{s}$ is an optimal Stackelberg strategy. More generally, the following holds:

Lemma 5 *Consider a Stackelberg strategy $\boldsymbol{s}$ of the form $\boldsymbol{s} = \bar{\boldsymbol{s}} + \boldsymbol{\varepsilon}$, where*

$$\boldsymbol{\varepsilon} = \left(\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_{\bar{k}-1}, -\sum_{n=1}^{\bar{k}-1} \varepsilon_n, \overset{\overset{\bar{k}+1}{\sqcap}}{0}, \ldots, 0 \right) \tag{25.18}$$

and $\boldsymbol{\varepsilon}$ is such that

$$\varepsilon_n \in [0, \hat{x}_n(\bar{k})] \qquad \forall n \in \left\{1, \ldots, \bar{k} - 1\right\} \tag{25.19}$$

$$\bar{s}_{\bar{k}} \geq \sum_{n=1}^{\bar{k}-1} \varepsilon_n. \tag{25.20}$$

Then $\boldsymbol{s}$ is an optimal Stackelberg strategy.

Proof We show that $\boldsymbol{s} = \bar{\boldsymbol{s}} + \boldsymbol{\varepsilon}$ is a feasible assignment of the compliant flow αr and that the induced equilibrium of the followers is $(\boldsymbol{t}(\boldsymbol{s}), \boldsymbol{m}(\boldsymbol{s})) = (\bar{\boldsymbol{t}} - \boldsymbol{\varepsilon}, \bar{\boldsymbol{m}})$.

Since $\sum_{n=1}^{N} \varepsilon_n = 0$ by definition (25.18) of $\boldsymbol{\varepsilon}$, we have $\sum_{n=1}^{N} s_n = \sum_{n=1}^{N} \bar{s}_n = \alpha r$. We also have

- $\forall n \in \{1, \ldots, \bar{k} - 1\}$, $s_n = \varepsilon_n \in [0, \hat{x}_n(\bar{k})]$ by Eq. (25.19). Thus $s_n \in [0, x_n^{\max}]$.
- $s_{\bar{k}} = \bar{s}_{\bar{k}} + \varepsilon_{\bar{k}} \geq 0$ by Eq. (25.20), and $s_{\bar{k}} \leq \bar{s}_{\bar{k}} \leq x_{\bar{k}}^{\max}$.
- $\forall n \in \{\bar{k} + 1, \ldots, N\}$, $s_n = \bar{s}_n \in [0, x_n^{\max}]$.

This shows that $\boldsymbol{s}$ is a feasible assignment. To show that $\boldsymbol{s}$ induces $(\bar{\boldsymbol{t}} - \boldsymbol{\varepsilon}, \bar{\boldsymbol{m}})$, we need to show that $\forall n \in \operatorname{supp}\left(\bar{\boldsymbol{t}} - \boldsymbol{\varepsilon}\right)$, $\forall k \in \{1, \ldots, N\}$,

$$\ell_n(\bar{s}_n + \varepsilon_n + \bar{t}_n - \varepsilon_n, \bar{m}_n) \leq \ell_k(\bar{s}_k + \varepsilon_k + \bar{t}_k - \varepsilon_k, \bar{m}_k)$$

This is true $\forall n \in \operatorname{supp}\left(\bar{\boldsymbol{t}}\right)$, by definition of $(\bar{\boldsymbol{t}}, \bar{\boldsymbol{m}})$ and Eq. (25.5). To conclude, we observe that $\operatorname{supp}\left(\bar{\boldsymbol{t}} - \boldsymbol{\varepsilon}\right) \subset \operatorname{supp}\left(\bar{\boldsymbol{t}}\right)$.

This shows that the NCF strategy is robust to perturbations: even if the strategy $\bar{\boldsymbol{s}}$ is not realized exactly, it may still be optimal if the perturbation $\boldsymbol{\varepsilon}$ satisfies the conditions given above.

The converse of the previous lemma is true. This gives a necessary and sufficient condition for optimal Stackelberg strategies, given in the following theorem.

Theorem 2 (Characterization of optimal Stackelberg strategies) *The set of optimal Stackelberg strategies* $S^\star(N, r, \alpha)$ *is the set of strategies* $\boldsymbol{s}$ *of the form* $\boldsymbol{s} = \bar{\boldsymbol{s}} + \boldsymbol{\varepsilon}$ *where* $\bar{\boldsymbol{s}} = \mathrm{NCF}(N, r, \alpha)$ *is the non-compliant first strategy, and* $\boldsymbol{\varepsilon}$ *satisfies Eqs.* (25.18)*,* (25.19)*, and* (25.20).

Proof We prove the converse of Lemma 5. Let $\boldsymbol{s} \in S^\star(N, r, \alpha)$ be an optimal Stackelberg strategy, $(\boldsymbol{t}, \boldsymbol{m}) = (\boldsymbol{t}(\boldsymbol{s}), \boldsymbol{m}(\boldsymbol{s}))$ the equilibrium of non-compliant flow induced by $\boldsymbol{s}$, $k = \max \operatorname{supp}(\boldsymbol{t})$ the last link in the support of $\boldsymbol{t}$, and $\boldsymbol{x} = \boldsymbol{s} + \boldsymbol{t}$ the total flow assignment.

We first show that $\boldsymbol{x} = \bar{\boldsymbol{x}}$. By optimality of both $\boldsymbol{s}$ and $\bar{\boldsymbol{s}}$, we have $C(\boldsymbol{x}, \boldsymbol{m}) = C(\bar{\boldsymbol{x}}, \bar{\boldsymbol{m}})$, and therefore inequalities (25.16) and (25.17) in the proof of Theorem 1 must hold with equality. In particular, to have equality in (25.16), we need to have

$$\sum_{n=1}^{l-1} x_n(\ell_n(x_n, m_n) - \ell_n(\bar{x}_n, \bar{m}_n)) + \sum_{n=l}^{N} x_n(\ell_n(x_n, m_n) - a_l) = 0. \tag{25.21}$$

The terms in both sums are nonnegative, therefore

$$x_n(\ell_n(x_n, m_n) - \ell_n(\bar{x}_n, \bar{m}_n)) = 0 \qquad \forall n \in \{1, \ldots, l-1\} \tag{25.22}$$

$$x_n(\ell_n(x_n, m_n) - a_l) = 0 \qquad \forall n \in \{l, \ldots, N\}, \tag{25.23}$$

and to have equality in (25.17), we need to have

$$(x_n - \bar{x}_n)(\ell_n(\bar{x}_n, \bar{m}_n) - a_l) = 0 \qquad \forall n \in \{1, \dots, l-1\}. \tag{25.24}$$

Let $n \in \{1, \dots, l-1\}$. From the expression (25.14) of the latencies under $\bar{\boldsymbol{x}}$, we have $\ell_n(\bar{x}_n, \bar{m}_n) < a_l$; thus from equality (25.24), we have $x_n - \bar{x}_n = 0$. Now let $n \in \{l+1, \dots N\}$. We have by definition of the latency functions, $\ell_n(x_n, m_n) \geq a_n > a_l$, thus from equality (25.23), $x_n = 0$. We also have from the expression (25.13), $\bar{x}_n = 0$. Therefore $x_n = \bar{x}_n$ $\forall n \neq l$, but since $\boldsymbol{x}$ and $\bar{\boldsymbol{x}}$ are both assignments of the same total flow r, we also have $x_l = \bar{x}_l$, which proves $\boldsymbol{x} = \bar{\boldsymbol{x}}$.

Next we show that $k = \bar{k}$. We have from the proof of Theorem 1 that $k \geq \bar{k}$. Assume by contradiction that $k > \bar{k}$. Then since $k \in \text{supp}(\boldsymbol{t})$, we have by definition of the induced followers' assignment in Eq. (25.5), $\forall n \in \{1, \dots, N\}$, $\ell_n(x_n, m_n) \geq \ell_k(x_k, m_k)$. And since $\ell_k(x_k, m_k) \geq a_k > a_{\bar{k}}$, we have (in particular for $n = \bar{k}$) $\ell_{\bar{k}}(x_{\bar{k}}, m_{\bar{k}}) > a_{\bar{k}}$, i.e., link $\bar{k}$ is congested under $(\bar{\boldsymbol{x}}, \bar{\boldsymbol{m}})$, thus $x_{\bar{k}} > 0$. Finally, since $\ell_{\bar{k}}(\bar{x}_{\bar{k}}, \bar{m}_{\bar{k}}) = a_{\bar{k}}$, we have $\ell_{\bar{k}}(\bar{x}_{\bar{k}}, \bar{m}_{\bar{k}}) > \ell_{\bar{k}}(\bar{x}_{\bar{k}}, \bar{m}_{\bar{k}})$. Therefore $x_{\bar{k}}(\ell_{\bar{k}}(x_{\bar{k}}, m_{\bar{k}}) - \ell_{\bar{k}}(\bar{x}_{\bar{k}}, \bar{m}_{\bar{k}})) > 0$, since $\bar{k} < k \leq l$; this contradicts (25.22).

Now let $\boldsymbol{\varepsilon} = \boldsymbol{s} - \bar{\boldsymbol{s}}$. We want to show that $\boldsymbol{\varepsilon}$ satisfies Eq. (25.18), (25.19), and (25.20).

First, we have $\forall n \in \left\{1, \dots, \bar{k}-1\right\}$, $\bar{s}_n = 0$, thus $\varepsilon_n = s_n - \bar{s}_n = s_n$. We also have $\forall n \in \left\{1, \dots, \bar{k}-1\right\}$, $0 \leq s_n \leq x_n$, $x_n = \bar{x}_n$ (since $\boldsymbol{x} = \bar{\boldsymbol{x}}$), and $\bar{x}_n = \hat{x}_n(\bar{k})$ (by Eq. (25.13)); therefore $0 \leq s_n \leq \hat{x}_n(\bar{k})$. This proves (25.19).

Second, we have $\forall n \in \left\{\bar{k}+1, \dots, N\right\}$, $t_n = \bar{t}_n = 0$ (since $k = \bar{k}$), and $x_n = \bar{x}_n$ (since $\boldsymbol{x} = \bar{\boldsymbol{x}}$); thus $\varepsilon_n = s_n - \bar{s}_n = x_n - t_n - \bar{x}_n + \bar{t}_n = 0$. We also have $\sum_{n=1}^{N} \varepsilon_n = 0$ since $\boldsymbol{s}$ and $\bar{\boldsymbol{s}}$ are assignments of the same compliant flow αr; thus $\varepsilon_{\bar{k}} = -\sum_{n \neq \bar{k}} \varepsilon_n = -\sum_{n=1}^{\bar{k}-1} \varepsilon_n$. This proves (25.18).

Finally, we readily have (25.20) since $s_{\bar{k}} \geq 0$ by definition of $\boldsymbol{s}$.

5 Price of Stability Under Optimal Stackelberg Routing

To quantify the inefficiency of Nash equilibria, and the improvement that can be achieved using Stackelberg routing, several metrics have been used including price of anarchy (Roughgarden and Tardos 2002, 2004) and price of stability (Anshelevich et al. 2004). We use price of stability as a metric, which is defined as the ratio between the cost of the best Nash equilibrium and the cost of the social optimum.[4] We start by characterizing the social optimum.

[4]Price of anarchy is defined as the ratio between the costs of the *worst* Nash equilibrium and the social optimum. For the case of nondecreasing latency functions, the price of anarchy and the price of stability coincide since all Nash equilibria have the same cost by the essential uniqueness property.

5.1 Characterization of Social Optima

Consider an instance (N, r) where the flow demand r does not exceed the maximum capacity of the network, i.e., $r \leq \sum_n x_n^{\max}$. A social optimal assignment is an assignment that minimizes the total cost function $C(\boldsymbol{x}, \boldsymbol{m}) = \sum_n x_n \ell_n(x_n, m_n)$, i.e., it is a solution to the following social optimum (SO) optimization problem:

$$\begin{aligned} \underset{\substack{\boldsymbol{x}\in\prod_{n=1}^{N}[0,x_n^{\max}] \\ \boldsymbol{m}\in\{0,1\}^N}}{\text{minimize}} \quad & \sum_{n=1}^{N} x_n \ell_n(x_n, m_n) \qquad (SO) \\ \text{subject to} \quad & \sum_{n=1}^{N} x_n = r \end{aligned}$$

Proposition 5 *$(\boldsymbol{x}^\star, \boldsymbol{m}^\star)$ is optimal for (SO) only if $\forall n \in \{1, \ldots, N\}$, $m_n^\star = 0$.*

Proof This follows immediately from the fact the latency on a link in congestion is always greater than the latency of the link in free-flow $\ell_n(x_n, 1) > \ell_n(x_n, 0)$ $\forall x_n \in (0, x_n^{\max})$.

As a consequence of the previous proposition, and using the fact that the latency is constant in free-flow, $\ell_n(x_n, 0) = a_n$, the social optimum can be computed by solving the following equivalent linear program:

$$\begin{aligned} \underset{\boldsymbol{x}\in\prod_{n=1}^{N}[0,x_n^{\max}]}{\text{minimize}} \quad & \sum_{n=1}^{N} x_n a_n \\ \text{subject to} \quad & \sum_{n=1}^{N} x_n = r \end{aligned}$$

Then since the links are ordered by increasing free-flow latency $a_1 < \cdots < a_N$, the social optimum is simply given by the assignment that saturates most efficient links first. Formally, if $k_0 = \max\left\{k | r \geq \sum_{n=1}^{k} x_n^{\max}\right\}$, then the social optimal assignment is given by $\boldsymbol{x}^\star = \left(x_1^{\max}, \ldots, x_{k_0}^{\max}, r - \sum_{n=1}^{k_0} x_n^{\max}, 0, \ldots, 0\right)$.

5.2 Price of Stability and Value of Altruism

We are now ready to derive the price of stability. Let $(\boldsymbol{x}^\star, \boldsymbol{0})$ denote the social optimum of the instance (N, r). Let $\bar{s}$ be the non-compliant first strategy NCF(N, r, α) and $(\boldsymbol{t}(\bar{s}), \boldsymbol{m}(\bar{s}))$ the induced equilibrium of the followers. The price of stability of the Stackelberg instance NCF(N, r, α) is

$$\text{POS}(N, r, \alpha) = \frac{C\,(\bar{\boldsymbol{s}} + \boldsymbol{t}(\bar{\boldsymbol{s}}), \boldsymbol{m}(\bar{\boldsymbol{s}}))}{C\,(\boldsymbol{x}^{\star}, \boldsymbol{0})},$$

where $\bar{\boldsymbol{s}}$ is the NCF strategy and $(\bar{\boldsymbol{t}}, \bar{\boldsymbol{m}})$ its induced equilibrium. The improvement achieved by optimal Stackelberg routing with respect to the Nash equilibrium ($\alpha = 0$) can be measured using the *value of altruism* (Aswani and Tomlin 2011), defined as

$$\text{VOA}\,(N, r, \alpha) = \frac{\text{POS}(N, r, 0)}{\text{POS}(N, r, \alpha)}.$$

This terminology refers to the improvement achieved by having a fraction α of altruistic (or compliant) players, compared to a situation where everyone is selfish. We give the expressions of price of stability and value of altruism in the case of a two-link network, as a function of the compliance rate $\alpha \in [0, 1]$ and demand r.

5.2.1 Case 1: $0 \le (1-\alpha)r \le x_1^{\max}$

In this case, link 1 can accommodate all the non-compliant flow; thus the induced equilibrium of the followers is

$$(\boldsymbol{t}(\bar{\boldsymbol{s}}), \boldsymbol{m}(\bar{\boldsymbol{s}})) = \big(((1-\alpha)r, 0), (0, 0)\big),$$

and by Eq. (25.7), the total flow induced by $\bar{\boldsymbol{s}}$ is $\bar{\boldsymbol{s}} + \boldsymbol{t}(\bar{\boldsymbol{s}}) = (x_1^{\max}, r - x_1^{\max})$ and coincides with the social optimum. Therefore, the price of stability is one.

5.2.2 Case 2: $x_1^{\max} < (1-\alpha)r \le x_2^{\max} + \hat{x}_1(2)$

Observe that this case can only occur if $x_2^{\max} + \hat{x}_1(2) > x_1^{\max}$. In this case, link 1 cannot accommodate all the non-compliant flow, and the induced Nash equilibrium $(\boldsymbol{t}(\bar{\boldsymbol{s}}), \boldsymbol{m}(\bar{\boldsymbol{s}}))$ is then supported on both links. It is equal to $(\boldsymbol{x}^{2,(1-\alpha)r}, \boldsymbol{m}^2) = \big((\hat{x}_1(2), (1-\alpha)r - \hat{x}_1(2)), (1, 0)\big)$, and the total flow is $\bar{\boldsymbol{s}} + \boldsymbol{t}(\bar{\boldsymbol{s}}) = (\hat{x}_1(2), r - \hat{x}_1(2))$, with total cost $a_2 r$ (Fig. 25.7b). The social optimum

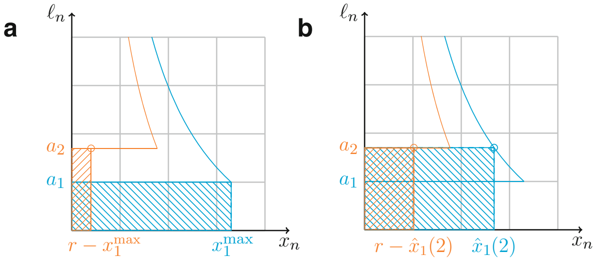


Fig. 25.7 Social optimum and best Nash equilibrium when the demand exceeds the capacity of the first link ($r > x_1^{\max}$). The area of the shaded regions represents the total costs of each assignment. (**a**) Social optimum. (**b**) Best Nash equilibrium

is $(\boldsymbol{x}^\star, \boldsymbol{m}^\star) = \big((x_1^{\max}, r - x_1^{\max}), (0,0)\big)$, with total cost $a_1 x_1^{\max} + a_2(r - x_1^{\max})$ (Fig. 25.7a). Therefore the price of stability is

$$\mathrm{POS}(2, r, \alpha) = \frac{r a_2}{r a_2 - x_1^{\max}(a_2 - a_1)} = \frac{1}{1 - \frac{x_1^{\max}}{r}\left(1 - \frac{a_1}{a_2}\right)}.$$

We observe that for a fixed flow demand $r > x_1^{\max}$, the price of stability is an increasing function of a_2/a_1. Intuitively, the inefficiency of Nash equilibria increases when the difference in free-flow latency between the links increases. And as $a_2 \to a_1$, the price of stability goes to 1.

When the compliance rate is $\alpha = 0$, the price of stability attains a supremum equal to a_2/a_1, at $r = (x_1^{\max})^+$ (Fig. 25.8a). This shows that selfish routing is most costly when the demand is slightly above critical value $r^{\mathrm{NE}}(1) = x_1^{\max}$. This also shows that for the general class of HQSF latencies on parallel networks, the price of stability is unbounded, since one can design an instance $(2, r)$ such that the maximal price of stability a_2/a_1 is arbitrarily large. Under optimal Stackelberg routing ($\alpha > 0$), the price of stability attains a supremum equal to $1/(\alpha + (1-\alpha)(a_1/a_2))$ at $r = \left(x_1^{\max}/(1-\alpha)\right)^+$. We observe in particular that the supremum is decreasing in α and that when $\alpha = 1$ (total control), the price of stability is identically one.

Therefore optimal Stackelberg routing can significantly decrease price of stability when $r \in (x_1^{\max}, x_1^{\max}/(1-\alpha))$. This can occur for small values of the compliance rate in situations where the demand slightly exceeds the capacity of the first link (Fig. 25.8c).

The same analysis can be done for a general network: given the latency functions on the links, one can compute the price of stability as a function of the flow demand r and the compliance rate α, using the form of the NCF strategy together with Algorithm 1 to compute the BNE. Computing the price of stability function reveals critical values of demand, for which optimal Stackelberg routing can lead to a significant improvement. This is discussed in further detail in the next section, using an example network with *four* links.

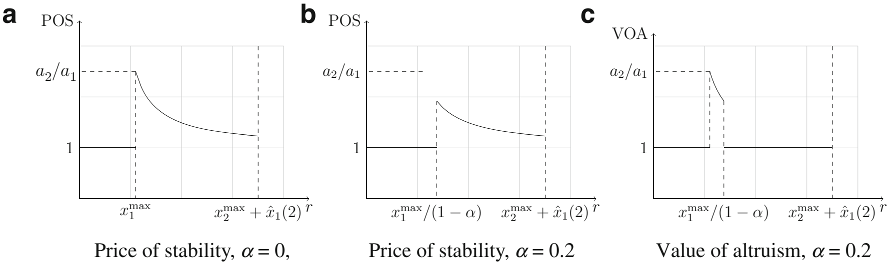


Fig. 25.8 Price of stability and value of altruism on a two-link network. Here we assume that $\hat{x}_1(2) + x_2^{\max} > x_1^{\max}$. (**a**) Price of stability, $\alpha = 0$. (**b**) Price of stability, $\alpha = 0.2$. (**c**) Value of altruism, $\alpha = 0.2$

6 Numerical Results

In this section, we apply the previous results to a scenario of freeway traffic from the San Francisco Bay Area. Four parallel highways are chosen starting in San Francisco and ending in San Jose: I-101, I-280, I-880, and I-580 (Fig. 25.9). We analyze the inefficiency of Nash equilibria and show how optimal Stackelberg routing (using the NCF strategy) can improve the efficiency.

Figure 25.10 shows the latency functions for the highway network, assuming a triangular fundamental diagram for each highway. Under free-flow conditions, I-101 is the fastest route available between San Francisco and San Jose. When I-101 becomes congested, other routes represent viable alternatives.

We computed price of stability and value of altruism (defined in the previous section) as a function of the demand r for different compliance rates. The results are shown in Fig. 25.11. We observe that for a fixed compliance rate, the price of stability is piecewise continuous in the demand (Fig. 25.11a), with discontinuities corresponding to an increase in the cardinality of the equilibrium's support (and a link transitioning from free-flow to congestion). If a transition exists for link n, it occurs at critical demand $r = r^{(\alpha)}(n)$, defined to be the infimum demand r such that n is congested under the equilibrium induced by NCF(N, r, α).

It can be shown that $r^{(\alpha)}(n) = r^{\text{NE}}(n)/(1-\alpha)$, and we have in particular $r^{\text{NE}}(n) = r^{(0)}(n)$. Therefore if a link n is congested under best Nash equilibrium ($r > r^{\text{NE}}(n)$), optimal Stackelberg routing can decongest n if $r^{(\alpha)}(n) \geq r$. In particular, when the demand is slightly above critical demand $r^{(0)}(n)$, link n can be decongested with a small compliance rate. This is illustrated by the numerical values of price of stability on Fig. 25.11a, where a small compliance rate ($\alpha = 0.05$) achieves high value of altruism when the demand is slightly above the critical values.

Fig. 25.9 Map of a simplified parallel highway network model, connecting San Francisco to San Jose

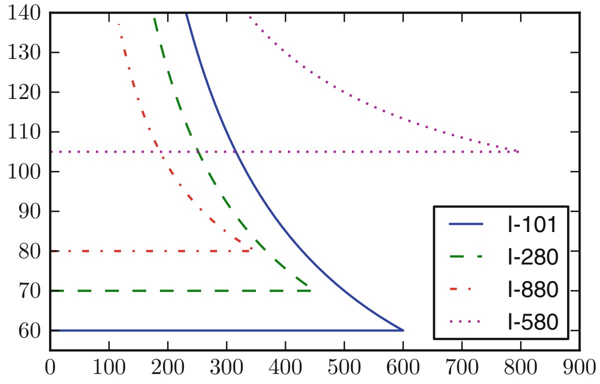


Fig. 25.10 Latency functions on an example highway network. Latency is in minutes, and demand is in cars/minute

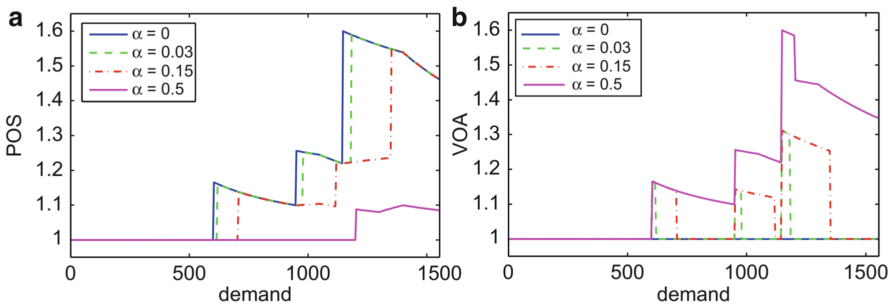


Fig. 25.11 Price of stability and value of altruism as a function of the demand r for different values of compliance rate α. (**a**) Price of stability. (**b**) Value of altruism

This shows that optimal Stackelberg routing can achieve a significant improvement in efficiency, especially when the demand is near one of the critical values $r^{(\alpha)}(n)$.

Figure 25.12 shows price of stability and value of altruism as a function of the demand $r \in [0, r^{\text{NE}}(N)]$ and compliance rate $\alpha \in [0, 1]$. We observe in particular that for a fixed value of demand, price of stability is a piecewise constant function of α. Computing this function can be useful for efficient planning and control, since it informs the central coordinator of the critical compliance rates that can achieve a strict improvement. For instance, if the demand on the example network is 1100 cars/min, price of stability is constant for compliance rates $\alpha \in [0.14, 0.46]$. Therefore if a compliance rate greater than 0.46 is not feasible, the controller may prefer to implement a control strategy with $\alpha = 0.14$, since further increasing the compliance rate will not improve efficiency and may incur additional external cost (due to incentivizing more drivers, for example).

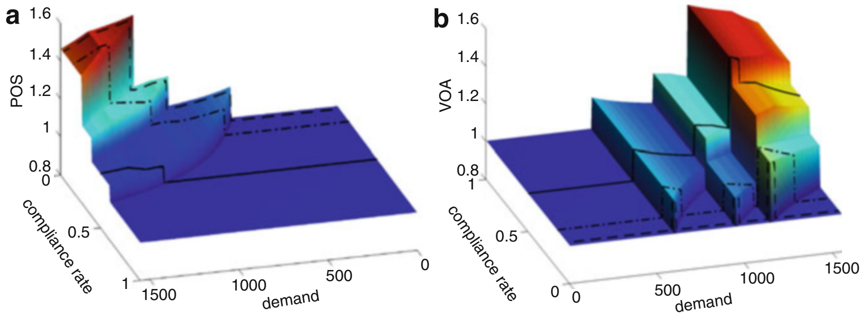


Fig. 25.12 Price of stability (**a**) and value of altruism (**b**) as a function of the compliance rate α and demand r. Iso-α lines are plotted for $\alpha = 0.03$ (*dashed*), $\alpha = 0.15$ (*dot-dashed*), and $\alpha = 0.5$ (*solid*)

7 Summary and Concluding Remarks

Motivated by the fundamental diagram of traffic for transportation networks, this chapter has introduced a new class of latency functions (HQSF) to model congestion with horizontal queues and studied the resulting Nash equilibria for non-atomic routing games on parallel networks. We showed that the essential uniqueness property does not hold for HQSF latencies and that the number of equilibria is at most $2N$. We also characterized the best Nash equilibrium. In the Stackelberg routing game, we proved that the non-compliant first (NCF) strategy is optimal and that it can be computed in polynomial time. Table 25.1 summarizes the main differences between the classical setting (vertical queues) and the HQSF setting.

We illustrated these results using an example network for which we computed the decrease in inefficiency that can be achieved using optimal Stackelberg routing. This example showed that when the demand is near critical values $r^{\text{NE}}(n)$, optimal Stackelberg routing can achieve a significant improvement in efficiency, even for small values of compliance rate.

On the one hand, these results show that careful routing of a small compliant population can dramatically improve the efficiency of the network. On the other hand, they also indicate that for certain demand and compliance values, Stackelberg routing can be completely ineffective. Therefore identifying the ranges where optimal Stackelberg routing does improve the efficiency of the network is crucial for effective planning and control.

This framework offers several directions for future research: the work presented here only considers parallel networks under static assumptions (constant flow demand r and static equilibria), and one question is whether these equilibria are stable in the dynamic sense and how one may steer the system from one equilibrium to a better one – consider, for example, the case where the players are in a congested equilibrium and assume a coordinator has control over a fraction of the flow. Can the

Table 25.1 Main assumptions and results for the Stackelberg routing game on a parallel network

Setting	Vertical queues	Horizontal queues, single valued in free-flow (HQSF)
Model	$x \mapsto \ell(x)$ latency is a function of the flow $x \in [0, x^{\max}]$	$(x, m) \mapsto \ell(x, m)$ latency is a function of the flow $x \in [0, x^{\max}]$ and the congestion state $m \in \{0, 1\}$
Assumptions	$x \mapsto \ell(x)$ is continuously nondecreasing $x \mapsto x\ell(x)$ is convex	$x \mapsto \ell(x, 0)$ is single valued. $x \mapsto \ell(x, 1)$ is continuously decreasing $\lim_{x \to x^{\max}} \ell(x, 1) = \ell(x^{\max}, 0)$
Set of Nash equilibria	Essential uniqueness: if x, x' are Nash equilibria, then $C(x) = C(x')$ (Beckmann et al. 1956)	No essential uniqueness in general The number of Nash equilibria is at most $2N$ (Proposition 4) The best Nash equilibrium is a single-link-free-flow equilibrium (Lemma 2)
Optimal Stackelberg strategy	NP hard (Roughgarden 2001)	The NCF strategy is optimal and can be computed in polynomial time (Theorem 1) The set of all optimal Stackelberg strategies can be computed in polynomial time (Theorem 2)

coordinator steer the system to a single-link-free-flow equilibrium by decongesting a link? And what is the minimal compliance rate needed to achieve this?

References

Anshelevich E, Dasgupta A, Kleinberg J, Tardos E, Wexler T, Roughgarden T (2004) The price of stability for network design with fair cost allocation. In: 45th annual IEEE symposium on foundations of computer science, Berkeley, pp 295–304

Aswani A, Tomlin C (2011) Game-theoretic routing of GPS-assisted vehicles for energy efficiency. In: American control conference (ACC). IEEE, San Francisco, pp 3375–3380

Babaioff M, Kleinberg R, Papadimitriou CH (2009) Congestion games with malicious players. Games Econ Behav 67(1):22–35

Beckmann M, McGuire CB, Winsten CB (1956) Studies in the economics of transportation. Yale University Press, New Haven

Benaïm M (2015) On gradient like properties of population games, learning models and self reinforced processes. Springer, Cham, pp 117–152

Blackwell D (1956) An analog of the minimax theorem for vector payoffs. Pac J Math 6(1):1–8

Blum A, Even-Dar E, Ligett K (2006) Routing without regret: on convergence to Nash equilibria of regret-minimizing algorithms in routing games. In: Proceedings of the twenty-fifth annual ACM symposium on principles of distributed computing (PODC'06). ACM, New York, pp 45–52

Blume LE (1993) The statistical mechanics of strategic interaction. Games Econ Behav 5(3): 387–424

Boulogne T, Altman E, Kameda H, Pourtallier O (2001) Mixed equilibrium for multiclass routing games. IEEE Trans Autom Control 47:58–74

Caltrans (2010) US 101 South, corridor system management plan. http://www.dot.ca.gov/hq/tpp/offices/ocp/pp_files/new_ppe/corridor_planning_csmp/CSMP_outreach/US-101-south/2_US-101_south_CSMP_Exec_Summ_final_22511.pdf

Cesa-Bianchi N, Lugosi G (2006) Prediction, learning, and games. Cambridge University Press, Cambridge

Dafermos S (1980) Traffic equilibrium and variational inequalities. Transp Sci 14(1):42–54

Dafermos SC, Sparrow FT (1969) The traffic assignment problem for a general network. J Res Natl Bur Stand 73B(2):91–118

Daganzo CF (1994) The cell transmission model: a dynamic representation of highway traffic consistent with the hydrodynamic theory. Transp Res B Methodol 28(4):269–287

Daganzo CF (1995) The cell transmission model, part II: network traffic. Transp Res B Methodol 29(2):79–93

Drighès B, Krichene W, Bayen A (2014) Stability of Nash equilibria in the congestion game under replicator dynamics. In: 53rd IEEE conference on decision and control (CDC), Los Angeles, pp 1923–1929

Evans CL (1998) Partial differential equations. Graduate studies in mathematics. American Mathematical Society, Providence

Farokhi F, Johansson HK (2015) A piecewise-constant congestion taxing policy for repeated routing games. Transp Res B Methodol 78:123–143

Fischer S, Vöcking B (2004) On the evolution of selfish routing. In: Algorithms–ESA 2004. Springer, Berlin, pp 323–334

Fischer S, Räcke H, Vöcking B (2010) Fast convergence to Wardrop equilibria by adaptive sampling methods. SIAM J Comput 39(8):3700–3735

Fox MJ, Shamma JS (2013) Population games, stable games, and passivity. Games 4(4):561–583

Friesz TL, Mookherjee R (2006) Solving the dynamic network user equilibrium problem with state-dependent time shifts. Transp Res B Methodol 40(3):207–229

Greenshields BD (1935) A study of traffic capacity. Highw Res Board Proc 14:448–477

Hannan J (1957) Approximation to Bayes risk in repeated plays. Contrib Theory Games 3:97–139

Hofbauer J, Sandholm WH (2009) Stable games and their dynamics. J Econ Theory 144(4):1665–1693.e4
Kleinberg R, Piliouras G, Tardos E (2009) Multiplicative updates outperform generic no-regret learning in congestion games. In: Proceedings of the 41st annual ACM symposium on theory of computing, Bethesda, pp 533–542. ACM
Korilis YA, Lazar AA, Orda A (1997a) Achieving network optima using Stackelberg routing strategies. IEEE/ACM Trans Netw 5:161–173
Korilis YA, Lazar AA, Orda A (1997b) Capacity allocation under noncooperative routing. IEEE Trans Autom Control 42:309–325
Koutsoupias E, Papadimitriou C (1999) Worst-case equilibria. In: Proceedings of the 16th annual conference on theoretical aspects of computer science. Springer, Heidelberg, pp 404–413
Krichene S, Krichene W, Dong R, Bayen A (2015a) Convergence of heterogeneous distributed learning in stochastic routing games. In: 53rd annual allerton conference on communication, control and computing, Monticello
Krichene W, Drighès B, Bayen A (2015b) Online learning of Nash equilibria in congestion games. SIAM J Control Optim (SICON) 53(2):1056–1081
Krichene W, Castillo MS, Bayen A (2016) On social optimal routing under selfish learning. IEEE Trans Control Netw Syst PP(99):1–1
Lam K, Krichene W, Bayen A (2016) On learning how players learn: estimation of learning dynamics in the routing game. In: 7th international conference on cyber-physical systems (ICCPS), Vienna
Lebacque JP (1996) The Godunov scheme and what it means for first order traffic flow models. In: International symposium on transportation and traffic theory, Lyon, pp 647–677
LeVeque RJ (2007) Finite difference methods for ordinary and partial differential equations: steady-state and time-dependent problems. Classics in applied mathematics. Society for Industrial and Applied Mathematics, Philadelphia
Lighthill MJ, Whitham GB (1955) On kinematic waves. II. A theory of traffic flow on long crowded roads. Proc R Soc Lond A Math Phys Sci 229(1178):317
Lo HK, Szeto WY (2002) A cell-based variational inequality formulation of the dynamic user optimal assignment problem. Transp Res B Methodol 36(5):421–443
Monderer D, Shapley LS (1996) Potential games. Games Econ Behav 14(1):124–143
Ozdaglar A, Srikant R (2007) Incentives and pricing in communication networks. In: Nisan N, Roughgarden T, Tardos E, Vazirani V (eds) Algorithmic game theory. Cambridge University Press, Cambridge
Papageorgiou M, Blosseville JM, Hadj-Salem H (1989) Macroscopic modelling of traffic flow on the Boulevard Périphérique in Paris. Transp Res B Methodol 23(1):29–47
Papageorgiou M, Blosseville J, Hadj-Salem H (1990) Modelling and real-time control of traffic flow on the southern part of boulevard peripherique in paris: part I: modelling. Transp Res A Gen 24(5):345–359
Richards PI (1956) Shock waves on the highway. Oper Res 4(1):42–51
Roughgarden T (2001) Stackelberg scheduling strategies. In: Proceedings of the thirty-third annual ACM symposium on theory of computing, Heraklion, pp 104–113. ACM
Roughgarden T, Tardos E (2002) How bad is selfish routing? J ACM (JACM) 49(2):236–259
Roughgarden T, Tardos E (2004) Bounding the inefficiency of equilibria in nonatomic congestion games. Games Econ Behav 47(2):389–403
Sandholm WH (2001) Potential games with continuous player sets. J Econ Theory 97(1):81–108
Sandholm WH (2010) Population games and evolutionary dynamics. MIT Press, Cambridge
Schmeidler D (1973) Equilibrium points of nonatomic games. J Stat Phys 7(4):295–300
Shamma JS (2015) Learning in games. Springer, London, pp 620–628
Swamy C (2007) The effectiveness of Stackelberg strategies and tolls for network congestion games. In: Proceedings of the eighteenth annual ACM-SIAM symposium on discrete algorithms (SODA'01), New Orleans. Society for Industrial and Applied Mathematics, Philadelphia, pp 1133–1142

Thai J, Hariss R, Bayen A (2015) A multi-convex approach to latency inference and control in traffic equilibria from sparse data. In: American control conference (ACC), Chicago, pp 689–695

Wang Y, Messmer A, Papageorgiou M (2001) Freeway network simulation and dynamic traffic assignment with METANET tools. Transp Res Rec J Transp Res Board 1776(-1):178–188

Wardrop JG (1952) Some theoretical aspects of road traffic research. In: ICE proceedings: engineering divisions, vol 1, pp 325–362

Weibull JW (1997) Evolutionary game theory. MIT Press, Cambridge

Work DB, Blandin S, Tossavainen OP, Piccoli B, Bayen AM (2010) A traffic model for velocity data assimilation. Appl Math Res eXpress 2010(1):1

Communication Networks: Pricing, Congestion Control, Routing, and Scheduling

26

Srinivas Shakkottai and R. Srikant

Contents

Abstract
This chapter considers three fundamental problems in the general area of communication networks and their relationship to game theory. These problems are (i) allocation of shared bandwidth resources, (ii) routing across shared links, and (iii) scheduling across shared spectrum. Each problem inherently

S. Shakkottai (✉)
Department of Electrical and Computer Engineering, Texas A&M University, College Station, TX, USA
e-mail: sshakkot@tamu.edu

R. Srikant
Department of Electrical and Computer Engineering, University of Illinois at Urbana-Champaign, Urbana, IL, USA
e-mail: rsrikant@illinois.edu

T. Başar, G. Zaccour (eds.), *Handbook of Dynamic Game Theory*,
https://doi.org/10.1007/978-3-319-44374-4_29

involves agents that experience negative externalities under which the presence of one degrades the utility perceived by others. Two approaches to solving such problems are (i) to find a globally optimal allocation and simply implement it in a fait accompli fashion, and (ii) request information from the competing agents (traffic flows) and construct a mechanism to allocate resources. Often, only the second option is viable, since a centralized solution using complete information might be impractical (or impossible) with many millions of competing flows, each one having private information about the application that it corresponds to. Hence, a game theoretical analysis of these problems is natural. In what follows, we will present results on each problem and characterize the efficiency loss that results from the mechanism employed.

Keywords
Communication networks · Utility maximization · Congestion control · Traffic routing · Packet scheduling

1 Introduction

Communication networks are possibly the largest control systems in existence. They consist of many millions of flows interacting with each other as well as the network infrastructure, and competing for available capacity on wired and wireless links. The most commonly used communication network today is the Internet, which is illustrated in schematic form in Fig. 26.1. Here, we have *flows*, each of which is between two hosts, commonly a Web server and an end user. A flow typically consists of data packets from the server end, and acknowledgements back from the end user. The packets traverse a *route* that consists of *communication links*, with the direction of packet forwarding being determined by routers. The network infrastructure itself is owned by several different Internet service providers, each of which implements different traffic shaping, scheduling, and pricing policies in their particular network.

In Fig. 26.1, the server must choose the rate of transmission of packets to each of the flows based on the available capacity in the end-to-end route of the flow. It obtains information on the state of the links though feedback from the routers that could either mark or drop packets if the rate of packet arrivals on a particular link is too high. This feedback is returned back to the server using the acknowledgement packets, which then results in the server reducing or increasing the rate based on the feedback received.

Control actions at different network routers are generally implemented via simple algorithms. The individual routers usually do not maintain per-flow information and take decisions on which packets to forward at each instant of time based on their perception of fairness and stability across packets arriving from different directions. These decisions result in packets being dropped or marked, which forms the feedback returned to the server. With the increasing prevalence of software-defined networking, however, it is increasingly possible to take decisions on a per-flow basis.

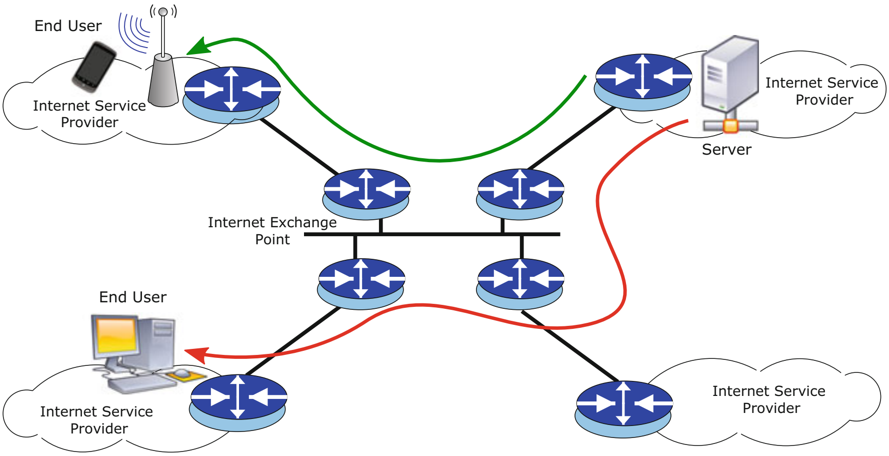


Fig. 26.1 Internet flows between a server and end users

Wireless links usually form the last hop of flow. The number of competing users in the case of WiFi is usually low, and simple randomized access is employed. However, in the cellular data context, the usage of wireless links across competing flows is carefully scheduled due to the limited availability of wireless spectrum. However, such scheduling decisions again may result in dropped or marked packets.

Communication networks have usually been designed with the idea of distributedly achieving an overall goal of fair resource allocation and good quality of service to all flows assuming a cooperative setup. For a more comprehensive study of Internet control systems, the reader is referred to Srikant (2004) and Shakkottai and Srikant (2007). However, given the inherent resource constraints in the system and the desire for each end user to get the best possible quality of service, it is natural to try to understand the system from a game theoretic perspective. This approach is becoming increasingly popular, particularly in the case of wireless resource allocation due to the perceived scarcity of the resource.

This chapter deals with the analysis of communication networks from the perspective of strategic agents. Our focus is on two main problems, namely, (i) allocation of capacity to competing flows and (ii) routing decisions by flows. We consider three questions with different interaction models between the users as follows:

1. *Nash equilibrium. Resource allocation across a finite number of agents.* Here, we consider the problem of resource sharing via an auction mechanism that requests bids from a finite set of agents and performs an allocation based on the responses. The efficient solution is to allocate resources in such a way that overall utility of the system is maximized. Our objective will be to quantify the efficiency loss

in this system where the agents are nonstrategic in that they do not consider the existence of other agents versus the case that they are strategic.

2. *Wardrop equilibrium. Routing with infinite agents.* Here, we consider the problem encountered in choosing between different routes on a per-packet basis. A single packet has effectively no impact on any other packet, but the total traffic along each route would impact the delay seen by each packet using that route. Each packet desires to reach the destination in the shortest possible time, while the overall efficient solution is to minimize the total delay in the system.
3. *Mean Field Equilibrium. Repeated resource allocation with infinite agents.* Finally, we consider the problem of a repeated auction of resources between agents that only compete against random subsets of other agents for any particular resource. Here, agents are unaware of whom they would compete against in the next auction, and hence model their competitors via a belief about what they are likely to bid. As before, we are interested in the question of whether an efficient allocation can be achieved at each step.

2 Congestion Control

As we described in the previous section, a communication network can be identified with a set of sources of traffic (or users) $\mathcal{R}$ and a set of links $\mathcal{L}$. Each link $l \in \mathcal{L}$ has a finite capacity c_l. Each source desires to communicate with a destination in the network and uses a route $r \subset \mathcal{L}$ to reach its destination. Thus, we can equivalently associate each source with the route it uses, and we will interchangeably refer to both by $r \in \mathcal{R}$. We denote the utility that a user obtains from transmitting data on route r at rate x_r by $U_r(x_r)$. The typical assumption is that the utility function is continuously differentiable, nondecreasing, and strictly concave. We further assume that $U(0) \geq 0$. The assumption on concavity of the utility function represents the fact that a user quality of experience has diminishing returns to per unit rate allocated on the links that it uses. For instance, the perceived value of a rate increase by 1 Mbps is much greater when the user has a low rate than at a high rate.

Consider a network planner who is interested in allocating resources to users with the goal of maximizing the sum of the users' utilities. The network planner can do this only if he knows the utility functions of all the users in the network, or if there is an incentive for the users to reveal their utility functions truthfully. In this section, we will first discuss an incentive mechanism called the Vickrey-Clarke-Groves (VCG) which makes it profitable for the users to reveal their true utilities to the central network planner. However, the amount of information that needs to be conveyed by the users and the amount of computation required on the part of the network planner make it difficult to implement the VCG mechanism. One can design a mechanism based on the idea of distributed resource allocation using a gradient approach. We call this the Kelly mechanism. However, this mechanism is truth-revealing only under the assumption that the network is very large so that it is difficult for each user to estimate its impact on the price decided by the network. Users that are unaware of their effect on the price are called *price taking*. On the other hand, in a small network such as a single link, a user may be able to assess its impact on the network price. In such a case, the user may act *strategically*, i.e., act in such a manner that influences the price to maximize its own benefit. We will review results that show that the inefficiency of the Kelly mechanism with strategic users is bounded by 25%, i.e., the Kelly mechanism loses at the most a factor of $1/4$ compared to the maximum possible network utility.

2.1 VCG Mechanism

Consider a network planner who wants to solve a utility maximization problem, where each user r is associated with a route r:

$$\max_{x \geq 0} \sum_r U_r(x_r)$$

subject to

$$\sum_{r:l \in r} x_r \leq c_l, \forall l.$$

Here, x_r is the rate allocated to user r, who has a strictly concave utility function given by U_r and c_l is the capacity of link l. Also, we use the notation $l \in r$ to denote the fact that link l is part of route r.

Suppose that the network planner asks each user to reveal their utilities and user r reveals its utility function as $\tilde{U}_r(x_r)$, which may or may not be the same as $U_r(x_r)$. Users may choose to lie about their utility function to get a higher rate than they would get by revealing their true utility function. Let us suppose that the network solves the maximization problem

$$\max_{x \geq 0} \sum_r \tilde{U}_r(x_r)$$

subject to

$$\sum_{r:l \in r} x_r \leq c_l, \forall l$$

and allocates the resulting optimal solution $\tilde{x}_r$ to user r. In return for allocating this rate to user r, the network charges a certain price p_r. The price is calculated as follows. The network planner calculates the reduction in the sum of the utilities obtained by other users in the network due to the presence of user r and collects this amount as the price from user r. Specifically, the network planner first obtains the optimal solution $\{\bar{x}_s\}$ to the following problem:

$$\max_{x \geq 0} \sum_{s \neq r} \tilde{U}_s(x_s)$$

subject to

$$\sum_{s \neq r : l \in s} x_s \leq c_l, \forall l.$$

In other words, the network planner first solves the utility maximization problem without including user r. The price p_r is then computed as

$$p_r = \sum_{s \neq r} \tilde{U}(\bar{x}_s) - \sum_{s \neq r} \tilde{U}(\tilde{x}_s),$$

which is the difference of sum utilities of all other users without $(\{\bar{x}\})$ and with $(\{\tilde{x}\})$ the presence of user r. The network planner announces this mechanism to the users of the network, i.e., the network planner states that once the users reveal their utilities, it will allocate resources by solving the utility maximization problem and will charge a price p_r to user r. Now the question for the users is the following: what utility function should user r announce to maximize its *payoff*? The payoff is the utility minus the price:

$$U_r(\tilde{x}_r) - p_r.$$

We will now see that an optimal strategy for each user is to truthfully reveal its utility function. We will show this by proving that announcing a false utility function cannot increase the payoff for user r.

Suppose user r reveals its utility function truthfully, while the other users may or may not. In this case, the payoff for user r is given by

$$\mathcal{U}^t = U_r(\tilde{x}_r^t) - \left(\sum_{s \neq r} \tilde{U}_s(\bar{x}_s^t) - \sum_{s \neq r} \tilde{U}_s(\tilde{x}_s^t) \right),$$

where $\{\tilde{x}_s^t\}$ is the allocation given to the users by the network planner and $\{\bar{x}_s^t\}$ is the solution of the network utility maximization problem when user r is excluded from the network. The superscript t indicates that user r has revealed its utility function truthfully. Next, suppose that user r lies about its utility function and denote the network planner's allocation by $\tilde{x}^l$. The superscript l indicates that user r has lied. Now, the payoff for user r is given by

$$\mathcal{U}^l = U_r(\tilde{x}_r^l) - \left(\sum_{s \neq r} \tilde{U}_s(\bar{x}_s^t) - \sum_{s \neq r} \tilde{U}_s(\tilde{x}_s^l) \right).$$

If truth-telling were not optimal, $\mathcal{U}^l > \mathcal{U}^t$. If this were true, by comparing the two expressions for $\mathcal{U}^t$ and $\mathcal{U}^l$, we get

$$U_r(\tilde{x}_r^l) + \sum_{s \neq r} \tilde{U}_s(\tilde{x}_s^l) > U_r(\tilde{x}_r^t) + \sum_{s \neq r} \tilde{U}_s(\tilde{x}_s^t),$$

which contradicts the fact that $\tilde{x}^t$ is the optimal solution to

$$\max_{x \geq 0} U_r(x_r) + \sum_{s \neq r} \tilde{U}_s(x_s)$$

subject to the capacity constraints. Thus, truth-telling is optimal under the VCG mechanism. Note that truth-telling is optimal for user r independent of the strategies of the other users. A strategy which is optimal for a user independent of the strategies of other users is called a *dominant strategy* in game theory. Thus, truth-telling is a dominant strategy under the VCG mechanism.

In the above discussion, note that $\bar{x}$ is somewhat irrelevant to the pricing mechanism. One could have chosen any $\bar{x}$ that is a function of the strategies of all users other than r in computing the price for user r, and the result would still hold, i.e., truth-telling would still be a dominant strategy. The reason for this is that the expression for $\mathcal{U}^t - \mathcal{U}^l$ is independent of $\bar{x}$ since the computation of $\bar{x}$ does not use either $U_r(.)$ or $\tilde{U}_r(.)$. Another point to note is that truth-telling is an optimal strategy. Given the strategies of all the other users, there may be other strategies for user r that are optimal as well. Such user strategies may result in allocations that are not optimal from the point of view of the network planner.

We have thus established that truth-telling is optimal under a VCG mechanism. However, the VCG mechanism is not used in networks. The reason for this is twofold:

- Each user is asked to reveal its utility function. Thus, an entire function has to be revealed by each user, which imposes a significant communication complexity in the information exchange required between the users and the network planner.
- The network planner has to solve many maximization problems: one to compute the resource allocation and one for user to compute the user's price. Each of these optimization problems can be computationally quite expensive to solve in a centralized manner.

In the next subsection, we show how one can design a distributed mechanism for utility maximization.

2.2 Kelly Mechanism

One simple method to reduce the communication burden required to exchange information between the users and the network planner is to ask the users to submit bids which are amounts that the users are willing to pay for the resource, i.e., the link capacities in the network. We refer to the mechanism of submitting bids for resource in the context of network resource allocation as the Kelly mechanism. We will describe the Kelly mechanism only for the case of a single link with capacity c.

Let the bid of user r be denoted by w_r. Given the bids, suppose that the network computes a price per unit amount of the resource as

$$q \triangleq \frac{\sum_k w_k}{c}. \tag{26.1}$$

and allocates an amount of resource x_r to user r according to $x_r = w_r/q$. This is a weighted proportionally fair allocation since it is equivalent to maximizing $\sum_r w_r \log x_r$ subject to the resource constraint.

The payoff that the user obtains is given by

$$U_r\left(\frac{w_r}{q}\right) - w_r \tag{26.2}$$

Since the user is rational, it would try to maximize the payoff. We assume that users are price taking and hence are unaware of the effect that their bids have on the price per unit resource. As far as they know, the central authority is selling them a resource at a price q, regardless of what their bid might be. The system is illustrated in Fig. 26.2.

What would a user bid given that the price per unit resource is q? Clearly, the user would try to maximize the payoff and try to solve the problem:

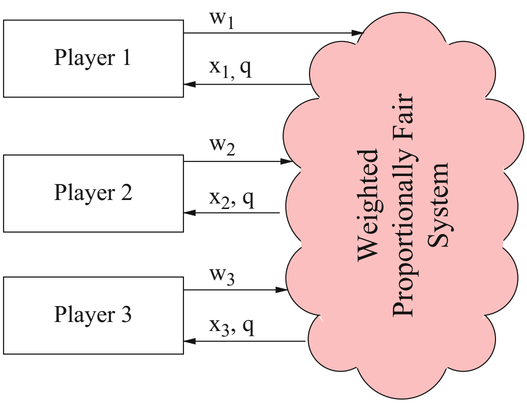


Fig. 26.2 In the price-taking paradigm, users are unaware of both how the price is calculated and of each others' bids. As far as the users are concerned, the system is a fair black box that accepts bids and outputs the price and their respective allocations

$$\max_{w_r \geq 0} U_r\left(\frac{w_r}{q}\right) - w_r.$$

Assuming, as we did earlier, that $U_r(x_r) \to -\infty$ as $x_r \to 0$, the optimal value of w_r is strictly greater than zero and is given by

$$U_r'\left(\frac{w_r}{q}\right) = q. \tag{26.3}$$

Since we know that $x_r = w_r/q$, the above equation can be equivalently written in two other forms:

$$q = U_r'(x_r) \tag{26.4}$$

and

$$w_r = x_r U_r'(x_r). \tag{26.5}$$

Now the utility maximization problem that the network planner wants to solve is given by

$$\max_{x \geq 0} \sum_r U_r(x_r),$$

subject to $\sum_r x_r \leq c$. The solution to the problem satisfies the KKT optimality conditions given by

$$U_r'(x_r) = q, \sum_r x_r = c \tag{26.6}$$

The weighted proportionally fair mechanism ensures that (26.6) is satisfied (we do not allocate more resource than is available). Also, we have just seen that coupled with rational price-taking users, the mechanism results in an allocation that satisfies (26.4), which is identical to (26.6). Thus, there exists a solution to the system of price-taking users that we call x^T, q^T (using T to denote "taking") that achieves the desired utility maximization by using the weighted proportionally fair mechanism.

2.2.1 Relation to Decentralized Utility Maximization

Now, suppose that the network wants to reduce its computational burden. Then, it can compute the price according to the following *dual algorithm*:

$$\dot{q} = \left(\sum_r x_r - c\right)_q^+.$$

Here, $(g(x))_y^+$ denotes

$$(g(x))_y^+ = \begin{cases} g(x), & y > 0, \\ \max(g(x), 0), & y = 0, \end{cases}$$

We use projection above to ensure that p_l never goes negative. The algorithm can be interpretted as the system responding to resource usage by giving back a price, which it increases or decreases depending on whether the resource is overutilized or underutilized, respectively. User r is then allocated a rate w_r/q in proportion to its bid w_r. Given the price, we have already seen that the user r's rational bid w_r assuming that it is a price-taking user is given by

$$U_r'\left(\frac{w_r}{q}\right) = q,$$

which is the same as $U'(x_r) = q$.

It is easy to show that the price update along with such a user response converges to the optimal solution of the network utility maximization problem. Hence, the algorithm can be thought of as a decentralized implementation of the Kelly mechanism. In summary, the Lagrange multiplier is a pricing incentive for users to behave in a socially responsible manner assuming that they are price taking.

2.3 Strategic or Price-Anticipating Users

In the price-anticipating paradigm, the users are aware of the effect that their bid has on the price of the resource. In this case the problem faced by the users is a game in which they attempt to maximize their individual payoffs *anticipating* the price change that their bid would cause. Each user strategically tries to maximize its payoff given by

$$P_r(w_r; w_{-r}) = \begin{cases} U_r\left(\frac{w_r}{\sum_k w_k} c\right) - w_r & \text{if } w_r > 0 \\ U_r(0) & \text{if } w_r = 0, \end{cases} \tag{26.7}$$

where w_{-r} is the vector of all bids, except w_r.

Our game is a system wherein there are R users and each user r can make a bid $w_r \in \mathbb{R}^+ \cup \{0\}$, i.e., the game consists of deciding a nonnegative real number. A *strategy* is a rule by which a user would make his bid. In our case, the strategy that users might use could be S_r ="set the bid $w_r = w_r^S$", where $w_r^S \in \mathbb{R}^+ \cup \{0\}$ is some constant. Since the strategy recommends playing the same bid all the time, it is called a *pure strategy*. A *strategy profile* is an element of the product-space of strategy spaces of each user, denoted by $\mathcal{S}$. We denote the strategy profile corresponding to all users using the strategy of setting their bids based on the vector w^G by $S \in \mathcal{S}$. We would like to know if a particular strategy profile is stable in some sense. For example, we might be interested in knowing if everyone knows everyone else's strategy, would they want to change in some way. The concept of the *Nash equilibrium* formally defines one kind of stability.

Definition 1. A pure strategy Nash equilibrium is a strategy profile from which no user has a unilateral incentive to change his strategy.

Note the term "unilateral". This means that users do not collude with each other – they are interested solely in their own payoffs. If they find that there is no point changing from the strategy that they are currently using, they would continue to use it and so remain at equilibrium. How do we tell if the strategy profile S defined above is actually a Nash equilibrium? The answer is to check the payoff obtained by using it. We denote the bid recommended by strategy S_r to user r as w_r^S and similarly the bid recommended by any other strategy G by w_r^G. So the users would not unilaterally deviate from the strategy profile S if

$$P_r(w_r^S; w_{-r}^S) \geq P_r(w_r^G; w_{-r}^S), \tag{26.8}$$

which would imply that the strategy profile S is a Nash equilibrium. We would like to know if there exists a vector w^S such that strategy profile S that recommends playing that vector would be a Nash equilibrium, i.e., whether there exists w^S that satisfies (26.8). We first find the conditions that need to be satisfied by the desired w^S and then show that there indeed exists a unique such vector.

Now, our first observation is that w^S must have at least two positive components. On the one hand, if there were exactly one user with a positive bid, it would want to decrease the bid towards zero and yet have the whole resource to itself (so there can't be exactly one positive bid). On the other hand, if there were no users with a positive bid, there would be an incentive for all users to increase their bids to some nonzero value to capture the resource. The next observation is that since w^S has at least two positive components, and since $w_r/(\sum_{k\neq r} w_k + w_r)$ is strictly increasing in w_r if there are, the payoff function is strictly concave in w_r. Assuming that the utility functions are continuously differentiable, so this means that for each user k, the maximizer w^S of (26.7) satisfies the conditions

$$U_r'\left(\frac{w^S}{\sum_k w_k^S}c\right)\left(1-\frac{w_r^S}{\sum_k w_k^S}\right) = \frac{\sum_k w_k^S}{c}, \text{ if } w_k^S > 0 \tag{26.9}$$

$$U_r'(0) \leq \frac{\sum_k w_k^S}{c}, \text{ if } w_k^S = 0, \tag{26.10}$$

which are obtained by simply differentiating (26.7) and setting to 0 (or ≤ 0 if $w_r^S = 0$) and multiplying by $\sum_k w_k^S / c$.

We now have the set of conditions (26.9) and (26.10) that must be satisfied by the bids that the Nash strategy profile suggests. But we don't know yet if there actually exists any such vector w^S that would satisfy the conditions. How do we go about showing that such a vector actually exists? Consider the conditions again. They look just like the KKT first-order conditions of a constrained optimization problem. Perhaps we could construct the equivalent optimization problem to which these are indeed the KKT conditions? Then if the optimization problem has a unique solution, the solution would be the desired vector of bids w^S. Consider the constrained optimization problem of maximizing

$$\sum_k \hat{U}_k(x_k), \tag{26.11}$$

subject to the constraints

$$\sum_k x_k \le c, \quad x_k \ge 0 \quad \forall k = 1, 2, 3 \ldots, R \tag{26.12}$$

where the utility function $\hat{U}(.)$ is defined as

$$\hat{U}_k(x_k) = \left(1 - \frac{x_k}{c}\right) U_k(x_k) + \left(\frac{x_k}{c}\right)\left(\frac{1}{x_k}\int_0^{x_k} U_k(z)dz\right). \tag{26.13}$$

It easy to see that $\hat{U}(.)$ is concave and increasing in $0 \le x_k \le c$ by differentiating it, which yields $\hat{U}_k'(x_k) = U_k'(x_k)(1 - x_k/c)$. Since $U_k(.)$ is concave and strictly increasing, we know that $U_k'(x_k) > 0$ and that $U_k'(.)$ is nonincreasing. Hence, we conclude that $\hat{U}_k'(x_k)$ is nonnegative and strictly decreasing in k over the region $0 \le x_k \le c$ as required.

We verify that the KKT first-order conditions are identical in form to the conditions (26.9) and (26.10). Directly from the optimization problem above, we have that there exists a unique vector w and a scalar (the Lagrange multiplier) λ such that

$$U_r'(x_k)\left(1 - \frac{x_k}{c}\right) = \lambda, \quad \text{if } x_k > 0 \tag{26.14}$$

$$U_k'(0) \le \lambda, \quad \text{if } x_k = 0 \tag{26.15}$$

$$\sum_k x_k = c \tag{26.16}$$

We check that at least two components of x above are positive. We have from (26.16), at least one of the $x_k > 0$. If only a single component $x_r > 0$ with all others being 0, then from (26.14) we have $\lambda = 0$, which in turn means from (26.15) that $U_k'(0) \le 0$ for some k. This is impossible since $U_k(.)$ was assumed to be concave, strictly increasing for all k. Then we see that as desired, the above conditions are identical to the Nash conditions with $\lambda = \sum_k w_k^S/c$ and $x_k = cw_k/\sum_k w_k$. Thus, we see that even in the case of price-anticipating users, there exists unique solution. We will denote the resulting allocation x^S, the S being used to denote “strategy.”

Notice that the solution found applies to an optimization problem that is different to the one in the price-taking case, so we would expect that in terms of utility maximization, the price-anticipating solution would probably be worse than the price-taking case. We will show how to bound the worst-case performance of the price-anticipating case in the next subsection. Since in the price-anticipating case, all users play strategically with complete knowledge, and what the central authority does is to allocate resources in a weighted proportionally fair manner, the system can be likened to anarchy with users operating with minimal control. We refer to the inefficiency in such a system as *the price of anarchy*.

We now examine the impact of price-anticipating users on the network utility, i.e., the sum of the utilities of all the users in the network. We use the superscript T to denote solution to the network utility maximization problem (we use T since the optimal network utility is also the network utility assuming price-taking users), and the superscript S to denote the solution for the case of price-anticipating users. We will show the following result.

Theorem 1. *Under the assumptions on the utility function given at the beginning of Sect. 2,*

$$\sum_r U_r\left(x_r^S\right) \geq \frac{3}{4} \sum_r U_r\left(x_r^T\right). \tag{26.17}$$

Hence, the price of playing a game versus the optimal solution is no greater than $1/4$ *of the optimal solution.*

Proof. The proof of this theorem consists of two steps:

- Showing that the worst-case scenario for the price-anticipating paradigm is when the utility functions U_r are all linear.
- Minimizing the value of the game under the above condition.

Step 1

Using concavity of $U_r(.)$ for any user k, we have for the general allocation vector z with $\sum_k z_k \leq c$, that $U_r\left(x_k^T\right) \leq U_k(z_r) + U_k'(z_k)\left(x_k^T - z_k\right)$. This means that

$$\frac{\sum_r U_r(z_r)}{\sum_r U_r(x_r^T)} \geq \frac{\sum_r (U_r(z_r) - U_r'(z_r)z_r) + \sum_r U_r'(z_r)z_r}{\sum_r (U_r(z_r) - U_r'(z_r)z_r) + \sum_r U_r'(z_r)x_r^T}$$

Since we know that $\sum_k x_k^T = c$, we know that $\sum_k U_k'(z_k)x_k^T \leq (\max_k U_r'(z_k))c$. Using this fact in the above equation, we obtain

$$\frac{\sum_k U_k(z_k)}{\sum_k U_k(x_r^T)} \geq \frac{\sum_k (U_k(z_k) - U_k'(z_k)z_k) + \sum_k U_k'(z_k)z_k}{\sum_k (U_k(z_k) - U_k'(z_k)z_k) + (\max_k U_k'(z_k))c}$$

The term $\sum_k (U_k(z_k) - U_k'(z_k)z_k)$ is nonnegative by concavity of U_k, and the assumption that $U(0) \geq 0$, which means that

$$\frac{\sum_k U_k(z_k)}{\sum_k U_k(x_k^T)} \geq \frac{\sum_k U_k'(z_k)z_k}{(\max_k U_k'(z_k))c} \tag{26.18}$$

The above inequality is of interest as it compares the utility function with its linear equivalent. If we substitute $z = x^S$ in (26.18), we obtain

$$\frac{\sum_k U_k(x_k^S)}{\sum_k U_k(x_k^T)} \geq \frac{\sum_k U_k'(x_k^S)x_k^S}{(\max_k U_k'(x_k^S))c} \tag{26.19}$$

Now, we notice that the left-hand side of the expression above is the price of anarchy that we are interested in, while the right-hand side of the expression looks like the price of anarchy for the case where the utility function is linear, i.e., when $U_r(x_r) = U_r'(x_r^S)x_r \triangleq \bar{U}_r(x_r)$. We verify this observation by noting that since the conditions in (26.14), (26.15), and (26.16) are linear, the numerator is the aggregate utility with price-anticipating users when the utility function is $\bar{U}(.)$. Also, the denominator is the maximum aggregate that can be achieved with the utility function $\bar{U}(.)$, which means that it corresponds to the price-taking case for utility $\bar{U}(.)$. Thus, we see that the linear utility function of the sort described above necessarily has a lower total utility than any other type of utility function.

Step 2

Since the worst-case scenario is for linear utility functions, we may take $U_r(x_r) = \alpha_r x_r$. From Step 1, the price of anarchy, i.e., the ratio of aggregate utility at the Nash equilibrium to the aggregate utility at social optimum is then given by

$$\frac{\sum_k \alpha_k x_k^S}{\{\max_k \alpha_k\}c}. \tag{26.20}$$

Without loss of generality, we may take the $\max_k \alpha_k = 1$ and $c = 1$. Since this means that the denominator of the above expression is 1, to find the worst-case ratio, we need to find $\alpha_2, \alpha_3, \ldots, \alpha_R$ such that the numerator is minimized. This would directly give the price of anarchy. So the objective is to

$$\min_{\{x^S,\alpha\}} x_1^S + \sum_{r=2}^{R} \alpha_r x_r^S \tag{26.21}$$

$$\text{subject to } \alpha_k \left(1 - x_k^S\right) = 1 - x_1^S, \text{ if } x_k^S > 0 \tag{26.22}$$

$$\alpha_k \leq 1 - x_1^S, \text{ if } x_k^S = 0 \tag{26.23}$$

$$\sum_k x_k^S = 1 \tag{26.24}$$

$$0 \leq \alpha_k \leq 1, \; k = 2, \ldots R \tag{26.25}$$

$$x_k^S \geq 0, \; k = 1, \ldots, R \tag{26.26}$$

Notice that the constraints on the above optimization problem follow from (26.14), (26.15), and (26.16) to ensure that x^S is a allocation vector that a Nash strategy profile would suggest. Since only the users with nonzero allocations contribute to the utility, we can consider the system with $N \leq R$ users, with every user getting a nonzero allocation. Equivalently, we could just assume that all R users get a nonzero allocation and observe what happens as R increases. Then $\alpha_k(1 - x_k^S) = 1 - x_1^S$ holds for all users in the new system, which implies $\alpha_k = (1 - x_1^S)/(1 - x_k^S)$. Let us fix x_1^S for now and minimize over the remaining x_r^S. Then we have

$$\min_{\{x_r^S : r \neq 1\}} x_1^S + \sum_{k=2}^{R} \frac{x_k^S(1 - x_1^S)}{1 - x_k^S} \tag{26.27}$$

$$\text{subject to} \sum_{k=2}^{R} x_r^S = 1 - x_1^S \tag{26.28}$$

$$0 \leq x_k^S \leq x_1^S, \quad k = 2, \ldots, R \tag{26.29}$$

The above problem is well defined only if $x_1^S \geq 1/R$ (otherwise the last constraint will be violated upon minimization). If we assume this condition, by symmetry, the minimum value occurs for all users $2, \ldots, R$ getting the same allocation equal to $(1 - x_1^S)/(R - 1)$. Substituting this value into the problem, we only have to minimize over x_1^S, i.e., the problem is now given by

$$\min_{x_1^S} x_1^S + (1 - x_1^S)^2 \left(1 - \frac{1 - x_1^S}{R - 1}\right)^{-1} \tag{26.30}$$

$$\text{subject to } 0 \leq x_1^S \leq 1. \tag{26.31}$$

The objective function above is decreasing in R and so the lowest value would occur as $R \to \infty$. So we finally have very simple problem to solve, namely,

$$\min_{x_1^S} x_1^S + (1 - x_1^S)^2 \tag{26.32}$$

$$\text{subject to } 0 \leq x_1^S \leq 1 \tag{26.33}$$

By differentiation we can easily see that the solution to the above problem is for $d_1^S = 1/2$, which yields a worst-case aggregate of 3/4. Thus, we see that the aggregate utility falls by no more than 25% when the users are price anticipating. □

3 Flow Routing

In the last section, we considered the problem of strategic agents that try to maximize their individual throughputs, and saw that the *the price of anarchy* can be tightly upper bounded in a game of such strategic behavior. We consider a different game involving routing of flows though a network in this section. As we saw in the last section, the load on a link can be thought of as imposing a cost on all the flows on that link. This cost can be thought of as the delay experienced by each packet of the flow, which in turn causes a reduction in the quality of the service being supported by that flow. Suppose that a set of routes is available between a source and destination and each packet makes a decision on which route to take based on the perceived delay on each alternative. What would be the effects of such per-packet selfish routing?

Figure 26.3 illustrates the setup of an example routing problem proposed by Pigou. Here, there is a total flow of 1 unit, between S and D. We can think of this flow as the total number of packets per second that are being injected into the network. Let the delay per unit flow function on link A be $p_A(y_A) = 1$, where y_A is flow on link A. Similarly, the delay per unit flow on link B is $p_B(y_B) = y_B$, with the corresponding flow being y_B. Thus, the upper route has a fixed delay regardless of the flow on it, whereas the lower one has a delay proportional to the flow on it. The total flow is $y_A + y_B = 1$.

We can characterize the total delay experienced by an average packet under selfish routing by deriving the equilibrium flows on the links under such selfish dynamics. We make the assumption that the flows are infinitely divisible, in which case the decision of each infinitesimal unit of flow's decision has a vanishingly small impact on the delay on any link. Effectively, this is as if each packet is price taking and simply greedily chooses the route that shows the smallest delay. Then the equilibrium conditions would simply be that any routes that are in use would have the same per-unit delay, since otherwise some of the flow would be diverted to a

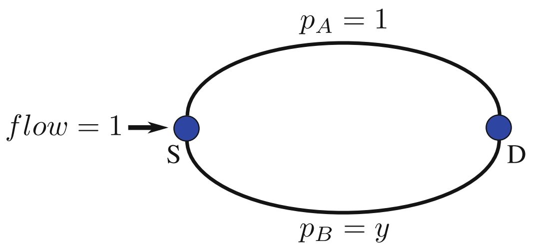


Fig. 26.3 A Pigovian Network

route that shows a smaller delay until equalization occurs. Hence, for *selfish routing* at equilibrium

- if A and B are both used then $p_A(y_A) = p_B(y_B)$.
- if only A is used, $p_A(y_A) \leq p_B(y_B)$.
- if only B is used, $p_B(y_B) \leq p_A(y_A)$.

Such a pair (y_A, y_B) is said to be in *Wardrop equilibrium*. In this example, the Wardrop equilibrium is $(y_A, y_B) = (0, 1)$. The average delay experienced by a packet is

$$\frac{y_A p_A(y_A) + y_B p_B(y_B)}{y_A + y_B} = y_A p_A(y_A) + y_B p_B(y_B) = 1. \tag{26.34}$$

Now, let us consider what would happen if a network planner were to route the flows in such a way that the average delay of the system as a whole is minimized. We refer to this case as *socially optimal routing*. We can determine the socially optimal flow assignment by choosing (y_A, y_B) to solve

$$\min y_A + y_B^2$$

subject to

$$y_A + y_B = 1,\ y_A \geq 0,\ y_B \geq 0$$

The problem is easy to solve by differentiating the objective function and setting it equal to zero as follows:

$$\min_{0 \leq y_B \leq 1} (1 - y_B) + y_B^2 \tag{26.35}$$

$$\Rightarrow -1 + 2y_B = 0 \Rightarrow y_B = 1/2 \tag{26.36}$$

$$\Rightarrow \text{optimal cost} = 1/2 + 1/4 = 3/4 \tag{26.37}$$

Combining the two results in (26.34) and (26.37), we have a characterization of the *PoA (price of anarchy)* as

$$\frac{\text{Optimal cost}}{\text{Cost under selfish routing}} = \frac{4}{3}.$$

The question arises as to whether the same kind of result would apply in the case of flow routing in a general network? For instance, one might think that in a network with a large number of routes to choose from, perhaps the delay incurred even with selfish routing might be comparatively small. We next present an example that shows that the intuition that adding links, even if the cost function of the additional link is zero can actually increase the delay of a system with selfish routing. This apparently paradoxical result is known as Braess' paradox, named after its discoverer.

3.1 Braess' Paradox

Consider the network shown in Fig. 26.4. We have the typical problem of routing a unit flow in a network with four links and a combination of fixed and linear link delay functions. Using the same logic that we used earlier, the Wardrop equilibrium must be such that all routes with a nonzero flow must have the same unit delay. By symmetry, we have that the Wardrop equilibrium is $(y_1, y_2) = (1/2, 1/2)$. The average delay per packet is $3/2 \cdot 1/2 + 3/2 \cdot 1/2 = 3/2$.

Now, let us add a link with zero delay (regardless of flow on it) between A and B, as shown in Fig. 26.5. We see immediately that Wardrop equilibrium is to route all the flow on $S \to A \to B \to D$. The average delay corresponding to this flow routing is $1 + 1 = 2$. We observe that the addition of a link with zero delay has increased the average delay under selfish routing!

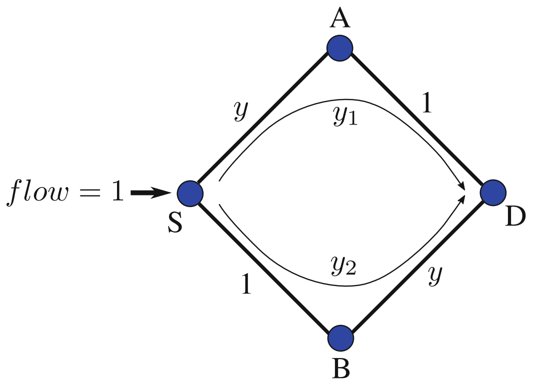


Fig. 26.4 A flow routing example with four links

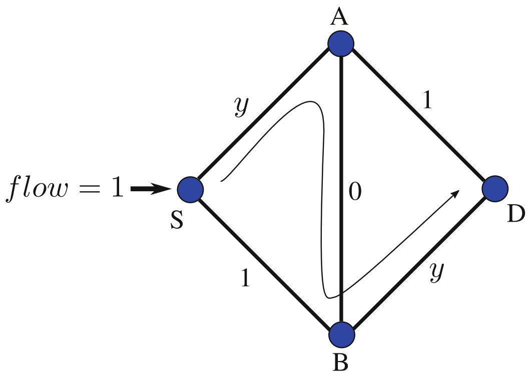


Fig. 26.5 Braess' Paradox. Adding a link with zero delay increases average delay

3.2 Flow Routing Game

We now consider the problem of tightly bounding the price of anarchy of selfish flow routing in a general network. We will focus only on linear cost functions. Surprisingly, it turns out that the bound of $4/3$ that we derived for the Pigovian network with some specific cost functions not only turns out to be accurate for a Pigovian network with general affine cost functions but also turns out to be the bound for a general network with affine cost functions. This subsection is devoted to proving that result.

Let F be the total flow between source, S and destination, D in a network. Let $\mathcal{R}$ be the set of routes between S and D. As before, we will use the notation $l \in r$ to indicate that link l is a part of route r. Also, x_r denotes the flow on route r, while y_l denotes flow on link l with

$$y_l = \sum_{r:l \in r} x_r.$$

The cost (or delay) of a route r per unit of flow on it is denoted by

$$q_r(\mathbf{y}) = \sum_{l \in r} p_l(y_l),$$

where p_l is the cost of link l, and is a nondecreasing function and $\mathbf{y} = (y_1, y_2, \cdots y_l, \cdots)$.

We define the *Socially optimal routing* problem as follows:

$$\min \sum_r \left(\sum_l p_l(y_l) \right) x_r \tag{26.38}$$

$$s.t. \tag{26.39}$$

$$\sum_{r:l \in r} x_r = y_l \tag{26.40}$$

$$\sum_r x_r = F \tag{26.41}$$

$$x_r, y_l \geq 0 \ \forall r, l \tag{26.42}$$

Definition 2. A Wardrop equilibrium is a vector $\mathbf{y} = \{y_l\}$ such that, if $x_r > 0$ for some $r \in \mathcal{R}$, then $q_r(\mathbf{y}) \leq q'_r(\mathbf{y}), \forall r' \in \mathcal{R}$.

In other words, a route has nonzero flow only if it is a minimum-cost route. Does a Wardrop equilibrium exist in the flow routing game? In order to answer this question, we will rewrite the conditions of the Wardrop equilibrium as the solution of a convex optimization problem to which a solution has to exist. First, according to the definition, the Wardrop equilibrium is equivalent to the condition that there exists a $\lambda \geq 0$ such that

$$q_r(y) = \lambda \; if \; x_r > 0 \tag{26.43}$$

$$\text{and } q_r(y) \geq \lambda \; if x_r = 0 \tag{26.44}$$

We will now show that the above condition is identical to solving the following problem.

$$\min \sum_l \int_0^{y_l} p_l(y)dy \tag{26.45}$$

$$s.t. \sum_{r:l \in r} x_r = y_l \tag{26.46}$$

$$\sum_{r \in R} x_r = F \tag{26.47}$$

$$x_r \geq 0, y_l \geq 0 \tag{26.48}$$

In order to show the equivalence between the conditions defined by (26.43), (26.44) and (26.45), (26.46), (26.47), (26.48), we will characterize the solution of the latter problem and show that it exactly satisfies the former. Using (26.46), y_l can be eliminated from the problem yielding

$$\min \sum_l \int_0^{\sum_{r:l \in r} x_r} p_l(y)dy \tag{26.49}$$

$$s.t. \sum_{r \in R} x_r = F \tag{26.50}$$

$$x_r \geq 0 \; \forall r. \tag{26.51}$$

First, we note that since $p_l(y)$ is a nondecreasing function of y

$$\int_0^{y_l} p_l(y)dy$$

is a convex function. Hence, the above formulation is a convex optimization problem. The Lagrange dual of the above problem is

$$\min_{x_r \geq 0} \underbrace{\sum_l \int_0^{\sum_{r:l \in r} x_r} p_l(y)dy}_{V(\mathbf{x})} - \lambda \left(\sum_r x_r - F \right), \tag{26.52}$$

where we have used the notation $\mathbf{x} = (x_1, x_2, \cdots, x_r, \cdots)$. The first-order KKT conditions corresponding to the solution are, therefore,

$$\frac{\partial V}{\partial x_r} = \lambda \; if x_r \geq 0 \tag{26.53}$$

$$\frac{\partial V}{\partial x_r} \geq \lambda \; if x_r = 0 \tag{26.54}$$

$$\frac{\partial V}{\partial x_r} = \sum_{l \in r} p_l(y_l)), \tag{26.55}$$

which are identical to the definition of the Wardrop equilibrium in (26.43) and (26.44). Thus, (26.45), (26.46), (26.47), and (26.48) can be used as an alternative definition of a Wardrop equilibrium.

3.2.1 PoA for Linear Latency Functions: Pigovian Network

We now reconsider the Pigovian network with two routes that we saw earlier, except with more general (but still affine) cost functions and any nonnegative flow F, as shown in Fig. 26.6. Let y_l be the flow on any link l in a graph. Assume that $p_l(y_l) = a_l y_l + b_l (a_l, b_l \geq 0)$. Also, let α be the worst-case PoA in a Pigou network. It is easily seen that the Wardrop equilibrium places all the flow on the top link and the cost is $(aF + b)F = aF^2 + bF$.

In order to find the socially optimal flows, we need to solve

$$\min_{0 \leq y \leq F} (ay + b)y + (aF + b)(F - y). \tag{26.56}$$

Note that we immediately have $\min_{0 \leq y \leq F} \Leftrightarrow \min_{y \geq 0}$ in this case. Let us ignore the fact that $y \geq 0$ and simply try to find the global minimum. Differentiating, we have

$$\Rightarrow 2ay + b - aF - b = 0 \tag{26.57}$$

$$\Rightarrow y = F/2, \tag{26.58}$$

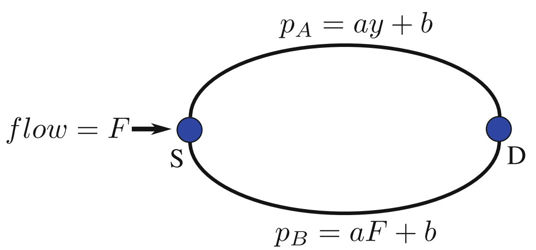


Fig. 26.6 A Pigovian network with linear delay functions

i.e., the solution satisfies $y \geq 0$. The cost of the social optimal is

$$= (aF/2 + b)F/2 + (aF + b)F/2 \tag{26.59}$$

$$= 3aF/4 + b. \tag{26.60}$$

Finally, we can identify the Price of Anarchy as

$$\alpha = \max_{F \geq 0, a \geq 0, b \geq 0} \frac{(aF + b)F}{F\left(\frac{3aF}{4} + b\right)} \tag{26.61}$$

$$= 4/3. \tag{26.62}$$

Note that the PoA is achieved when $F \to \infty$.

3.2.2 PoA for Linear Latency Functions: General Network

We now extend the results of the Pigovian network to a general network. The main result is stated below.

Theorem 2. *Consider any network with affine link delay costs, with $a_l, b_l \geq 0$. The PoA is upper bounded by* $4/3$.

Proof. Let $\hat{\mathbf{x}}$ be the WE and $\mathbf{x}^*$ be socially optimal for a network with a flow between a source and destination of F. If $\hat{x}_r > 0, q_r(\hat{\mathbf{x}}) \leq q_{r'}(\hat{\mathbf{x}})\ \forall r'$, which implies that for all used routes $q_r(\hat{x_r})$ are equal. Call this value as M. Thus,

$$\sum_r q_r(\hat{\mathbf{x}})\hat{x}_p = \sum_{r:\hat{x}_r \neq 0} q_r(\hat{\mathbf{x}})\hat{x}_r = MF. \tag{26.63}$$

Now, since the allocation $\mathbf{x}^*$ might use routes not used in $\hat{\mathbf{x}}$,

$$\sum_r q_r(\hat{x}_r)x_r^* \geq M \sum_r x_r^* = MF \tag{26.64}$$

$$\Rightarrow \sum_r q_r(\hat{x})(x_r^* - \hat{x}_r) \geq 0, \tag{26.65}$$

i.e., if all route costs are fixed as $q_r(\hat{x})$, then Wardrop equilibrium is the smallest cost.

Recall from our discussion on the Pigovian network that

$$\frac{4}{3} = \max_{F \geq 0, a \geq 0, b \geq 0} \frac{Fp(F)}{\min_{0 \leq y \leq F} yp(y) + (F - y)p(F)} \tag{26.66}$$

$$= \max_{F \geq 0, a \geq 0, b \geq 0} \max_{0 \leq y \leq F} \frac{Fp(F)}{yp(y) + (F - y)p(F)} \tag{26.67}$$

Recall that the upper bound on y is slack, and hence we can equivalently write

$$\frac{4}{3} = \max_{F \geq 0, a \geq 0, b \geq 0, y \geq 0} \frac{Fp(F)}{yp(y) + (F - y)p(F)} \tag{26.68}$$

Since the above inequality is true in general, we can substitute $\hat{y}_l$ for F and y_l^* for y to obtain

$$\frac{4}{3} \geq \frac{\hat{y}_l p_l(\hat{y}_l)}{y_l^* p_l(y_l^*) + (\hat{y}_l - y_l^*) p_l(\hat{y}_l)} \tag{26.69}$$

$$\Rightarrow \quad y_l^* p_l(y_l^*) \geq \frac{3}{4} \hat{y}_l p_l(\hat{y}_l) + (y_l^* - \hat{y}_l) p_l(\hat{y}_l) \tag{26.70}$$

Now, summing over all links we get

$$\sum_l y_l^* p_l(y_l^*) \geq \frac{3}{4} \sum_l \hat{y}_l p_l(\hat{y}_l) + \underbrace{\sum_l (y_l^* - \hat{y}_l) p_l(\hat{y}_l)}_{(A)} \tag{26.71}$$

If the term labelled (A) above is nonnegative, the theorem immediately follows. In order to show that this is indeed the case, we use the following argument. We have

$$\sum_l (y_l^* - \hat{y}_l) p_l(\hat{y}_l) \tag{26.72}$$

$$= \sum_l p_l(\hat{y}_l) \left(\sum_{l \in r} x_r^* - \sum_{l \in r} \hat{x}_r \right) \tag{26.73}$$

$$= \sum_r \left(\sum_{l \in r} p_l(\hat{y}_l) \right) (x_r^* - \hat{x}_r) \tag{26.74}$$

$$= \sum_r q_r(\hat{x})(x_r^* - \hat{x}_r) \tag{26.75}$$

$$\geq 0, \tag{26.76}$$

where the final inequality follows from (26.65). Hence, from (26.71)

$$PoA \leq \frac{4}{3}.$$

We may note in conclusion that the socially optimal solution, and WE both may have multiple solutions but have unique costs since they are both convex programs. □

4 Pricing Approach to Scheduling

In the last two sections, we considered two kinds of agent models in resource sharing games and characterized the efficiency loss incurred in each case. In Sect. 2, the focus was on a finite set of agents, with each having an impact on others' payoffs, whereas in Sect. 3, the number of agents was infinite, and each agent had an infinitesimal impact on the payoffs of others. In this section, we will consider a model in which the number of agents is infinitely large, but each agent only interacts with a finite subset of these agents at each time. Further, we now consider a repeated game in which an agent's state changes at each step according to a discrete time Markov process.

Our context is a system consisting of smartphone users whose apps are modeled as queues that come into existence when the user starts the app and vanish when the user terminates that app. The app generates packets (either uplink or downlink) that get buffered in the corresponding queue and which need to be served by a cellular base station. If a user is scheduled for service, the queue is decremented by a unit amount. We can associate a cost with the instantaneous queue length, which captures the idea that the quality of service of the app suffers with increasing packet delays. The user moves around, meaning that it changes the cell that it is present in. The shared resource across users in a particular cell is the spectrum allocated to that cell, and we assume for simplicity that this is such that only one queue can be served at any time. Then a question arises as to how to schedule queues in each cell in such a way that a good performance is obtained.

It is well known that the longest-queue-first (LQF) algorithm has attractive properties, such as minimizing the expected value of the longest queue in the system. One would expect that such an algorithm would ensure good performance across all apps. While the queues corresponding to downlink are present at the base station itself, the uplink queues are present at the user devices themselves. However, if the base station announces an LQF policy and then polls the queues seeking their lengths, there is clearly an incentive for each queue to try and receive additional service by misreporting its value. How should we design a system wherein the queues accurately reveal their information to the base stations?

Suppose that we hold an auction in which each queue placed a bid and the highest bidder gets service. The mechanism used can be chosen as a second-price scheme in which the winning bidder pays the value of the second highest bid. It is well known that such an auction promotes truth-telling about valuation in the single-step case. But in our system, the queue makes bids in each time step during its lifetime, and so we have a repeated game setting. Further, the queue has to estimate the likely bids made by the other queues at each step and choose its bid by trading off the value of winning (in terms of decrementing its queue length) and the payment to be made. What kind of bids would be seen in such as system? Would conducting such an auction repeatedly over time with queues arriving and departing result in some form of equilibrium? Would the scheduling decisions resulting from such auctions resemble that of LQF?

4.1 Mean Field Games

We will investigate the existence of an equilibrium using the theory of mean field games (MFGs). In this setup, the players assume that each opponent would play an action drawn *independently* from a fixed distribution over its action space. This distribution is called the *belief* of the players. The player then chooses an action that, given its own state, is the best response against actions drawn in this fashion. We say that the system is at Mean Field Equilibrium (MFE) if this best response action is itself a sample drawn from the assumed distribution. Thus, the player's action should be consistent with the assumed distribution.

The MFG framework offers a structure to approximate so-called Perfect Bayesian Equilibrium (PBE) in dynamic games. PBE requires each player to keep track of their beliefs on the future plays of every other opponent in the system and play the best response to that belief. This makes the computation of PBE intractable when the number of players is large. In our case, the state of each player is the current queue length, while the action is the bid made to the base station in which the phone is currently located. A PBE would require each app to estimate the queue lengths and the associated bids of every other queue that it is competing against – something that is clearly quite hard. The MFG approximation would assume that the bids made by the opponents are drawn independently from some bid distribution and to place a bid in response. The MFG framework simplifies computation of the best response and often turns out to be asymptotically optimal.

Analysis of a typical mean field game problem involves the following steps:

1. Characterization of the best response policy under a fixed belief. We view the system from the perspective of a single agent that takes an action at each time step based on its state, its belief about other agents' actions, and its knowledge about the transition probabilities of its state. In the repeated game case, this process involves characterizing the optimal action in a Markov decision process (MDP).
2. Showing the existence of the MFE using fixed point arguments. This step involves verifying that the map between the belief distribution and the invariant distribution of the mean field agent's actions has a fixed point. In general, the properties to be verified are continuity of the map and compactness of the range space.
3. Showing asymptotic independence of the queues. One of the assumptions in computing the MFE is that, when a given agent interacts with another agent, their queue length distributions are independent. The first step in justifying this assumption is to prove that the queue lengths of any finite set of queues are independent when the numbers of users goes to infinity. This independence result is proved for a class of policies which includes the policy obtained using the MFE assumption. Then, the result is extended to show that independence continues to hold in steady state.

4. Showing that the MFE is an ϵ-Nash equilibrium asymptotically. Here, the idea is that if the mean field policy were employed by all agents except a particular agent, that agent would gain no more than ϵ payoff by a unilateral deviation employing any other policy.
5. Dynamics of the many particle system. One hurdle to implementing the policy obtained from the MFE approach is that the equilibrium distribution of each user's queue length distribution is unknown. In principle, this is estimated by the stations from a histogram of the users' queue lengths. Indeed, simulations indicate that a policy computed from such an empirical estimate of the distribution converges to the MFE policy when the number of users is large. However, proving such a result is an open problem at this time.

In what follows, we will focus only on Steps 1 and 2 above and will provide full results on the first and a sketch of the second. In the problem that we consider, Step 4 and the independence result in Step 3 over finite time horizons follow in a straightforward fashion, while research into independence assumption in steady state in Step 3 and all of Step 5 is ongoing. We provide some references into these aspects at the end of the chapter.

4.2 Mean Field Auction Model

We consider a system consisting of N cells, each with a base station, and a total of NM agents. The agents are the smartphone apps, in which each one is associated with a queue. These agents are randomly permuted in the cells at each discrete time instant k, with each cell having exactly M agents. The model can be thought of as representing the idea that each cell in a cellular system typically has about 1000 devices, of which about 10 devices might be active at any time. The devices are mobile, which means that the pool from which active devices appear is constantly changing. Each cell contains a base station, which conducts a second-price auction to choose which agent to serve. Each agent must choose its bid in response to its state and its belief over the bids of its competitors.

Figure 26.7 illustrates the MFG approximation, which is accurate in the limit as N becomes large. As mentioned earlier, an MFG is described from the perspective of a single agent, which assumes that the actions of all its competitors are drawn independently from some distribution called the belief.

Auction system: The agent of interest competes in a second-price auction against $M - 1$ other agents, whose bids are assumed to be independently drawn from a continuous, finite mean (cumulative) bid distribution ρ, with support $\mathbb{R}^+$. The state of the agent is its current queue length q (the random variable is represented by Q). The queue length induces a holding cost $C(q)$, where $C(.)$ is a strictly convex and increasing function. Suppose that the agent bids an amount $w \in \mathbb{R}^+$. The outcomes of the auction are that the agent would obtain a unit of service with probability $p_\rho(w)$ and would have to pay an expected amount of $r_\rho(w)$ when all the other bids are drawn independently from ρ. Further, the queue has future job arrivals according

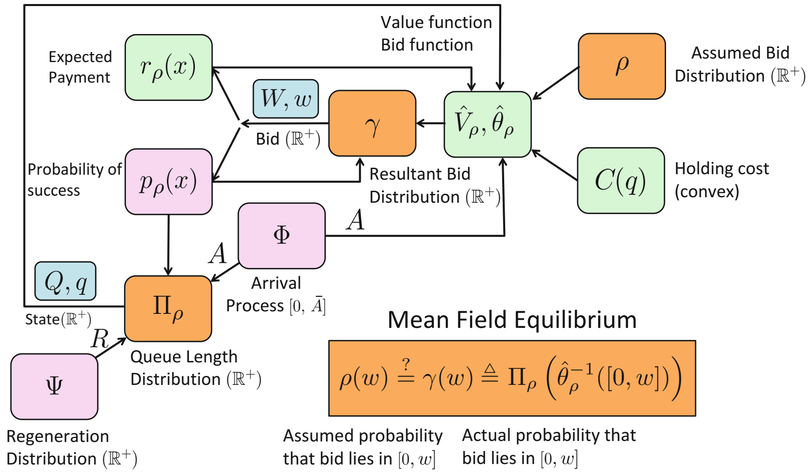


Fig. 26.7 The game consists of an auction part (inner loop) and a queue dynamics part (outer loop). The system is at MFE if the resultant bid distribution γ is the same as the assumed bid distribution ρ

to distribution Φ, with the random job size being denoted by A. Finally, the app can terminate at any time instant with probability $1-\beta$. Based on these inputs, the agent needs to determine the value of its current state $\hat{V}_\rho(q)$, and the best response bid to make $w = \hat{\theta}_\rho(q)$.

Queueing system: The queueing dynamics are driven by the arrival process Φ and the probability of obtaining service being $p_\rho(w)$ as described above. When the user terminates an app, he/she immediately starts a fresh app, i.e., a new queue takes the place of a departing queue. The initial condition of this new queue is drawn from a regeneration distribution Ψ, whose support is $\mathbb{R}^+$. The invariant distribution associated with this queueing system (if it exists) is denoted by Π_ρ.

Mean field equilibrium: The probability that the agent's bid (represented by the random variable W) lies in the interval $[0, w]$ is equal to the probability that the agent's queue length lies in some set whose best response is to bid between $[0, w]$. Thus, the probability of the bid lying in the interval $[0, w]$, denoted by the cumulative probability distribution $\gamma(w)$, is $\Pi_\rho(\hat{\theta}_\rho^{-1}([0, w]))$. According to the assumed (cumulative) bid distribution, the probability of the same event is $\rho(w)$. If $\rho(w) = \gamma(w)$, it means that the assumed bid distribution is consistent with the best response bid distribution, and we have an MFE.

4.2.1 Agent's Decision Problem

Under the mean field regime, we are interested in the decision and state evolution a particular agent i that has a belief that the bid of each other agent (opponent) has cumulative distribution ρ, independent of each other. We assume that $\rho \in \mathcal{P}$ where, $\mathcal{P}$ is the set of distributions with a continuous c.d.f. and a finite mean, upper bounded by some $E < \infty$. Suppose that the random variable representing the bid made by agent i at time k is denoted by $W_{i,k}$, with the realized value being w. Also, let

$$\bar{W}_{-i,k} = \max_{j \in M_{i,k}} W_{j,k},$$

represent the maximum value of $M - 1$ draws from the distribution ρ. Thus, $\bar{W}_{-i,k}$ is the value of the highest opposing bid.

Since the time of regeneration T_i^k is a geometric random variable, the expected cost of agent i can be can be written as

$$V_{i,\rho}(H_{i,k}; \theta_i) = \mathbb{E}\left[\sum_{t=k}^{\infty} \beta^t [C(Q_{i,t}) + r_\rho(W_{i,t})]\right], \tag{26.77}$$

where $H_{i,k}$ is the history observed by agent i until time k, θ_i is the bid function that it employs, and the expectation is over future state evolutions. Also,

$$r_\rho(w) = \mathbb{E}[\bar{W}_{-i,k}\mathbf{I}\{\bar{W}_{-i,k} \leq w\}]$$

is the expected payment when i bids w under the assumption that the bids of other agents are distributed according to ρ. Hence, given ρ, the probability that agent i wins in the auction is

$$p_\rho(w) = \mathbb{P}(\bar{W}_{-i,k} \leq w) = \rho(w)^{M-1}. \tag{26.78}$$

The expected payment when bidding w is

$$\begin{aligned} r_\rho(w) &= \mathbb{E}[\bar{W}_{-i,k}\mathbf{I}\{\bar{W}_{-i,k} \leq w\}] \\ &= w p_\rho(w) - \int_0^w p_\rho(u) du. \end{aligned} \tag{26.79}$$

The state process $Q_{i,k}$ is Markov and has a transition kernel

$$\begin{aligned} \mathbb{P}(Q_{i,k+1} \in B | Q_{i,k} = q, W_{i,k} = w) &= \beta p_\rho(w)\mathbb{P}((q-1)^+ + A_k \in B) \\ &+ \beta(1 - p_\rho(w))\mathbb{P}(q + A_k \in B) + (1-\beta)\Psi(B), \end{aligned} \tag{26.80}$$

where $B \subseteq \mathbb{R}^+$ is a Borel set and $x^+ \triangleq \max(x, 0)$. Recall that $A_k \sim \Phi$ is the arrival between $(k)^{th}$ and $(k+1)^{th}$ auction and Ψ is density function of the regeneration process. In the above expression, the first term corresponds to the event that agent wins the auction at time k, while the second corresponds to the event that it does not. The last term captures the event that the agent regenerates after auction k. The agent's decision problem can be modeled as an infinite horizon discounted cost MDP. Standard results can be used to show that there exists an optimal Markov deterministic policy for our MDP; (Strauch 1966). Then, from (26.77), the optimal value function of the agent can be written as

$$\hat{V}_{i,\rho}(q) = \inf_{\theta_i \in \Theta} \mathbb{E}\left[\sum_{t=1}^{\infty} \beta^t [C(Q_{i,t}) + r_\rho(W_{i,t})] \,|\, Q_{i,0} = q\right], \tag{26.81}$$

where Θ is the space of Markov deterministic policies. Once we have the above formulation, the index of the agent is redundant as we are concerned with a single agent's decision problem. Hence, we will omit the agent subscript i in what follows.

4.2.2 Existence of a Stationary Distribution

Given cumulative bid distribution ρ and a Markov policy $\theta \in \Theta$, the transition kernel given by (26.80) can be rewritten as,

$$\begin{aligned}\mathbb{P}(Q_{k+1} \in B | Q_k = q) &= \beta p_\rho(\theta(q))\mathbb{P}((q-1)^+ + A_k \in B) \\ &\quad + \beta(1 - p_\rho(\theta(q)))\mathbb{P}(q + A_k \in B) + (1-\beta)\Psi(B).\end{aligned} \tag{26.82}$$

A basic question is whether a stationary distribution Π_ρ exists under an arbitrary Markov policy θ. This is critical if we are to characterize the map between the assumed bid distribution and ρ and the resultant bid distribution γ. It turns out that under our formulation, the existence of the invariant state distribution follows immediately from Meyn et al. (2009).

4.2.3 Mean Field Equilibrium

The mean field equilibrium is essentially a consistency check that the bid distribution γ induced by the stationary distribution Π_{ρ,θ_ρ} is identical to the bid distribution that forms the belief of the agent, i.e., ρ. Hence, we have the following definition of MFE:

Definition 3 (Mean field equilibrium). Let ρ be a bid distribution and θ_ρ be a stationary policy for an agent. Then, we say that (ρ, θ_ρ) constitutes a mean field equilibrium if

1. θ_ρ is an optimal policy of the decision problem in (26.81), given bid distribution ρ.
2. $\rho(x) = \gamma(w) \triangleq \Pi_\rho(\theta_\rho^{-1}([0, w])), \forall w \in \mathbb{R}^+$, where $\Pi_\rho = \Pi_{\rho,\theta_\rho}$.

We now characterize the best response policy and describe the steps involved in proving existence of the MFE.

4.3 Properties of Optimal Bid Function

The decision problem given by (26.81) is an infinite horizon, discounted Markov decision problem (MDP). The optimality equation or Bellman equation corresponding to the decision problem is

$$\hat{V}_\rho(q) = C(q) + \beta \mathbb{E}_A(\hat{V}_\rho(q+A)) + \inf_{x \in R^+} \left[r_\rho(w) - p_\rho(w)\beta \mathbb{E}_A \left(\hat{V}_\rho(q+A) - \hat{V}\rho((q-1)^+ + A) \right) \right], \tag{26.83}$$

where $A \sim \Phi$, and we use the notation $\max(0, z) = z^+$.

We define the set of functions

$$\mathcal{V} = \left\{ f : \mathbb{R}^+ \mapsto \mathbb{R}^+ : \sup_{q \in \mathbb{R}^+} \left| \frac{f(q)}{\mu(q)} \right| < \infty \right\}, \tag{26.84}$$

where $\mu(q) = \max\{C(q), 1\}$. Clearly, $\mathcal{V}$ is a Banach space with μ-*norm*,

$$\|f\|_\mu = \sup_{q \in \mathbb{R}^+} \left| \frac{f(q)}{\mu(q)} \right| < \infty.$$

We define the Bellman operator T_ρ as

$$(T_\rho f)(q) = C(q) + \beta \mathbb{E}_A f(q+A) + \inf_{w \in \mathbb{R}^+} \left[r_\rho(w) - p_\rho(w)\beta(\mathbb{E}_A(f(q+A) - f((q-1)^+ + A))) \right], \tag{26.85}$$

where $f \in \mathcal{V}$. It is straightforward to show that the infimum in the above operator occurs at

$$\beta \Delta f(q)^+, \tag{26.86}$$

where $\Delta f(q) = \mathbb{E}_A(f(q+A) - f((q-1)^+ + A))$. Then, substituting from (26.78), (26.79) and (26.86), (26.85) can be rewritten as

$$(T_\rho f)(q) = C(q) + \beta \mathbb{E}_A f(q+A) - \int_0^{\beta \Delta f(q)^+} p_\rho(u) du. \tag{26.87}$$

Our first step is to show that an optimal solution exists for this problem. The MDP is in discrete time, but state consists of all nonnegative real numbers. There exist standard regularity conditions under which such an MDP has a solution. For instance, our problem setup can be posed as a slightly modified version of that in Theorem 8.3.6 of Hernández-Lerma and de Ozak (1992). The result is as follows:

Lemma 1. *Given a cumulative bid distribution* ρ,

1. *There exists a* $j \in \mathbb{N}$ *such that* $T_\rho^j : \mathcal{V} \to \mathcal{V}$ *is a contraction mapping. Hence, there exists a unique* $f_\rho^* \in \mathcal{V}$ *such that* $T_\rho f_\rho^* = f_\rho^*$*, and for any* $f \in \mathcal{V}$, $T_\rho^n f \to f_\rho^*$ *as* $n \to \infty$.
2. *The fixed point* f_ρ^* *of operator* T_ρ *is the unique solution to the optimality Eq.* (26.83)*, i.e.,* $f_\rho^* = \hat{V}_\rho$.
3. *Letting* $\hat{\theta}_\rho(q) = \beta \Delta \hat{V}_\rho(q)^+$, $\hat{\theta}_\rho$ *is an optimal policy.*

Corollary 1. *An optimal policy of the agent's decision problem* (26.81) *is given by*

$$\hat{\theta}_\rho(q) = \beta \mathbb{E}_A \left[\hat{V}_\rho(q + A) - \hat{V}_\rho((q-1)^+ + A) \right].$$

We now establish that $\hat{V}_\rho$ and $\hat{\theta}_\rho$ are continuous and increasing functions.

Lemma 2. *Given a cumulative bid distribution function* ρ

1. $\hat{V}_\rho$ *is a continuous increasing function.*
2. $\hat{\theta}_\rho$ *is a continuous strictly increasing function.*

Proof. Let $f \in \mathcal{V}$. Suppose f is a continuous monotone increasing function. We first prove that $T_\rho f$ is also continuous monotone increasing function. Since, the $n - step$ Bellman operator $T_\rho^n f \to \hat{V}_\rho$ according to Statement 2 of Lemma 1, we conclude that $\hat{V}_\rho$ also has the same property.

Let $q > q'$. Then,

$$\begin{aligned}
T_\rho f(q) - T_\rho f(q') &= C(q) - C(q') + \beta \mathbb{E}_A (f(q + A) - f(q' + A)) \\
&\quad + \inf_w [r_\rho(w) - \beta p_\rho(w) \mathbb{E}_A (f(q + A) - f((q-1)^+ + A))] \\
&\quad - \inf_w [r_\rho(w) - \beta p_\rho(w) \mathbb{E}_A (f(q' + A) - f((q'-1)^+ + A))] \\
&\overset{(a)}{\geq} \beta \mathbb{E}_A (f(q + A) - f(q' + A)) + \beta \inf_w \left[p_\rho(w) \mathbb{E}_A (f(q' + A) \right. \\
&\quad \left. - f((q'-1)^+ + A) - f(q + A) + f((q-1)^+ + A)) \right] \\
&\geq \beta \min \left\{ \mathbb{E}_A (f(q + A) - f(q' + A)), \mathbb{E}_A (f((q-1)^+ + A) \right. \\
&\quad \left. - f((q'-1)^+ + A)) \right\} \\
&\overset{(b)}{\geq} 0,
\end{aligned}$$

where (a) follows from the assumption that $C(.)$ is an increasing function, and (b) follows from the assumption that $f(.)$ is an increasing function.

To prove that $T_\rho f$ is continuous consider a sequence $\{q_n\}$ such that $q_n \to q$. Since f is a continuous function, $f(q_n+a) \to f(q+a)$. Then, by using dominated convergence theorem, we have $\mathbb{E}_A f(q_n+A) \to \mathbb{E}_A f(q+A)$ and $\mathbb{E}_A f((q_n-1)^+ + A) \to \mathbb{E}_A f((q-1)^+ + A)$. Also, $\Delta f(q_n) \geq 0$ as f is an increasing function. Then, from (26.87), we get that

$$T_\rho f(q_n) = C(q_n) + \beta\mathbb{E}_A f(q_n + A) - \int_0^{\beta\Delta f(q_n)} p_\rho(u)du$$

$$\to C(q) + \beta\mathbb{E}_A f(q + A) - \int_0^{\beta\Delta f(q)} p_\rho(u)du = T_\rho f(q).$$

Hence, $T_\rho f$ is a continuous function. This yields Statement 1 in the lemma.

Now, to prove the second part, assume that Δf is an increasing function. First, we show that $\Delta T_\rho f$ is an increasing function. Let $q > q'$. From (26.87), for any $a < \bar{A}$ we can write

$$\begin{aligned}
(T_\rho f)(q+a) &- (T_\rho f)((q-1)^+ + a) - (T_\rho f)(q'+a) + (T_\rho f)((q'-1)^+ + a) \\
&= C(q+a) - C((q-1)^+ + a) - C(q'+a) + C((q'-1)^+ + a) \\
&+ \beta\mathbb{E}_A f(q+a+A) - \beta\mathbb{E}_A f((q-1)^+ + a + A) \\
&- \beta\mathbb{E}_A f(q'+a+A) + \beta\mathbb{E}_A f((q'-1)^+ + a + A) \\
&- \int_{\beta\Delta f(q'+a)}^{\beta\Delta f(q+a)} p_\rho(u)\,du + \int_{\beta\Delta f((q'-1)^+ + a)}^{\beta\Delta f((q-1)^+ + a)} p_\rho(u)\,du \\
&= C(q+a) - C((q-1)^+ + a) \\
&- C(q'+a) + C((q'-1)^+ + a) \\
&+ \beta\mathbb{E}_A f((q+a-1)^+ + A) - \beta\mathbb{E}_A f((q-1)^+ + a + A) \\
&- \beta\mathbb{E}_A f((q'+a-1)^+ + A) + \beta\mathbb{E}_A f((q'-1)^+ + a + A) \\
&+ \int_{\beta\Delta f(q'+a)}^{\beta\Delta f(q+a)} 1 - p_\rho(u)\,du + \int_{\beta\Delta f((q'-1)^+ + a)}^{\beta\Delta f((q-1)^+ + a)} p_\rho(u)\,du
\end{aligned}$$

It can be easily verified that

$$\mathbb{E}_A(f(q+a-1)^+ + A) - \mathbb{E}_A(f(q-1)^+ + a + A) - \mathbb{E}_A(f(q'+a-1)^+ + A) + \mathbb{E}_A(f(q'-1)^+ + a + A) \geq 0,$$

as f is increasing (due to Statement 1 of this lemma). From the assumption that Δf is increasing, the last two terms in the above expression are also nonnegative. Now, taking expectation on both sides, we obtain $\Delta T_\rho f(q) - \Delta T_\rho f(q') \geq \Delta C(q) - \Delta C(q') > 0$. Therefore, from Statements 2 and 3 of Lemma 1, we have

$$\hat{\theta}_\rho(q) - \hat{\theta}_\rho(q') = \Delta \hat{V}_\rho(q) - \Delta \hat{V}_\rho(q') \geq \Delta C(q) - \Delta C(q') > 0.$$

Here, the last inequality holds since C is a strictly convex increasing function.□

4.4 Existence of MFE

We now describe the steps involved in showing the existence of the MFE. In many cases we will only provide proof sketches to show how the argument proceeds.

Theorem 3. *There exists an MFE* $\left(\rho, \hat{\theta}_\rho\right)$ *such that*

$$\rho(x) = \gamma(x) \triangleq \Pi_\rho\left(\hat{\theta}_\rho^{-1}[0, x]\right), \forall x \in \mathbb{R}^+.$$

We first introduce some useful notation. Let $\Theta = \{\theta : \mathbb{R} \mapsto \mathbb{R}, \sup_{q \in \mathbb{R}^+} \left|\frac{\theta(q)}{w(q)}\right| < \infty\}$. Note that Θ is a normed space with w-norm. Also, let Ω be the space of absolutely continuous probability measures on $\mathbb{R}^+$. We endow this probability space with the topology of weak convergence.

We define $\theta^* : \mathcal{P} \mapsto \Theta$ as $(\theta^*(\rho))(q) = \hat{\theta}_\rho(q)$, where $\hat{\theta}_\rho(q)$ is the optimal bid given by Corollary 1. It can be easily verified that $\hat{\theta}_\rho \in \Theta$. Also, define the mapping Π^* that takes a bid distribution ρ to the invariant workload distribution $\Pi_\rho(\cdot)$. Later, using Lemma 3 we will show that $\Pi_\rho(\cdot) \in \Omega$. Therefore, $\Pi^* : \mathcal{P} \to \Omega$. Finally, define $\mathcal{F}$ as $(\mathcal{F}(\rho))(x) = \gamma(x) = \Pi_\rho(\hat{\theta}_\rho^{-1}([0, x]))$. Lemma 5 will show that $\mathcal{F}$ maps $\mathcal{P}$ into itself.

In order to prove the above theorem, we need to show that $\mathcal{F}$ has a fixed point, i.e., $\mathcal{F}(\rho) = \rho$.

Theorem 4 (Schauder Fixed Point Theorem). *Suppose $\mathcal{F}(\mathcal{P}) \subset \mathcal{P}$. If $\mathcal{F}(\cdot)$ is continuous and $\mathcal{F}(\mathcal{P})$ is contained in a convex and compact subset of $\mathcal{P}$, then $\mathcal{F}(\cdot)$ has a fixed point.*

We will show that the mapping $\mathcal{F}$ satisfies the conditions of the above theorem, and hence it has a fixed point. Note that $\mathcal{P}$ is a convex set. Therefore, we only need to verify the other two conditions.

To prove the continuity of mapping $\mathcal{F}$, we first show that θ^* and Π^* are continuous mappings. To that end, we will show that for any sequence $\rho_n \to \rho$ in uniform norm, we have $\theta^*(\rho_n) \to \theta^*(\rho)$ in w-norm and $\Pi^*(\rho_n) \Rightarrow \Pi^*(\rho)$

(where $\Rightarrow$ denotes weak convergence). Finally, we use the continuity of θ^* and Π^* to prove that $\mathcal{F}(\rho_n) \to \mathcal{F}(\rho)$.

Step 1: Continuity of the map θ^*

Theorem 5. *The map θ^* is continuous.*

Proof. Define the map $V^* : \mathcal{P} \mapsto \mathcal{V}$ that takes ρ to $\hat{V}_\rho(\cdot)$. We begin by showing that $\|\hat{\theta}_{\rho_1} - \hat{\theta}_{\rho_2}\|_\mu \le K\|\hat{V}_{\rho_1} - \hat{V}_{\rho_2}\|_\mu$, which means that the continuity of the map V^* implies the continuity of the map θ^*. Next, we show two simple properties of the Bellman operator. The first is that for any $\rho \in \mathcal{P}$ and $f_1,\ f_2 \in \mathcal{V}$,

$$\|T_\rho f_1 - T_\rho f_2\|_\mu \le \hat{K}\|f_1 - f_2\|_\mu \tag{26.88}$$

for some large $\hat{K}$, independent of ρ.

Second, let T_{ρ_1} and T_{ρ_2} be the Bellman operators corresponding to $\rho_1,\ \rho_2 \in \mathcal{P}$ and let $f \in \mathcal{V}$. We show that

$$\|T_{\rho_1} f - T_{\rho_2} f\|_\mu \le 2(M-1)K_1\|f\|_\mu\|\rho_1 - \rho_2\|. \tag{26.89}$$

We then have

$$\|T_{\rho_1}^j \hat{V}_{\rho_2} - T_{\rho_2}^j \hat{V}_{\rho_2}\|_\mu \le \tag{26.90}$$

$$\begin{aligned}
\|T_{\rho_1}^j \hat{V}_{\rho_2} - T_{\rho_1}^{j-1} T_{\rho_2} \hat{V}_{\rho_2}\|_\mu & \\
&+ \|T_{\rho_1}^{j-1} T_{\rho_2} \hat{V}_{\rho_2} - T_{\rho_1}^{j-2} T_{\rho_2}^2 \hat{V}_{\rho_2}\|_\mu + \cdots \\
&+ \|T_{\rho_1} T_{\rho_2}^{j-1} \hat{V}_{\rho_2} - T_{\rho_2}^j \hat{V}_{\rho_2}\|_\mu \\
&\le \hat{K}^{j-1}\|T_{\rho_1}\hat{V}_{\rho_2} - T_{\rho_2}\hat{V}_{\rho_2}\|_\mu + \cdots \\
&+ \|T_{\rho_1} T_{\rho_2}^{j-1} \hat{V}_{\rho_2} - T_{\rho_2}^j \hat{V}_{\rho_2}\|_\mu \qquad (26.91) \\
&\le (\hat{K}^{j-1} + \cdots + 1)\|T_{\rho_1}\hat{V}_{\rho_2} - T_{\rho_2}\hat{V}_{\rho_2}\|_\mu \\
&\le 2(M-1)K\|\rho_1 - \rho_2\|(\hat{K}^{j-1} + \cdots + 1)\|\hat{V}_{\rho_2}\|_\mu \qquad (26.92)
\end{aligned}$$

Here, (26.91) and (26.92) follow from (26.88) and (26.89), respectively.

Now, let j be such that $T_{\rho_1}^j$ is an α-contraction, which is guaranteed to exist by Lemma 1. Note that Statement 1 of Lemma 1 implies that such a $j < \infty$ exists. Then we have

$$\|\hat{V}_{\rho_1} - \hat{V}_{\rho_2}\|_\mu = \|T_{\rho_1}^j \hat{V}_{\rho_1} - T_{\rho_2}^j \hat{V}_{\rho_2}\|_\mu$$

$$\le \|T^j_{\rho_1}\hat{V}_{\rho_1} - T^j_{\rho_1}\hat{V}_{\rho_2}\|_\mu + \|T^j_{\rho_1}\hat{V}_{\rho_2} - T^j_{\rho_2}\hat{V}_{\rho_2}\|_\mu$$

$$\Longrightarrow (1-\alpha)\|\hat{V}_{\rho_1} - \hat{V}_{\rho_2}\|_\mu \le \|T^j_{\rho_1}\hat{V}_{\rho_2} - T^j_{\rho_2}\hat{V}_{\rho_2}\|_\mu \tag{26.93}$$

Finally, from (26.92) and (26.93), we get

$$\begin{aligned}&\|\hat{V}_{\rho_1} - \hat{V}_{\rho_2}\|_\mu \\ &\le \frac{2(m-1)K(\hat{K}^{j-1} + \cdots + 1)\|\rho_1 - \rho_2\|}{1-\alpha}\|\hat{V}_{\rho_2}\|_\mu \\ &\le \frac{2(m-1)K(\hat{K}^{j-1} + \cdots + 1)\|\rho_1 - \rho_2\|}{1-\alpha} \\ &\quad \times (\|\hat{V}_{\rho_1}\|_\mu + \|\hat{V}_{\rho_1} - \hat{V}_{\rho_2}\|_\mu).\end{aligned}$$

Therefore, if $\frac{2(m-1)K(\hat{K}^{j-1}+\cdots+1)}{1-\alpha}\|\rho_1 - \rho_2\| < \frac{1}{2}$, then

$$\begin{aligned}&\|\hat{V}_{\rho_1} - \hat{V}_{\rho_2}\|_\mu \\ &\le \frac{4(m-1)K(\hat{K}^{j-1} + \cdots + 1)}{1-\alpha}\|\hat{V}_{\rho_1}\|_\mu\|\rho_1 - \rho_2\|\end{aligned}$$

Hence, the maps V^* and θ^* are continuous. □

Step 2: Continuity of the map Π^*

Let $\Pi_{\rho,\theta}(.)$ be the invariant distribution generated by any θ. Recall that Π^* takes $\rho \in \mathcal{P}$ to probability measure $\Pi_\rho(.) = \Pi_{\rho,\hat{\theta}_\rho}(.)$. First, we show that $\Pi_{\rho,\theta}(.) \in \Omega$, where Ω is the space of absolutely continuous measures (with respect to Lebesgue measure) on $\mathbb{R}^+$.

Lemma 3. *For any $\rho \in \mathcal{P}$ and any $\theta \in \Theta$, $\Pi_{\rho,\theta}(\cdot)$ is absolutely continuous with respect to the Lebesgue measure on $\mathbb{R}^+$.*

Proof. $\Pi_{\rho,\theta}(\cdot)$ can be expressed as the invariant queue-length distribution of the dynamics

$$q \to \begin{cases} Q' + A & \text{with probability } \beta \\ R & \text{with probability } (1-\beta), \end{cases}$$

where $A \sim \Phi$ and $R \sim \Psi$, and Q' is a random variable with distribution generated by the conditional probabilities

$$\mathbb{P}(Q' = q|q) = 1 - p_\rho(\hat{\theta}(q))$$

$$\mathbb{P}(Q' = (q-1)^+|q) = p_\rho(\hat{\theta}(q))$$

Let Π' be the distribution of Q'. Then for any Borel set B, Π can be expressed using the convolution of Π' and Φ :

$$\Pi_{\rho,\theta}(B) = \beta \int_{-\infty}^{\infty} \Phi(B-y)d\Pi'(y) + (1-\beta)\Psi(B). \tag{26.94}$$

If B is a Lebesgue null set, then so is $B-y \;\; \forall y$. So, $\Phi(B-y) = 0$ and $\Psi(B) = 0$ and therefore $\Pi_\rho(B) = 0$. □

We now develop a useful characterization of $\Pi_{\rho,\theta}$. Let

$$\Upsilon_{\rho,\theta}^{(k)}(B|q) = \mathbb{P}(Q_k \in B|\text{no regeneration, } Q_0 = q)$$

be the distribution of queue length Q_k at time k induced by the transition probabilities given in (26.82) conditioned on the event that $Q_0 = q$ and that there are no regenerations until time k. We can now express the invariant distribution $\Pi_{\rho,\theta}(\cdot)$ in terms of $\Upsilon_{\rho,\theta}^{(k)}(\cdot|q)$ as in the following lemma.

Lemma 4. *For any bid distribution $\rho \in \mathcal{P}$ and for any stationary policy $\theta \in \Theta$, the Markov chain described by the transition probabilities given in (26.82) has a unique invariant distribution $\Pi_{\rho,\theta}(\cdot)$. Also $\Pi_{\rho,\theta}$ and $\Upsilon_{\rho,\theta}^{(k)}$ are related as follows:*

$$\Pi_{\rho,\theta}(B) = \sum_{k\geq 0}(1-\beta)\beta^k \mathbb{E}_\Psi(\Upsilon_{\rho,\theta}^{(k)}(B|Q)), \tag{26.95}$$

where $\mathbb{E}_\Psi(\Upsilon_{\rho,\theta}^{(k)}(B|Q)) = \int \Upsilon_{\rho,\theta}^{(k)}(B|q)d\Psi(q)$.

Proof. $\Upsilon_{\rho,\theta}^{(k)}(B|q)$ is the queue length distribution assuming no regeneration has happened yet, and the regeneration event occurs with probability β independently of the rest of the system. It is then easy to find $\Pi_{\rho,\theta}(B)$ in terms of $\Upsilon_{\rho,\theta}^{(k)}(B|q)$ by simply using the properties of the conditional expectation, and the theorem follows. Note that in $\mathbb{E}_\Psi(\Upsilon_{\rho,\theta}^{(k)}(B|Q))$, the random variable is the initial condition of the queue, as generated by Ψ. □

We next prove the continuity of Π^* in ρ.

Theorem 6. *The mapping $\Pi^* : \mathcal{P} \mapsto \Omega$ is continuous.*

Proof. By Portmanteau theorem, (Billingsley 2009), we only need to show that for any sequence $\rho_n \to \rho$ in w-norm and any open set B, $\liminf_{n\to\infty} \Pi_{\rho_n}(B) \geq \Pi_\rho(B)$. By Fatou's lemma,

$$\begin{aligned}\liminf_{n\to\infty} \Pi_{\rho_n}(B) \\ &= \liminf_{n\to\infty} \sum_{k=0}^{\infty}(1-\beta)\beta^k \mathbb{E}_{\Psi_R}[\Upsilon_{\rho_n}^{(k)}(B|Q)] \\ &\geq \sum_{k=0}^{\infty}(1-\beta)\beta^k \mathbb{E}_{\Psi_R}[\liminf_{n\to\infty} \Upsilon_{\rho_n}^{(k)}(B|Q)] \end{aligned} \tag{26.96}$$

where $Q \sim \Psi_R$. Let $\Upsilon_{\rho}^{(k)} = \Upsilon_{\rho,\hat{\theta}_\rho}^{(k)}$. We finally show that $\liminf_{n\to\infty} \Upsilon_{\rho_n}^{(k)}(B|q) \geq \Upsilon_{\rho}^{(k)}(B|q)$ for every $q \in \mathbb{R}^+$, and the proof follows. □

Step 3: Continuity of the mapping $\mathcal{F}$ Now, using the results from Step 1 and Step 2, we establish continuity of the mapping $\mathcal{F}$. First, we show that $\mathcal{F}(\rho) \in \mathcal{P}$.

Lemma 5. *For any $\rho \in \mathcal{P}$, let $\gamma(w) = (\mathcal{F}(\rho))(w) = \Pi_\rho(\hat{\theta}_\rho^{-1}([0,w])), w \in \mathbb{R}^+$. Then, $\gamma \in \mathcal{P}$.*

Proof. From the definition of Π_ρ, it is easy to see that γ is a distribution function. Since $\hat{\theta}_\rho$ is continuous and strictly increasing function as shown in Lemma 2, $\hat{\theta}_\rho^{-1}(\{w\})$ is either empty or a singleton. Then, from Lemma 3, we get that $\Pi_\rho(\hat{\theta}_\rho^{-1}(\{w\})) = 0$. Together, we get that $\gamma(w)$ has no jumps at any w and hence it is continuous.

To complete the proof, we need to show that the expected bid under $\gamma(.)$ is finite. In order to do this, we construct a new random process $\tilde{Q}_k$ that is identical to the original queue length dynamics Q_k, except that it never receives any service. We show that this process stochastically dominates the original, and use this property to bound the mean of the original process by a finite quantity independent of ρ. □

We now have the main theorem.

Theorem 7. *The mapping $\mathcal{F} : \mathcal{P} \mapsto \mathcal{P}$ given by $(\mathcal{F}(\rho))(w) = \Pi_\rho(\hat{\theta}_\rho^{-1}([0,w]))$ is continuous.*

Proof. Let $\rho_n \to \rho$ in uniform norm. From previous steps, we have $\hat{\theta}_{\rho_n} \to \hat{\theta}_\rho$ in μ-norm and $\Pi_{\rho_n} \Rightarrow \Pi_\rho$. Then, using Theorem 5.5 of Billingsley (2009), one can show that the push forwards also converge:

$$\Pi_{\rho_n}(\hat{\theta}_{\rho_n}^{-1}(\cdot)) \Rightarrow \Pi_\rho(\hat{\theta}_\rho^{-1}(\cdot)).$$

Then, $\mathcal{F}(\rho_n)$ converges point-wise to $\mathcal{F}(\rho)$ as it is continuous at every w, i.e., $(\mathcal{F}(\rho_n))(w) \to (\mathcal{F}(\rho))(w)$ for all $w \in \mathbb{R}^+$.

Finally, it is easy to show that in the norm space $\mathcal{P}$, point-wise convergence implies convergence in uniform norm, which completes the proof. □

Step 4: $\mathcal{F}(\mathcal{P})$ is contained in a compact subset of $\mathcal{P}$ We show that the closure of the image of the mapping $\mathcal{F}$, denoted by $\overline{\mathcal{F}(\mathcal{P})}$, is compact in $\mathcal{P}$. As $\mathcal{P}$ is a normed space, sequential compactness of any subset of $\mathcal{P}$ implies that the subset is compact. Hence, we just need to show that $\overline{\mathcal{F}(\mathcal{P})}$ is sequentially compact. Sequential compactness of a set $\overline{\mathcal{F}(\mathcal{P})}$ means the following: if $\{\rho_n\} \in \overline{\mathcal{F}(\mathcal{P})}$ is a sequence, then there exists a subsequence $\{\rho_{n_j}\}$ and $\rho \in \overline{\mathcal{F}(\mathcal{P})}$ such that $\rho_{n_j} \to \rho$. We use Arzelà-Ascoli theorem and uniform tightness of the measures in $\mathcal{F}(\mathcal{P})$ to show the sequential compactness. The version that we will use is stated below:

Theorem 8 (Arzelà-Ascoli Theorem). *Let X be a σ-compact metric space. Let $\mathcal{G}$ be a family of continuous real valued functions on X. Then the following two statements are equivalent:*

1. *For every sequence $\{g_n\} \subset \mathcal{G}$ there exists a subsequence g_{n_j} which converges uniformly on every compact subset of X.*
2. *The family $\mathcal{G}$ is equicontinuous on every compact subset of X, and for any $x \in X$, there is a constant C_x such that $|g(x)| < C_x$ for all $g \in \mathcal{G}$.*

Suppose a family of functions $\mathcal{D} \subseteq \mathcal{P}$ satisfies the equivalent conditions of the Arzelá-Ascoli theorem and in addition satisfy the uniform tightness property, i.e., $\forall \epsilon > 0$, there exists an x_ϵ such that for all $f \in \mathcal{D}$ $1 \geq f(x_\epsilon) \geq 1-\epsilon$. Then, for any sequence $\{\rho_n\} \subset \mathcal{D}$, there exists a subsequence $\{\rho_{n_j}\}$ that converges uniformly on every compact set to a continuous increasing function ρ on $\mathbb{R}^+$. As $\mathcal{D}$ is uniformly tight it can be shown that ρ_{n_j} converges uniformly to ρ and that $\rho \in \mathcal{P}$. Therefore, $\overline{\mathcal{D}}$ is sequentially compact in the topology of uniform norm.

In the following, we show that $\mathcal{F}(\mathcal{P})$ satisfies uniform tightness property and condition 2 in Arzelá-Ascoli theorem. First verifying the conditions of Arzelá-Ascoli theorem, note that the functions in consideration are uniformly bounded by 1. To prove equicontinuity, consider a $\gamma = \mathcal{F}(\rho)$ and let $x > y$.

$$\begin{aligned}\gamma(x) - \gamma(y) &= \Pi_\rho(\theta_\rho(q) \leq x) - \Pi_\rho(\theta_\rho(q) \leq y) \\ &= \Pi_\rho(y < \theta_\rho(q) \leq x) \end{aligned} \tag{26.97}$$

Lemma 6. *For any interval $[a,b]$, $\Pi_\rho([a,b]) < c \cdot (b-a)$, for some large enough c.*

Proof. The proof follows easily from our characterization of Π_ρ in terms of $\Upsilon_\rho^{(k)}$.□

The above lemma and Eq. (26.97) imply that $\gamma(x)-\gamma(y) \leq c(\theta_\rho^{-1}(x)-\theta_\rho^{-1}(y))$. To show equicontinuity, it is enough to show that $\limsup_{y\uparrow x} \frac{\gamma(x)-\gamma(y)}{x-y} \leq K(x)$ for some K independent of ρ. This property follows from our characterization of the optimal bid function.

Finally, we have the following lemma showing that $\mathcal{F}(\mathcal{P})$ is uniformly tight.

Lemma 7. *$\mathcal{F}(\mathcal{P})$ is uniformly tight, i.e., for any $\epsilon > 0$ and any $f \in \mathcal{F}(\mathcal{P})$, there exists an $x_\epsilon \in \mathbb{R}$ such that $1 - \epsilon \leq f(x_\epsilon) \leq 1$.*

Proof. From Lemma 5, we have $\mathcal{F}(\mathcal{P}) \subseteq \mathcal{P}$. Hence, the expectation of the bid distributions in $\mathcal{F}(\mathcal{P})$ is bounded uniformly. An application of Markov inequality will give uniform tightness. □

4.5 Properties of MFE

As we showed above, the bid function $\hat{\theta}_\rho(q)$ is monotone increasing in q regardless of ρ. This property implies that the service regime corresponding to MFE is identical to the LQF policy. The result essentially says that there is no price of anarchy induced by the auction-based scheduling policy! In other words, the desirable properties of LQF are a natural result of auction-based scheduling.

5 Notes

The VCG mechanism was developed by Clarke (1971) and Groves (1973) as a generalization of an auction called the second-price auction due to Vickrey (1961). The Kelly mechanism is due to Kelly (1997). The price of anarchy for strategic users using the Kelly mechanism was computed by Johari and Tsitsiklis in (2004). The interest in the Kelly mechanism is due to the fact that it has a simple decentralized implementation when the users are price taking. If one is more generally interested in truth-revealing mechanisms using one-dimensional bids, then there is recent work on the design of such mechanisms: the interested reader is referred to the works of Maheswaran and Basar (2006), Yang and Hajek (2007) and Johari and Tsitsiklis (2005).

The general class of selfish routing with infinitesimal agents (called non-atomic) was first discussed by Pigou (1920). The Baress' paradox was discovered by Braess (1968). The results on selfish routing discussed in this chapter is a simplified version of the work of Roughgarden and Tardos (2002). They also showed in the same work that the result can be generalized to networks with affine cost functions, with a price of anarchy of at most $4/3$. Our development follows the presentation in Roughgarden (2016). More recent developments in this area and further generalizations can be found in Roughgarden (2015) and Roughgarden and Schoppmann (2015).

The topic of "mean field games" has been covered in general terms in this Handbook in a chapter by that name, where detailed information on historical developments can be found. More relevant to the topic of this chapter, the MFG approach has recently been used in several different problems on games with a large number of agents, each subset of which meet infrequently. Examples include Tembine et al. (2009), Borkar and Sundaresan (2012), Xu and Hajek (2012), Adlakha and Johari (2013), Iyer et al. (2014), Manjrekar et al. (2014), and Li et al. (2015, 2017). The framework lends itself readily to the modeling an analysis of many realistic systems. For example, Iyer et al. (2014) consider advertisers competing via a second-price auction for spots on a webpage. The bid must lie in a finite real interval, and the winner can place an ad on the webpage. In the space of queueing systems, Xu and Hajek (2012) consider the game of sampling a number of queues and joining one. The mean field results on scheduling presented in this chapter are based primarily on work by Manjrekar et al. (2014). The idea of infrequent interactions between subsets of players is exploited in a recent application of the mean field game framework for mechanism design to incentivize truth-telling about one's ability to help peer devices in a device-to-device (D2D) network setting by Li et al. (2017). Another application is on designing "nudge systems" for modifying societal behavior through providing incentives such as lottery tickets. An application of this kind in the context of electricity networks is studied by Li et al. (2015), where the objective is demand response, i.e., modifying one's usage pattern when demand is high; demand response management in power networks in the large population regime has been covered in another chapter in the Handbook. Here, the agents are electricity consumers who must tradeoff the cost of modifying ones' usage (say by resetting their air conditioner temperature) versus the probability of a reward by obtaining many lottery tickets, under some belief about what the other consumers would do.

The asymptotic validity of the mean field assumption usually follows form a so-called *chaos hypothesis*, which essentially says that the correlation between the states of any finite subset of agents decays as the number of agents become large. Results of this nature are available in work by Graham and Méléard (1994) and can be used in the context of our scheduling game. There has been recent work studying the question of the conditions required to ensure that the mean field model is indeed the limiting case of the finite system. Work by Benaïm and Le Boudec (2008) and Borkar and Sundaresan (2012) provide regularity conditions under which the passage to the mean field is valid.

Conclusion

In this chapter, we considered three applications of game theory to problems related to routing and resource allocation in communication networks. In doing so, we explored three game theoretic equilibrium concepts in different settings, namely, (i) Nash Equilibrium in the problem of resource allocation to a finite set of agents, (ii) Wardrop Equilibrium in the context of selfish routing by an infinite number of agents, and (iii) Mean Field Equilibrium in the setting of repeated resource allocation to an infinite number of agents. In each case, we presented a model that pertains to a particular layer of the networking stack, and attempted to characterize the effects of strategic decision making on system performance as relevant to that layer. We noted that strategic decision making by agents can degrade system performance, and showed tight bounds on the performance degradation in these cases.

References

Adlakha S, Johari R (2013) Mean field equilibrium in dynamic games with strategic complementarities. Oper Res 61(4):971–989

Benaïm M, Le Boudec J-Y (2008) A class of mean field interaction models for computer and communication systems. Perform Eval 65(11–12):823–838

Billingsley P (2009) Convergence of probability measures. Wiley-Interscience, Hoboken

Borkar V, Sundaresan R (2012) Asymptotics of the invariant measure in mean field models with jumps. Stoch Syst 2(2):322–380

Braess D (1968) Über ein paradoxon aus der verkehrsplanung. Unternehmensforschung 12(1):258–268

Clarke EH (1971) Multipart pricing of public goods. Public Choice 11(1):17–33

Graham C, Méléard S (1994) Chaos hypothesis for a system interacting through shared resources. Probab Theory Relat Fields 100(2):157–174

Groves T (1973) Incentives in teams. Econometrica J Econom Soc 41(4):617–631

Hernández-Lerma O, de Ozak MM (1992) Discrete-time Markov control processes with discounted unbounded costs: optimality criteria. Kybernetika 28(3):191–212

Iyer K, Johari R, Sundararajan M (2014) Mean field equilibria of dynamic auctions with learning. Manag Sci 60(12):2949–2970

Johari R, Tsitsiklis JN (2004) Efficiency loss in a network resource allocation game. Math Oper Res 29(3):407–435

Johari R, Tsitsiklis JN (2005) Communication requirements of VCG-like mechanisms in convex environments. In: Proceedings of the Allerton Conference on Control, Communications and Computing, pp 1191–1196

Kelly FP (1997) Charging and rate control for elastic traffic. Eur Trans Telecommun 8:33–37

Li J, Xia B, Geng X, Ming H, Shakkottai S, Subramanian V, Xie L (2015) Energy coupon: a mean field game perspective on demand response in smart grids. ACM SIGMETRICS Perform Eval Rev 43(1):455–456

Li J, Bhattacharyya R, Paul S, Shakkottai S, Subramanian V (2017) Incentivizing sharing in realtime D2D streaming networks: a mean field game perspective. IEEE/ACM Trans Netw 25(1):3–17

Maheswaran RT, Basar T (2006) Efficient signal proportional allocation (ESPA) mechanisms: decentralized social welfare maximization for divisible resources. IEEE J Sel Areas Commun 24(5):1000–1009

Manjrekar M, Ramaswamy V, Shakkottai S (2014) A mean field game approach to scheduling in cellular systems. In: Proceedings IEEE INFOCOM, 2014, Toronto, pp 1554–1562

Meyn SP, Tweedie RL, Glynn PW (2009) Markov chains and stochastic stability, vol 2. Cambridge University Press, Cambridge

Pigou AC (1920) The economics of welfare. Palgrave Macmillan, New York

Roughgarden T (2015) Intrinsic robustness of the price of anarchy. J ACM 62(5):32

Roughgarden T (2016) Twenty lectures on algorithmic game theory. Cambridge University Press, Cambridge

Roughgarden T, Schoppmann F (2015) Local smoothness and the price of anarchy in splittable congestion games. J Econ Theory 156:317–342

Roughgarden T, Tardos E (2002) How bad is selfish routing? J ACM 49(2):236–259

Shakkottai S, Srikant R (2007) Network optimization and control. Found Tr Netw 2(3):271–379

Srikant R (2004) The mathematics of internet congestion control. Birkhauser, New York
Strauch RE (1966) Negative dynamic programming. Ann Math Stat 37(4):871–890
Tembine H, Le Boudec JY, El-Azouzi R, Altman E (2009) Mean field asymptotics of Markov decision evolutionary games and teams. In: International Conference on Game Theory for Networks (GameNets), pp 140–150
Vickrey W (1961) Counterspeculation, auctions, and competitive sealed tenders. J Financ 16(1):8–37
Xu J, Hajek B (2012) The supermarket game. In: IEEE International Symposium on Information Theory (ISIT, 2012), Cambridge, pp 2511–2515
Yang S, Hajek B (2007) VCG-Kelly mechanisms for allocation of divisible goods: adapting VCG mechanisms to one-dimensional signals. IEEE J Sel Areas Commun 25(6):1237–1243

Power System Analysis: Competitive Markets, Demand Management, and Security

27

Anibal Sanjab and Walid Saad

Contents

Abstract

In recent years, the power system has undergone unprecedented changes that have led to the rise of an interactive modern electric system typically known as the smart grid. In this interactive power system, various participants such as generation owners, utility companies, and active customers can compete, cooperate, and exchange information on various levels. Thus, instead of being centrally operated as in traditional power systems, the restructured operation is expected to rely on distributed decisions taken autonomously by its various interacting constituents. Due to their heterogeneous nature, these constituents can possess different objectives which can be at times conflicting and at other times aligned. Consequently, such a distributed operation has introduced various technical challenges at different levels of the power system ranging from energy management to control and security. To meet these challenges, game theory provides a plethora of useful analytical tools for the modeling and analysis of complex distributed decision making in smart power systems.

A. Sanjab (✉) · W. Saad
Virginia Tech, Blacksburg, VA, USA
e-mail: anibals@vt.edu; walids@vt.edu

T. Başar, G. Zaccour (eds.), *Handbook of Dynamic Game Theory*,
https://doi.org/10.1007/978-3-319-44374-4_28

The goal of this chapter is to provide an overview of the application of game theory to various aspects of the power system including: i) strategic bidding in wholesale electric energy markets, ii) demand-side management mechanisms with special focus on demand response and energy management of electric vehicles, iii) energy exchange and coalition formation between microgrids, and iv) security of the power system as a cyber-physical system presenting a general cyber-physical security framework along with applications to the security of state estimation and automatic generation control. For each one of these applications, first an introduction to the key domain aspects and challenges is presented, followed by appropriate game-theoretic formulations as well as relevant solution concepts and main results.

Keywords

Smart grid · Electric energy markets · Demand-side management · Power system security · Energy management · Distributed power system operation · Dynamic game theory

1 Introduction

The electric grid is the system responsible for the delivery of electricity from its production site to its consumption location. This electric system is commonly referred to as a "grid" due to the intersecting interconnections between its various elements. The three main constituents of any electric grid are generation, transmission, and distribution.

The *generation* side is where electricity is produced. The electricity generation is in fact a transformation of energy from chemical energy (as in coal or natural gas), kinetic energy (delivered through moving water or air), or atomic energy (for the case of nuclear energy) to electric energy. The *transmission* side of the grid is responsible for the transmission of this bulk generated electric energy over long distances from a source to a destination through high-voltage transmission lines. The use of high transmission voltage aims at reducing the involved electric losses. The *distribution* side, on the other hand, is the system using which electricity is routed to its final consumers. In contrast with the transmission side, on the distribution side, electricity is transported over relatively short distances at relatively low voltages to meet the electric demand of local customers.

In recent years, the electric grid has undergone unprecedented changes, aiming at increasing its reliability, efficiency, and resiliency. This, in turn, gave rise to an interactive and intelligent electric system known as the *smart grid*. As defined by the European Technology Platform for Smart Grids in their 2035 Strategic Research Agenda,

> "A smart electric grid is an electricity network that can intelligently integrate the actions of all users connected to it – generators, consumers and those that do both – in order to efficiently deliver sustainable, economic and secure electricity supplies."

The main drivers behind this evolution toward a smarter power system stem from an overhaul of three key elements of the grid: generation, transmission, and distribution. These major transformations can be summarized as follows:

1. *Restructuring of the electric energy market*: the recent restructuring of the electric generation and transmission markets transformed them from a centralized regulated market, in which the dispatch of the generation units is decided by a central entity that runs the system, into a competitive deregulated market. In this market, various generator owners, i.e., generation companies (GENCOs) and load serving entities (LSEs) submit offers and bids in a competitive auction environment. This market is cleared by an independent system operator based on which the amount of power that each of the GENCOs supplies and each of the LSEs receives along with the associated electricity prices are specified.
2. *Integration of renewable energy (RE) sources into the grid*: the vast integration of small-scale RE has transformed the distribution side and its customers from being mere consumers of electricity into electric "prosumers" who can consume as well as produce power and possibly feed it back to the distribution grid. In fact, distributed generation (DG) consists of any small-scale generation that is connected at the distribution side of the electric grid and which can, while abiding by the implemented protocols, pump electric energy into the distribution system reaping financial benefit to its owner. Examples of DG include photovoltaic cells, wind turbines, micro turbines, as well as distributed storage units (including electric vehicles). Similarly, large-scale RE sources have also penetrated the wholesale market in which their bulk produced energy is sold to the LSEs.
3. *The power grid as a cyber-physical system (CPS)*: the evolution of the traditional power system into a CPS is a byproduct of the massive integration of new sensing, communication, control, and data processing technologies into the traditional physical system. In the new cyber-physical power system, accurate and synchronized data is collected from all across the grid and sent through communication links for data processing which generates an accurate monitoring of the real-time state of operation of the system and enables a remote transmission of control signals for a more dependable, resilient, and efficient wide-area operation of the grid. Moreover, the integration of communication and information technologies into the grid has allowed utility companies to interact with their customers leading to a more economical and efficient operation of the distribution system and giving rise to the concepts of demand-side management (DSM), demand response (DR), and real-time pricing.

Clearly, the electric grid is gradually moving from a centralized operation, in which decisions are taken by a centralized entity, to a distributed architecture. In fact, in this new architecture, decisions governing the operation of the grid are taken in a distributed manner by the constituents of the grid such as GENCOs and LSEs, on the market side, and utility companies, DG owners, and consumers on the distribution side. Given that these constituents can have different objectives, their decisions might be conflicting, and, as such, they will engage in noncooperative and competitive interactions whose outcome impacts the revenues of each constituent as well as the overall performance of the grid. On the other hand, at other times, those constituents can seek to cooperate in order to reach a mutual goal. This gives rise to strategic interactions between the various components of the grid through which the decision taken by one component to meet its objective needs to take into consideration the decisions that can be taken by others.

Beyond energy management and market scenarios, the study of strategic behavior will also play a very important role in securing the grid. In fact, the interconnectivity between the various elements of the power system and its critical reliance on its underlying communication and computational systems make the grid more vulnerable to many types of attacks, of cyber and physical nature, aiming at compromising its functionality. In this regard, by exploiting hidden vulnerabilities and the dense system interconnectivity, an attacker uses its limited resources to devise an attack strategy aiming at penetrating the system and spreading failures to maximize the inflicted damage on the grid. On the other hand, using limited resources, the grid operator (defender) must protect the system against such attacks by choosing proper defense mechanisms. In this regard, when devising a defense strategy, the defender needs to study and account for the strategic behavior of potential attackers and vice versa. As a result, this shows the importance of strategic modeling and analysis for having a better understanding of the security state of the grid and devising strategies to improve this security.

Clearly, the overall operation and security of a power system will involve complex interactive decisions between its various elements. Modeling these decision making processes is essential for anticipating and optimizing the overall performance of the grid. In this respect, game theory provides a suitable framework to model the emerging interactive decision making processes in modern power systems (For a general overview of game-theoretic approach to the smart grid, the reader is referred to Saad et al. (2012)). In particular, given the dynamic nature of the grid's operation and control, dynamic game theory provides very useful modeling tools to analyze the grid's dynamic and distributed decision making. For example, participants in an energy market as well as demand-side management mechanisms periodically and repeatedly make decisions based on information available to them from past time periods and projections of future system behavior. Moreover, the power system itself is a dynamic system whose states evolve over time depending on the various operational and control actions taken. As a result, the outcome of any strategic interaction depends not only on the decisions taken by the grid's constituents but also on the state of the system. In this respect, dynamic game models provide the required analytical tools to capture such dynamic decision making over the dynamic power system.

The goal of this chapter is to introduce various game-theoretic techniques that can be applied to model the strategic behavior in various areas of the power system. First, the strategic interactions between different market participants in a wholesale competitive electric energy market is studied. Second, the distribution side is considered shedding the light on the strategic behavior of the various prosumers and their overall effect on the distribution system in demand-side management schemes as well as on microgrids and their energy exchange. Finally, the security of the cyber-physical power system is studied using game-theoretic techniques presenting a general dynamic security framework for power systems in addition to considering specific security areas pertaining to the security of automatic generation control and power system state estimation.

In summary, the goals of this chapter are threefold:

1. Discuss the areas of the power system in which the strategic behavioral modeling of the involved parties is necessary and devise various game-theoretic models that reflect this strategic behavior and enables its analysis.
2. Investigate different types of games which can be developed depending on the studied power system application.
3. Provide reference to fundamental research works which have applied game-theoretic methods for studying the various sides of the smart grid covered.

One should note that even though proper definitions and explanations of the various concepts and models used are provided in this chapter, prior knowledge of game theory, which can be acquired through previous chapters of this book, would be highly beneficial to the reader. In addition, the various power system models used will be clearly explained in this book chapter. Thus, prior knowledge of power systems is not strictly required. However, to gain a deeper knowledge of power system concepts and models, the reader is encouraged to explore additional resources on power system analysis and control such as for power system analysis (Glover et al. 2012), for power system operation and control (Wood and Wollenberg 2012), for electric energy markets (Gomez-Exposito et al. 2009), for power system state estimation (Abur and Exposito 2004), for power system protection (Horowitz et al. 2013), and for power system dynamic modeling (Sauer and Pai 1998), among many others.

Here, since each section of this chapter treats a different field of power system analysis, the notations used in each section are specific to that section and are chosen to align with typical notations used in the literature of their corresponding power system field. To this end, a repeated symbol in two different sections can correspond to two different quantities, but this should not create any source of ambiguity.

2 Competitive Wholesale Electric Energy Markets

In a traditional power system, to operate the system in the most economical manner, the regulating entity operating the whole grid considers the cost of operation of every generator and dispatches these generators to meet the load while minimizing the total generation cost of the system. In fact, a generator cost function $f_i(P_{G_i})$, which reflects the cost of producing P_{G_i} MW by a certain generator G_i, is usually expressed as a polynomial function of the form:

$$f_i(P_{G_i}) = a_i P_{G_i}^2 + b_i P_{G_i} + c_i. \tag{27.1}$$

To minimize the total cost of generation of the system, the operator solves an economic dispatch problem or what is commonly known as an optimal power flow (OPF) problem, whose linearized version is usually defined as follows[1]:

$$\min_{\mathbf{P}} \sum_{i=1}^{N_G} f_i(P_{G_i}) \tag{27.2}$$

subject to:

$$\sum_{i=1}^{N}(P_i - D_i) = 0, \tag{27.3}$$

$$P_{G_i}^{\min} \leq P_{G_i} \leq P_{G_i}^{\max}, \forall i \in \{1, \ldots, N_G\}, \tag{27.4}$$

$$F_l^{\min} \leq F_l \leq F_l^{\max}, \forall l \in \{1, \ldots, L\}, \tag{27.5}$$

where $\mathbf{P} = [P_{G_1}, \ldots, P_{G_{N_G}}]$ and N_G, N, and L represent, respectively, the number of generators, buses,[2] and transmission lines in the system. P_i and D_i are, respectively, the power injection and load at a bus i. Thus, $P_i = 0$ ($D_i = 0$) corresponds to the case in which no generator (or load) is connected to bus i. F_l in (27.5) corresponds to the power flow over a transmission line l which is constrained by the operational and thermal limits, $\{F_l^{\min}, F_l^{\max}\}$, of the associated line. In the lossless MW-only simplified power flow formulation (known as DC power flow), the flow over a transmission line l connecting buses i and j can be expressed as:

$$F_l \triangleq P_{ij} = (\theta_i - \theta_j)/X_{ij}, \tag{27.6}$$

where X_{ij} is the transmission line reactance (expressed in p.u.[3]) while θ_i and θ_j correspond to the phase angle of the voltage at buses i and j, respectively, expressed

[1]This is the linearized lossless OPF formulation commonly known as the DCOPF. The more general nonlinear OPF formulation, known as the ACOPF, has more constraints such as limits on voltage magnitudes and reactive power generation and flow. Moreover, the ACOPF uses the AC power flow model rather than the linearized DC one. The ACOPF is a more complex problem to solve whose global solution can be very complex, and computationally challenging to compute, with the increase in the number of constraints involved. Hence, practitioners often tend to use the DCOPF formulation for market analyses. Here, the use of the terms AC and DC is just a notation that is commonly used in energy markets and does not, in any way, reflect that the used current in DCOPF or DC power flow is actually a direct current.

[2]In power system jargon, a bus is an electric node.

[3]p.u. corresponds to per-unit which is a relative measurement unit expressed with respect to a predefined base value (Glover et al. 2012).

in radians. As such, the power injection (power entering the bus) at a certain bus i, P_i, is equal to the summation of powers flowing out of that bus. Thus,

$$P_i = \sum_{j \in \mathscr{O}_i} P_{ij}, \tag{27.7}$$

where $\mathscr{O}_i$ is the set of neighboring buses connected to bus i.

The expressions in (27.6) and (27.7) correspond to the MW-only power flow model. This MW-only model is valid under the following assumptions: (i) the system is lossless (transmission lines' resistances are much smaller than their reactances), (ii) the shunt susceptance in the transmission lines π-model[4] is negligible, (iii) the voltage magnitudes of the buses are assumed not to deviate much from their flat profile (voltage magnitude at each bus is equal to 1 p.u.), and (iv) the voltage phase angles are assumed to be small. In steady-state operation, such assumptions hold to a reasonable extent.[5] As a result, this linearized model is commonly used in energy market applications in which steady-state system operation is assumed and fast OPF solutions are needed.

In a competitive energy market environment, rather than minimizing the total cost of generation, the system operator, referred to as ISO (independent system operator) hereinafter, does not own nor operate the generators. The ISO, in fact, receives generation offers from the GENCOs and demand bids from the LSEs and aims to maximize the underlying social welfare (to be defined next). The generation offers are normally an increasing function of the delivered power P_{G_i}, whereas the demand bids are represented by a decreasing function of the demanded power D_j. In various models, offers and bids are submitted block-wise[6] as shown in Fig. 27.1. In this regard, consider a GENCO owning a generator G_i. The submitted offer of G_i, $\gamma_i(P_{G_i})$, takes the following form:

$$\gamma_i(P_{G_i}) = \begin{cases} \rho_{G_{i,1}} & \text{if } P_{G_i}^{\min} \le P_{G_i} \le P_{G_{i,1}}, \\ \rho_{G_{i,2}} & \text{if } P_{G_{i,1}} \le P_{G_i} \le P_{G_{i,2}}, \\ \cdots \\ \rho_{G_{i,K_i}} & \text{if } P_{G_{i,K_i-1}} \le P_{G_i} \le P_{G_i}^{\max}, \end{cases} \tag{27.8}$$

where $P_{G_{i,x}} > P_{G_{i,y}}$ and $\rho_{G_{i,x}} > \rho_{G_{i,y}}$ when $x > y$. The offer of G_i is said to be composed of K_i blocks. In a similar manner, a bid of an LSE j, $\omega_j(D_j)$, composed of K_j blocks is expressed as follows:

[4]The π-model is a common model of transmission lines (Glover et al. 2012).

[5]If such assumptions do not hold, the standard nonlinear power flow equations will have to be used. The nonlinear power flow equations can be found in Glover et al. (2012) and Wood and Wollenberg (2012).

[6]Various offer structures are considered in the literature and in practice, including block-wise, piece-wise linear, as well as polynomial structures. Here, a block-wise offer and bid structures are used; however, a similar strategic modeling can equally be carried out for any of the other structures.

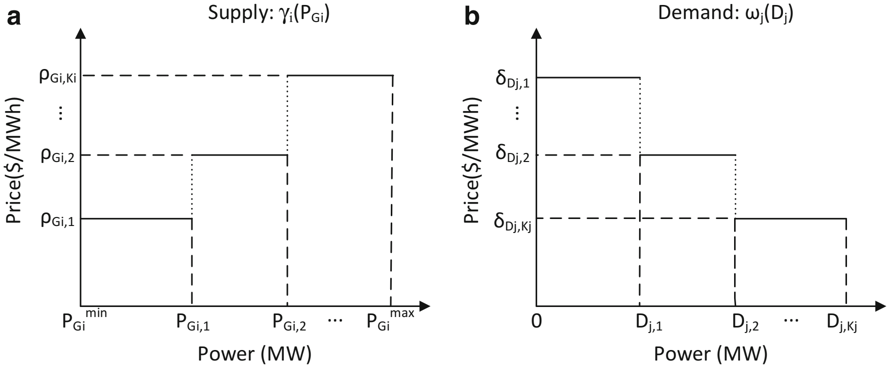


Fig. 27.1 (**a**) $\gamma_i(P_{G_i})$: Generation offer curve of generator G_i, (**b**) $\omega_j(D_j)$: Demand curve for load D_j

$$\omega_j(D_j) = \begin{cases} \delta_{D_{j,1}} & \text{if } 0 \le D_j \le D_{j,1}, \\ \delta_{D_{j,2}} & \text{if } D_{j,1} \le D_j \le D_{j,2}, \\ \cdots & \\ \delta_{D_{j,K_j}} & \text{if } D_{j,K_j-1} \le D_j \le D_{j,K_j}, \end{cases} \tag{27.9}$$

where $D_{j,x} > D_{j,y}$ and $\delta_{j,x} < \delta_{j,y}$ when $x > y$. Thus, after receiving N_G different supply offers and N_D demand bids, assuming N_D to be the number of load buses in the system, from the various GENCOs and LSEs of the market, the ISO dispatches the system in a way that maximizes the social welfare, W, by solving

$$\max_{\mathbf{P},\mathbf{D}} \; W = \sum_{j=1}^{N_D} \sum_{k=1}^{K_j} \delta_{D_{j,k}} D_k^{(j)} - \sum_{i=1}^{N_G} \sum_{k=1}^{K_i} \rho_{G_{i,k}} P_k^{(G_i)}, \tag{27.10}$$

subject to:

$$0 \le D_k^{(j)} \le D_{k-\max}^{(j)} \; \forall j,k, \text{ and } 0 \le P_k^{(G_i)} \le P_{k-\max}^{(G_i)} \; \forall i,k, \tag{27.11}$$

while meeting the same operational constraints as those in (27.3) and (27.5). The constraints in (27.11) insure preserving the demand and generation block sizes, as shown in Fig. 27.1. In this regard, $D_{k-\max}^{(j)} = D_{j,k} - D_{j,k-1}$ is the MW-size of block k of load j, $D_k^{(j)}$ is a decision variable specifying the amount of power belonging to block k that j obtains, and $\delta_{D_{j,k}}$ is the price offered by j for block k. Similarly, $P_{k-\max}^{(G_i)}$ is the size of block k offered by generator G_i, $\rho_{G_{i,k}}$ is the price demanded by G_i for that block, and $P^{(G_i)}$ is a decision variable specifying the amount of power corresponding to block k that G_i is able to sell. In this formulation, **P** and **D** is composed of $D_k^{(j)}$ $\forall j, k$ and $P_k^{(G_i)}$ $\forall i, k$. By solving this problem, the ISO clears the market and announces the power output level, $P_{G_i}^* = \sum_{k=1}^{K_i} P_k^{(G_i)*}$, requested from generation unit i, for $i = \{1, \ldots, N_G\}$, and the amount of power, $D_j^* = \sum_{k=1}^{K_j} D_k^{(j)*}$ allocated to each load D_j for $j = \{1, \ldots, N_D\}$. This market-clearing procedure generates what is known as locational marginal prices (LMPs) where the LMP at bus n, μ_n, denotes the price of electricity at n[7]. Consider a GENCO i owning n_{G_i} generation units and an LSE j serving n_{D_j} nodes. The profit Π_i of GENCO i and cost C_j of LSE j are computed, respectively, as follows:

$$\Pi_i = \sum_{r=1}^{n_{G_i}} [\mu^{(G_r)} P_{G_r}^* - f_i(P_{G_r}^*)], \tag{27.12}$$

$$C_j = \sum_{r=1}^{n_{D_j}} \mu^{(D_r)} D_r^*, \tag{27.13}$$

where $f_i(P_{G_r}^*)$ is the cost of producing $P_{G_r}^*$ MW by generator G_r while $\mu^{(G_r)}$ and $\mu^{(D_r)}$ correspond, respectively, to the LMP at the bus at which G_r or D_r is connected.

In some models, the demand is not considered to be flexible and thus always needs to be met. In that case, no demand bids are considered and (27.10) turns into

$$\min_{\mathbf{P}} \; W = \sum_{i=1}^{N_G} \sum_{k=1}^{K_i} \rho_{G_{i,k}} P_k^{(G_i)}, \tag{27.14}$$

while meeting the operational and block size constraints.

Given that the energy market is not a fully competitive market (as studied using game-theoretic models in Guan et al. (2001); Li and Shahidehpour (2005); Nanduri and Das (2007)), the submitted offer (bid) of each GENCO (LSE) affects the market-clearing process and hence affects the profit (cost) of the GENCO (LSE) itself as well as the profit and costs of other GENCOs and LSEs participating in the market.

2.1 Strategic Bidding as a Static Game

Strategic bidding is typically a periodically repeated process following the adopted market architecture. At every period at which the market opens for bids, participants submit new offers and bids, and the market is cleared accordingly. Hence, to be able to study this dynamic bidding structure and its evolution over time, it is important to have an understanding of the bidding process at every separate period. This is achieved by using a static strategic bidding model in which a snapshot of the energy

---

[7] This LMP-based nodal electricity pricing is a commonly used pricing technique in competitive markets of North America. Here, we note that alternatives to the LMP pricing structure are implemented in a number of other markets and mainly follow a zonal-based pricing approach.

market is considered. In this snapshot, market participants submit their offers and bids with no regard to previously taken actions or to future projections. Such an analysis provides interesting insights and a set of tools to be used when a general dynamic bidding model is derived.

The problem of choosing the optimal offers that each GENCO, from a set of $\mathcal{G}$ GENCOs, should submit for its n_{G_i} generators, in competition with other GENCOs, has been considered and analyzed in Li and Shahidehpour (2005) which is detailed next.

In this work, each GENCO, i, submits a generation offer for each of its generators, G_j for $j \in \mathcal{G}_i$ where $\mathcal{G}_i$ is the set of all generators owned by i and $|\mathcal{G}_i| = n_{G_i}$, considering three different supply blocks, $K_i = 3$, while choosing $\rho_{G_{j,k}}$ as follows:

$$\rho_{G_j,k} = \tau_{i,j} \frac{\partial f_j(P_{G_j})}{\partial P_{G_j}}|_{(P_{G_j}=P_{G_j,k})} = \tau_{i,j}(2a_j P_{G_j} + b_j)|_{(P_{G_j}=P_{G_j,k})}, \qquad (27.15)$$

where $\tau_{i,j}$ is the bidding strategy of GENCO i corresponding to its generator j with $\tau_{i,j} = 1$ in case G_j is a price taker. $\boldsymbol{\tau}_i$ is the vector of bidding strategies of all generators owned by GENCO i.

Thus, this decision making process of the different GENCOs can be modeled as a noncooperative game $\Xi = \langle \mathcal{I}, (\mathcal{S}_i)_{i\in\mathcal{I}}, (\Pi_i)_{i\in\mathcal{I}} \rangle$. Here, $\mathcal{I}$ is the set of GENCOs participating in the market, which are the players of this game. $\mathcal{S}_i$ is the set of actions available to player $i \in \mathcal{I}$ which consists of choosing a bidding strategy $\boldsymbol{\tau_i}$ which has limits set by GENCO i such that $\tau_{i,j}^{\min} \le \tau_{i,j} \le \tau_{i,j}^{\max}$. Π_i is the utility function of player $i \in \mathcal{I}$ corresponding to its payoff given in (27.12).

Thus, the different GENCOs submit their generation offers to the ISO, i.e., choose their bidding strategy $\boldsymbol{\tau}_i$ for each of their generators; the ISO receives those offers and sends dispatch decisions for the generators in a way that minimizes the cost of meeting the load subject to the existing constraints, i.e., by solving problem (27.14).

To this end, in the case of complete information, a GENCO can consider any strategy profile of all GENCOs, i.e., $\boldsymbol{\tau} = [\boldsymbol{\tau}_1, \boldsymbol{\tau}_2, \ldots, \boldsymbol{\tau}_G]$ and subsequently run an OPF to generate the resulting dispatch schedules, LMPs, and achieved payoffs. Based on this acquired knowledge, a GENCO can characterize its best bidding strategy, $\boldsymbol{\tau}_i^*$, facing any strategy profile, $\boldsymbol{\tau}_{-i}$, chosen by its opponents.

The most commonly adopted equilibrium concept for such noncooperative games is the Nash equilibrium (NE). The NE is a strategy profile of the game in which each GENCO chooses the strategy $\boldsymbol{\tau}_i^*$ which maximizes its payoff (27.12) in response to $\boldsymbol{\tau}_{-i}^*$ chosen by the other GENCOs. In other words, the bidding process reaches an NE when all the GENCOs simultaneously play best response strategies against the bidding strategies of the others.

2.1.1 Game Solution and Main Results

Obtaining a closed-form expression for the NE of this strategic bidding game is highly challenging since it requires analytical solutions of two optimization problems. In the first, each GENCO chooses its bidding strategy to maximize its profit (expected profit); while in the second, the ISO receives the GENCOs' offers and runs an OPF to meet the load while maximizing the social welfare. Hence, to solve the proposed game, the authors in Li and Shahidehpour (2005) introduce an iterative numerical solution algorithm which is guaranteed to reach an NE when it exists.

For analysis, the work in Li and Shahidehpour (2005) treats an eight-bus power system with six generation units and three GENCOs where each GENCO owns two generation units. The results show the advantage of performing strategic bidding, following the NE strategies, as compared to always bidding the marginal cost of production of each unit. In fact, in perfect competitive market models, all suppliers are price takers, and their optimal supply quantity is one that equates the marginal cost of production to the given price. However, the wholesale electric energy market is considered not to be a fully competitive market due to the physical features and limitations of the power system such as the transmission congestion and the insufficiency of transmission capacity. Thus, following the NE strategies, a strategic GENCO achieves a payoff that is at least as good as the one it obtains by submitting offers equal to its marginal cost.

This static game formulation can form the basis for a dynamic modeling of strategic behavior in wholesale electric energy market. This dynamic process is explained next.

2.2 Dynamic Strategic Bidding

Most competitive electric energy markets adopt a forward market known as the day-ahead (DA) market. In the DA market, participants submit their energy offers and bids for the next day, usually on an hourly basis, and the market is cleared based on those submitted offers and a forecast[8] of next day's state of operation (loads, outages, or others) using the OPF formulation provided in (27.10). This process is repeated on a daily basis requiring a dynamic strategic modeling.

Consider the DA market architecture with non-flexible demands, and let $\mathscr{I}$ be the set of GENCOs participating in the market each assumed to own one generation unit; thus, $|\mathscr{I}| = N_G$. Each GENCO at day $T-1$ submits its generation offers for the 24 h of T. Let $t \in \{1, \ldots, 24\}$ be the variable denoting a specific hour of T. Let $s_i^{t,T} \in \mathscr{S}_i$ be the offer submitted by GENCO i for time period t of day T. $s_i^{t,T}$ corresponds to choosing a $\gamma_i(P_{G_i})$ as in (27.8) for generator i at time t of day T.

[8]The DA market is followed by a real-time (RT) market to account for changes between the DA projections and the RT actual operating conditions and market behavior. In this section, the focus is on dynamic strategic bidding in the DA market.

Thus, the bidding strategy of GENCO i for day T corresponds to the 24 h bidding profile $\mathbf{s}_i^T = (s_i^{1,T}, s_i^{2,T}, \dots, s_i^{24,T})$ where $\mathbf{s}_i^T \in \mathscr{S}_i^{24}$. This process is repeated daily, i.e., for incremental values of T. However, based on the changes in load and system conditions, the characteristics of the bidding environment, i.e., the game differ from one day to the other. These changing characteristics of the game are known as the game's states. A description of the "state" of the game is provided next.

Let the state of the game for a given time t of day T, $\mathbf{x}^{t,T}$, be the vector of hourly loads and LMPs at all buses N of the system during t of T. Let $\mathscr{N}$ be the set of all buses in the system with $|\mathscr{N}| = N$, $\mathscr{N}_G$ be the set of all generators with $|\mathscr{N}_G| = N_G$, and $d_n^{t,T}$ and $\mu_n^{t,T}$ be, respectively, the load and LMP at bus $n \in \mathscr{N}$ for time t of day T. In this regard, let $\mathbf{d}^{t,T} = (d_1^{t,T}, d_2^{t,T}, \dots, d_N^{t,T})$ and $\boldsymbol{\mu}^{t,T} = (\mu_1^{t,T}, \mu_2^{t,T}, \dots, \mu_N^{t,T})$ be, respectively, the vectors of loads and LMPs at all buses at hour t of day T. Thus, $\mathbf{x}^{t,T} = (\mathbf{d}^{t,T}, \boldsymbol{\mu}^{t,T})$ is the state vector for time t of T and $\mathbf{x}^T = (\mathbf{x}^{1,T}, \mathbf{x}^{2,T}, \dots, \mathbf{x}^{24,T})$ is the vector of states of day T.

It is assumed that the DA offers for day $T + 1$ are submitted at the end of day T. Thus, $\mathbf{x}^T$ is available to each GENCO $i \in \mathscr{I}$ to forecast $\mathbf{x}^{T+1}$ and choose $\mathbf{s}_i^T$. The state of the game at $T + 1$ depends on the offers presented in day T, $\mathbf{s}_i^T$ for $i \in \{1, \dots, N_G\}$, as well as on realizations of random loads. Hence, the state $\mathbf{x}^{T+1}$ of the game at day $T + 1$ is a random variable.

Based on the submitted offers and a forecast of the loads, the ISO solves an OPF problem similar to that in (27.14) while meeting the associated constraints to produce the next day's hourly generator's outputs, $P_{G_i}^{t,(T+1)*}$ for $i \in \mathscr{N}_G$, as well as hourly LMPs, $\boldsymbol{\mu}^{t,(T+1)*}$, for $t \in \{1, \dots, 24\}$. Thus, the daily payoff of a GENCO i for a certain offer strategy $\mathbf{s}_i^T$ and state $\mathbf{x}_i^T$ based on the DA schedule is given by

$$\Pi_i^T(\mathbf{s}_i^T, \mathbf{x}^T) = \sum_{t=1}^{24} \left[\mu_i^{t,(T+1)*} P_{G_i}^{t,(T+1)*} - f_i\left(P_{G_i}^{t,(T+1)*}\right)\right], \tag{27.16}$$

where $f_i(P_{G_r})$ is as defined in (27.1).

Such a dynamic modeling of strategic bidding has been presented and analyzed in Nanduri and Das (2007) in which dynamic decision making has been modeled using a stochastic game. In contrast to deterministic games in which the outcome is solely dictated by the strategies chosen by the players, a stochastic game is one in which the outcome depends on the players' actions as well as the "state" of the game. In stochastic games, the game transitions from one state to the other (i.e., from one game to the other) based on the current state (i.e., current game) and on transition probabilities influenced by the actions chosen by the players.

A stochastic game is defined by the set of players, set of actions available to each player, state vector, state transition probabilities, and payoffs. In the dynamic bidding model, the set of players is the set of GENCOs, the set of actions of each player i is defined by $\mathscr{S}_i^{24}$ (a certain action is given by $\mathbf{s}_i^T$), and the state vector for a day T is defined by $\mathbf{x}^T$ ($\mathbf{s}_i^T$ and $\mathbf{x}^T$ should be discretized to fit in a discrete stochastic game model). As for the transition probabilities from one state to the other, the work in Nanduri and Das (2007) discusses the complexity of calculating these transition probabilities for this dynamic strategic bidding model and uses a simulation based environment to represent state dynamics. The payoff of each GENCO i over the multiple number of days in which it participates in the market can take many forms such as the summation, discounted summation or average of the stage payoff Π_i^T, defined in (27.16), over the set of days $\mathscr{T}$. The goal of each GENCO is to maximize this aggregate payoff over the multi-period time frame.

2.2.1 Game Solution and Main Results

This dynamic strategic bidding game is a nonzero-sum stochastic game. This type of stochastic games is known to be very challenging to solve. To this end, the authors in Nanduri and Das (2007) propose a solution framework based on reinforcement learning (RL) to obtain an approximation of the game's solution. This algorithm is applied in Nanduri and Das (2007) to a 12-bus test electric system with eight generators and four loads. The derived numerical results compute the market power (MP) that each of the GENCOs can have under various auction architectures (uniform auction, discriminatory auction, and second price auction) while investigating the effect of system conditions, such as transmission line congestion, on the rise of these market powers. Moreover, the sensitivity of the MP to the levels of demand, congestions, and type of auction used has been analyzed. This sensitivity analysis has shown that MP levels are most affected by the load level followed by the auction type and then the congestion level.

Hence, the use of this dynamic bidding strategy model provides the system operator with the analytical tools to assess market powers and the sensitivity of these market powers to various market characteristics and system operating conditions.

3 Strategic Behavior at the Distribution System Level

The integration of distributed and renewable generation into the distribution side of the grid and the proliferation of plug-in electric vehicles, along with the widespread deployment of new communication and computation technologies, have transformed the organization of the distribution system of the grid from a vertical structure between the electric utility company and the consumers to a more interactive and distributed architecture. In this interactive architecture, the consumers and the utility company can engage in a two-way communication process in which the utility company can send real-time price signals and energy consumption recommendations to the consumers to implement an interactive energy production/-consumption demand-side management. For example, using DSM, electric utilities seek to implement control policies over the consumption behavior of the consumers by providing them with price incentives to shift their loads from peak hours to off-peak hours. Participating in DSM mechanisms must simultaneously benefit the two parties involved. In fact, by this shift in load scheduling, utility companies can reduce energy production costs by reducing their reliance on expensive generation units with fast ramping abilities which are usually used to meet peak load levels. On

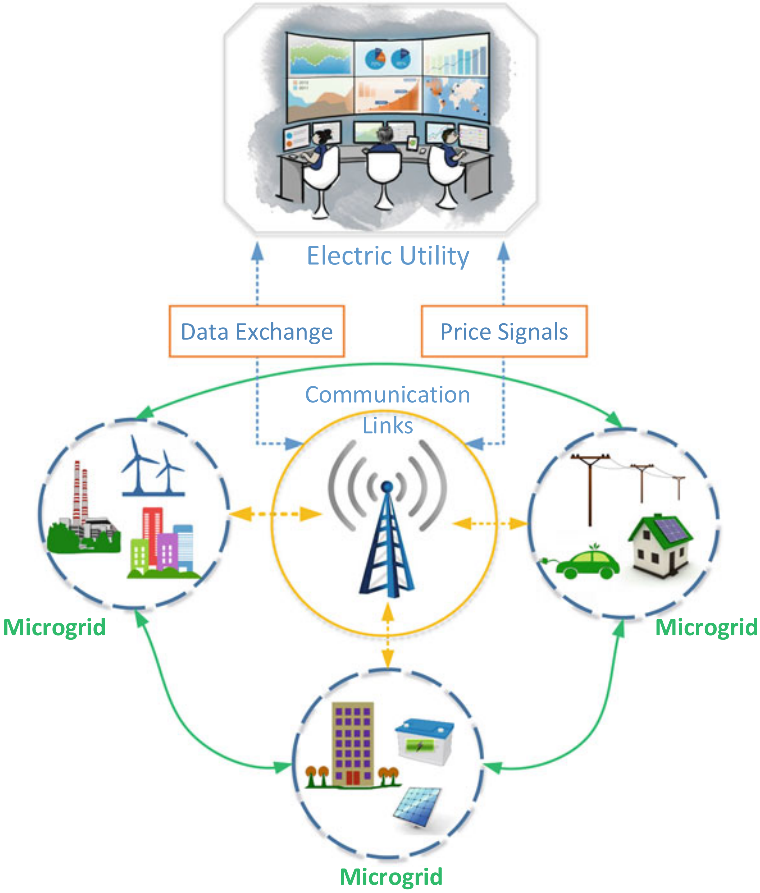


Fig. 27.2 Illustration of the various interactions in a smart distribution system

the other hand, by shifting their consumption, the consumers will receive reduced electricity prices.

Moreover, the proliferation of distributed generation has given rise to what is known as microgrids. A microgrid is a local small-scale power system constituted of locally interconnected DG units and loads and which is connected to the main distribution grid but can also self-operate, as an island, in the event of disconnection from the main distribution system. Thus, microgrids clearly reflect the distributed nature of the smart distribution system as they facilitate decomposing the macro-distribution system into a number of small-scale microgrids which can operate independently as well as exchange energy following implemented energy exchange protocols. This distributed nature of the distribution system and the interactions between its various constituents is showcased in Fig. 27.2.

As a result, this distributed decision making of individual prosumers, which represent customers who can produce and consume energy, as well as the decision making of groups of prosumers forming a microgrid introduce various operational challenges. The understanding and analysis of these challenges require a global distributed decision making model in which various entities with different objectives interact. Such modeling can be formulated using the tools of game theory.

This section focuses on the application of game-theoretic tools to study demand-side management as well as the problems of energy exchange and coalition formation between microgrids. As part of DSM, this section will focus on modeling the analysis and management of the energy consumption trends of electric vehicles. Throughout this section, the use of cooperative as well as noncooperative game theory is highlighted following the treated application.

3.1 Demand-Side Management

Due to the possibility of direct interactions between electricity providers and customers within the smart grid, DSM has emerged as a framework using which electric utility companies implement energy policies to guide the consumption patterns of the customers with the goal of improving the aggregate load profile of the distribution system. Hence, the essence behind DSM is understanding how a price signal sent by the utility company will affect the electricity demand and consumption patterns of each of the consumers. In this respect, in a DSM model, the utility company has to take into consideration the possible reaction of the consumers to the price it declares before announcing that price. When the price is decided, and a price signal is sent to the customers, these customers then react by choosing their levels of consumption. This hierarchical interaction can be modeled using a Stackelberg game model as introduced in Maharjan et al. (2013) and further developed in Maharjan et al. (2016).

Here, the focus will be shed on the response of the customers to dynamic real-time pricing through studying charging and consumption patterns of electric vehicles (EVs). Electric vehicles will be an integral component of tomorrow's power systems. EVs are typically equipped with batteries which can be charged via a plug-in connection to the local electricity grid. Allowing the EVs to sell their stored power back to the grid will turn them into a mobile type of energy prosumers who can play a vital part in DSM mechanisms. The charging strategy of an EV owner (referred to simply as EV hereinafter) is a dynamic control policy aiming at minimizing a cost function incorporating the real-time price of electricity which, in turn, depends on the aggregate demand for electricity by all consumers. Thus, the optimal charging strategies of all EVs are coupled through their resulting market price. As such, dynamic game-theoretic techniques can play a vital role in modeling and analyzing the charging policies of the various EVs. However, in practice, as the number of EVs in the system significantly grows, the effect of the charging strategy of a single EV owner on the actual aggregate real-time price significantly decreases. Thus, the charging policy of each EV is not a response to the charging actions of each electric vehicle in the system but rather to the aggregate charging actions of all the EVs, requiring a mean field game analysis to model their strategic behavior. To this end, the work in Couillet et al. (2012) proposes a differential mean field game to model the charging policies of EV owners as detailed next.

Even though mean field games have been formulated also for the nonhomogeneous case, we will restrict our discussion here to the homogeneous case, which assumes that all EVs are identical and enter the game symmetrically.[9] The validity of this assumption directly stem from the type of EVs connected to a particular grid.

Consider a distribution grid spanning a geographical area containing a set $\mathscr{V} = \{1, 2, \ldots, V\}$ of EVs. Let $x_t^{(v)}$ be the fraction of the battery's full capacity stored at the battery of EV $v \in \mathscr{V}$ at time $t \in [0, \ldots, T]$ where $x_t^{(v)} = 0$ denotes an empty battery and $x_t^{(v)} = 1$ denotes that the battery is fully charged. Let $g_t^{(v)}$ be the consumption rates of v at time t and $\alpha_t^{(v)}$ be its energy charging rate (or discharging rate in case EVs are allowed to sell energy back to the grid). The evolution of $x_t^{(v)}$ from an initial condition $x_0^{(v)}$ is governed by the following differential equation:

$$\frac{dx_t^{(v)}}{dt} = \alpha_t^{(v)} - g_t^{(v)}. \tag{27.17}$$

In (27.17), the evolution of $x_t^{(v)}$ for a deterministic consumption rate $g_t^{(v)}$ is governed by v's chosen charging rate $\alpha_t^{(v)}$ for $t \in [0, \ldots, T]$. To this end, the goal of v is to choose a control strategy $\alpha^{(v)} = \{\alpha_t^{(v)}, 0 \le t \le T\}$ which minimizes a charging cost function, J_v. This cost function incorporates the energy charging cost (which would be negative in case energy is sold to the grid) at the real-time energy price at $t \in [0, \ldots, T]$ given by the price function $p_t(\boldsymbol{\alpha}_t) : \mathbb{R}^V \to \mathbb{R}$ where $\boldsymbol{\alpha}_t = [\alpha_t^{(1)}, \ldots, \alpha_t^{(V)}]$ is the energy charging rates of all EVs at time t. In addition to the underlying energy purchasing costs, an EV also aims at minimizing other costs representing battery degradation at a given charging rate, $\alpha_t^{(v)}$, given by $h_t^{(v)}(\alpha_t^{(v)}) : \mathbb{R} \to \mathbb{R}$, as well as the convenience cost for storing a portion $x_t^{(v)}$ of the full battery capacity at time t captured by the function $f_t^{(k)}(x_t^{(k)}) : [0, 1] \to \mathbb{R}$. As such, the cost function of EV v can be expressed as

$$J_v(\alpha^{(v)}, \boldsymbol{\alpha}^{(-v)}) = \int_0^T \left(\alpha_t^{(v)} p_t(\boldsymbol{\alpha}_t) + h_t^{(v)}(\alpha_t^{(v)}) + f_t^{(v)}(x_t^{(v)})\right) dt + \kappa^{(v)}(x_T^{(v)}), \tag{27.18}$$

where $\boldsymbol{\alpha}^{(-v)}$ denotes the control strategies of all EVs except for v and $\kappa^{(v)}(x_T^{(v)}) : [0, 1] \to \mathbb{R}$ is a function representing the cost for v to end the trading period with $x_T^{(v)}$ battery level and which mathematically guarantees that the EVs would not sell all their stored energy at the end of the trading period.

Such a coupled multiplayer control problem can be cast as a differential game in which the set of players is $\mathscr{V}$, the state of the game at time t is given by

[9]For a discussion on general mean field games with heterogeneous players, see ▶ Chap. 7, "Mean Field Games" in this Handbook.

$\mathbf{x}_t = [x_t^{(1)}, \ldots, x_t^{(V)}]$, and whose trajectory is governed by the initial condition $\mathbf{x}_0$ and the EVs' control inputs as indicated in (27.17). To this end, the objective of each EV is to choose the control strategy $\alpha^{(v)}$ to minimize its cost function given in (27.18).

To solve this differential game for a relatively small number of EVs, and assuming that each EV has access only to its current state, one can attempt to find an optimal multiplayer control policy, $\boldsymbol{\alpha}^* = [\alpha^{*(1)}, \ldots, \alpha^{*(V)}]$, known as the own-state feedback Nash equilibrium which is a control profile such as for all $v \in \mathscr{V}$ and for all admissible control strategies:

$$J_v(\alpha^{*(v)}, \boldsymbol{\alpha}^{*(-v)}) \leq J_v(\alpha^{(v)}, \boldsymbol{\alpha}^{*(-v)}). \tag{27.19}$$

However, as previously stated, with a large number of EVs, the effect of the control strategy of a single EV on the aggregate electricity price is minimal which promotes the use of a mean field game approach that, to some extent, simplifies obtaining a solution to the differential game. Under the assumption of EVs' indistinguishability, the state of the game at time t can be modeled as a random variable, x_t, following a distribution $m(t, x)$ corresponding to the limiting distribution of the empirical distribution:

$$m^{(V)}(t, x) = \frac{1}{V} \sum_{v=1}^{V} \delta_{x_t^{(v)}=x}. \tag{27.20}$$

Considering the consumption rate of each EV to be a random variable (with mean $g_t dt$), given by $g_t(dt + \sigma_t dW_t)$ where W_t is a Brownian motion, the state evolution can be represented by a stochastic differential equation as follows:

$$dx_t = \alpha_t dt - g_t(dt + \sigma_t dW_t) + dN_t, \tag{27.21}$$

with initial condition x_0 being a random variable following a distribution $m_0 = m(0, .)$ and where dN_t is a reflective variable which ensures that x_t remains in the range $[0, 1]$.

Under the mean field game formulation, the multiplayer differential game control problem reduces to a single player stochastic control problem with a cost function

$$J(\alpha) = \mathbb{E}\left[\int_0^T \big(\alpha_t p_t(m_t) + h_t(\alpha_t) + f_t(x_t)\big)dt + \kappa(x_T)\right], \tag{27.22}$$

initial condition (x_0, m_0), and state dynamic equation given in (27.21); while $m_t(t, .)$ is the distribution of EVs among the individual states. Here, the expectation operation in (27.22) is performed over the random variable W_t.

The equilibrium of this mean field game proposed in Couillet et al. (2012) is denoted as a mean field equilibrium (MFE) in own-state feedback strategies and is one in which the control strategy α^* satisfies:

$$J(\alpha^*; m^*) \leq J(\alpha; m^*), \tag{27.23}$$

for all admissible control strategies α and where m^* is the distribution induced by α^* following the dynamic equation in (27.21) and initial state distribution m_0.

3.1.1 Game Solution and Main Results

The value function, $v : [0, T] \times [0, 1] \to \mathbb{R}$ for a given m_t is defined as follows:

$$v(u, y) = \inf_{\alpha \in \mathscr{A}} \mathbb{E}\left[\int_u^T \big(\alpha_t p_t(m_t) + h_t(\alpha_t) + f_t(x_t)\big)dt + \kappa(x_T)\right], \tag{27.24}$$

where the dynamics of x_t follow (27.21), $\mathscr{A}$ is the set of admissible controls, and $x_u = y$. As detailed in Couillet et al. (2012), an MFE is a solution of the following backward Hamilton-Jacobi-Bellman (HJB) equation:

$$\begin{aligned}\partial_t v(t, x) = -\inf_{\alpha \in \mathbb{R}} \{\alpha \partial_x v(t, x) + \alpha p_t(m_t^*) + h_t(\alpha_t) + f_t(x_t)\}\\ + g_t \partial_x v(t, x) - \frac{1}{2} g_t^2 \sigma_t^2 \partial_{xx}^2 v(t, x),\end{aligned} \tag{27.25}$$

in which $m_t^* = m(t, .)^*$ is the solution of the forward Fokker-Planck-Kolmogorov (FPK) equation defined as

$$\partial_t m(t, x) = -\partial_x[(\alpha_t^* - g_t)m(t, x)] + \frac{1}{2} g_t^2 \sigma_t^2 \partial_{xx}^2 m(t, x). \tag{27.26}$$

The work in Couillet et al. (2012) solves the HJB and FPK equations using a fixed-point algorithm and presents a numerical analysis over a 3-day period with varying average energy consumption rates and a quadratic price function in total electricity demand (from EVs and other customers). The obtained distribution evolution, m^*, shows sequences of increasing and decreasing battery levels with a tendency to consume energy during daytime and charge during nighttime. In fact, simulation results show that the maximum level of purchased energy by the EVs occur during nighttime when the aggregate electricity demand is low. However, the simulation results indicate that, even though during peak daytime demand electricity prices are higher, the level of purchased energy by the EVs remains significant since the EV owners have an incentive not to completely empty their batteries (via the function $f(t, x)$). Moreover, the obtained results showcase how price incentives can lead to an overall decrease in peak demand consumption and an increase in demand at low consumption periods indicating the potential for EVs to participate in demand-side management mechanisms to improve the consumption profile of the distribution system.

3.2 Microgrids' Cooperative Energy Exchange

A microgrid is an interconnected group of distributed energy sources and local loads at the distribution system's level through which local generation units can meet the electric energy demand of a local geographical area. The analysis and understanding of the operation of microgrids and their integration in the distribution system requires an accurate distributed decision making model. In fact, in a classical power system, the operator optimizes the system operation by solving a system-wide optimization problem based on a centralized objective function. However, due to the formation of autonomous microgrids, each of these microgrids may need to optimize its own objective function which can differ in nature from that of other microgrids. The difference in objectives stems from the versatile nature of the microgrid network which often incorporates different components such as electric vehicles, energy storage, and renewable energy units. As a result, it is imperative to use distributed analytical methods, such as those provided by game theory, in order to optimize and control a distribution system containing a network of microgrids.

One key purpose of microgrids is to relieve the demand on the main distribution system by serving their local geographical areas. However, due to their heavy reliance on intermittent type of generation sources, such as RE, as well as due to the consumers' random demand (especially with the integration of EVs), a microgrid can fall short on meeting its local electric load. As a result, this deficit needs to be met by the distribution grid.

However, in future power systems, it is envisioned that the distribution system will encompass a large number of interconnected microgrids. Hence, due to this interconnection, a microgrid in need for power can obtain power from another interconnected microgrid that has a power surplus instead of obtaining the power from the main grid. One should note, here, that such an energy exchange is contingent upon the introduction of proper energy exchange protocols to alleviate the technical challenges and operational risks that it may introduce. If properly implemented, this energy exchange can be beneficial to participants in both microgrids and can introduce various advantages ranging from achieving reduction in power transmission losses, due to the local exchange of power, to reducing the peak load of the main distribution system.

As a result, there is a need for devising a mathematical mechanism which enables this local energy exchange between microgrids experiencing a lack of energy, referred to as buyers, and microgrids that have a surplus of energy, referred to as sellers. To this end, the work in Saad et al. (2011) has devised such a scheme and showed that cooperative game theory can be used to model cooperative energy exchange mechanisms in the smart grid. The model presented in Saad et al. (2011), a game formulation, and the main results that ensue are introduced next. Here, the focus is on modeling the cooperative behavior between the different microgrids. For modeling the noncooperative behavior between different microgrids competing to meet a local load, the reader is referred to El Rahi et al. (2016).

Consider an interconnection of N microgrids in a distribution network. Each microgrid $n \in \mathcal{N}$, $\mathcal{N}$ referring to the set of all microgrids, has a set of DG units and loads that it serves. Let P_n be the difference between the generation and demand at n. Here, P_n is a random variable due to the random nature of the RE and the load. As a result, at a given time instant, microgrid n can be in one of three states: (1) $P_n > 0$ and n aims at selling this surplus, (2) $P_n < 0$ and n is searching to buy energy from another power source to meet the deficit, or (3) $P_n = 0$ and n does not take part in any energy exchange. To this end, when $P_n < 0$, n is faced with the options of covering the deficit by either buying energy from the distribution system or from another microgrid $n' \in \mathcal{N}$ with $n' \neq n$. When n buys the deficit P_n from the main grid, the incurred total electric losses are denoted by P_{n0}^{loss}, and the total cost of these losses is given by

$$u(\{n\}) = -\omega_n P_{n0}^{\text{loss}}, \tag{27.27}$$

where ω_n is the price per unit of power loss. P_{n0}^{loss} is dependent on many factors including the amount of power exchange, the voltage at which power is transmitted, the associated distribution lines' resistances and distance, as well as losses at corresponding substations. $u(\{n\})$ is hence considered as the utility function of a noncooperative microgrid $n \in \mathcal{N}$.

However, rather than buying power from the main grid, a microgrid can buy power from other microgrids. To this end, microgrids form a cooperative group, referred to as coalition, through which energy can be exchanged between the constituents of this coalition. Such a local exchange of power serves to reduce the incurred power losses mainly due to the short distances involved. Hence, based on their respective geographic locations and power needs, a number of microgrids would have an incentive to cooperate and exchange power locally. In this regard, a coalitional game can be formulated to model this cooperative energy exchange.

To this end, let $S \subseteq \mathcal{N}$ be a coalition between a number of cooperating microgrids. S is hence composed of a set of buyers, S_b, and a set of seller, S_s, such that $S = S_b \cup S_s$. The utility of a coalition S depends on the members of S as well as on the way in which the matching between buyers and sellers is done within S. The optimal matching mechanism within S can be modeled as a double-auction mechanism in analogy with the wholesale energy market model described in Sect. 2. The participants in this auction are the sellers and buyers, while their strategies are, respectively, to specify the amount of power to sell (buy) and the associated price. Thus, the end result of this auction specifies the optimal amount of power that each seller (buyer) is willing to sell (buy) and the equilibrium price (resulting in an optimal matching between the different sellers and buyers). Let Π be such an association between buyers and sellers within S; the utility of S can be expressed as follows:

$$u(S,\Pi) = -\left(\sum_{n \in S_s, n' \in S_b} \omega_n P_{nn'}^{\text{loss}} + \sum_{n \in S_s} \omega_n P_{n0}^{\text{loss}} + \sum_{n' \in S_b} \omega_{n'} P_{n'0}^{\text{loss}} \right), \tag{27.28}$$

where $P^{\text{loss}}_{nn'}$ corresponds to the power loss associated with the local energy exchange between the seller $n \in S_s$ and buyer $n' \in S_b$ while P^{loss}_{n0} and $P^{\text{loss}}_{n'0}$ correspond to the power loss resulting from power transfer, if any, between n or n' and the main grid.[10] The expression of $u(S, \Pi)$ in (27.28), hence, reflects the aggregate power lost through the various power transfers within S. Let $\mathscr{P}_s$ be the set of all possible Π associations between buyers and sellers in S. Accordingly, the value function $v(S)$ of the microgrids' coalitional game, which corresponds to the maximum total utility achieved by any coalition $S \subset \mathscr{N}$ (minimum aggregate power loss within S), is defined as

$$v(S) = \max_{\Pi \in \mathscr{P}_s} u(S, \Pi). \tag{27.29}$$

This value function represents the total cost of power loss incurred by a coalition S which, hence, needs to be split between the members of S. Since this cost can be split in an arbitrary way between the members of S, this game is known to have a transferable utility. The work in Saad et al. (2011) adopts a proportional fair division of the value function. Hence, every possible coalition S has, within, an optimal matching between the buyers and sellers which subsequently leads to an achieved value function and corresponding cost partitions. However, one important question that is yet to be answered is how can the different microgrids form the coalitions? The answer to this question can be obtained through the use of a coalitional formation game framework. Hence, the solution to the proposed energy exchange problem requires solving a coalition formation game and an auction game (or a matching problem).

3.2.1 Game Solution and Main Results

For the energy exchange matching problem within each coalition S, the authors in Saad et al. (2011) propose a sequential approach using which each buyer $b_i \in S_b = \{b_1, \ldots, b_k\}$, following a given order, purchases its needed power in a way that minimizes its incurred losses. This approach can be described as follows. A buyer $b_i \in S_b$ builds a preference relation among sellers S_s based on the amount of power lost during exchange between b_i and $s_j \in S_s$. Following this relation, b_i requests the amount of power it needs from its preferred seller, s_l (which corresponds to the seller yielding the smallest power loss which can be, for example, the geographically closest seller to b_i). If s_l can fully meet the power demand P_i, then b_i takes no further actions. However, if not all of P_i can be obtained from s_l, b_i purchases the highest possible fraction of P_i from s_l and tries to obtain the remaining power from its next preferred seller. This process goes on until b_i meets all of its demand deficit P_i. This approach is considered to be a simple approach through which buyers

[10]Energy exchange with the main grid happens in two cases: (1) if the total demand in S cannot be fully met by the supply in S, leading some buyers to buy their unmet load from the distribution system, or (2) if the total supply exceeds the total demand in S; and hence, the excess supply is sold to the distribution system.

meet their energy needs within the coalition. A more general approach would be to consider buyers and sellers as competing players in an auction mechanism as described previously in this section. An example of competitive behavior modeling between microgrids is presented in El Rahi et al. (2016).

With regard to the coalition formation problem, the authors in Saad et al. (2011) propose a distributed learning algorithm following which microgrids can, in an autonomous manner, cooperate and self-organize into a number of disjoint coalitions. This coalition formation algorithm is based on the rules of merge and split defined as follows. A group of microgrids (or a group of coalitions) would merge to form a larger coalition if this merge increases the utility (reduces the power loss) of at least one of the participating microgrids without decreasing the utilities of any of the others. Similarly, a coalition of microgrids splits into smaller coalitions if this split increases the payoff of one of the participants without reducing the payoff of any of the others.

This coalition formation algorithm and the energy exchange matching heuristic have been applied in Saad et al. (2011) to a case study consisting of a distribution network including a numbers of microgrids. In this respect, it was shown that coalition formation served to reduce the incurred electric losses associated with the energy exchange as compared with the case in which the microgrids decided not to cooperate. In fact, the numerical results have shown that this reduction in losses can reach an average of 31% per microgrid. The numerical results have also shown that the decrease in losses due to cooperation improves when the total number of microgrids increases.

Consequently, the introduced cooperative game-theoretic framework provides the mathematical tools needed to analyze coalition formation between microgrids and its effect on improving the efficiency of the distribution system through achieving a decrease in power losses.

4 Power System Security

The integration of communication and data processing technologies into the traditional power system promotes a more efficient, resilient, and dependable operation of the grid. However, this high interconnectivity and the fundamental reliance of the grid on its underlying communication and computational system render the power system more vulnerable to many types of attacks, of cyber and physical natures, aiming at disrupting its functionality. Even though the power system is designed to preserve certain operational requirements when subjected to external disturbances which are (to some extent) likely to happen, the system cannot naturally withstand coordinated failures which can be orchestrated by a malicious attacker.

To this end, the operator must devise defense strategies to detect, mitigate, and thwart potential attacks using its set of available resources. Meanwhile, an adversary will use its own resources to launch an attack on the system. The goal of this attack can range from reaping financial benefit through manipulation of electricity prices to triggering the collapse of various components of the system leading to a wide-range blackout. However, when devising a defense (attack) strategy, the defender (attacker) has to take into consideration the potential attack (defense) strategies that the attacker (defender) might admit so as to choose the defense (attack) strategy that minimizes (maximizes) the effect of that attack on the system. In this regard, given that the actions of the attackers have direct impact on the objective of the defender and vice versa, game-theoretic tools can be used to capture and analyze this intertwined strategic interaction. Moreover, given that the power system is a dynamic system, dynamic game-theoretic techniques provide a plethora of mathematical tools which can be used in power system security analyses.

The application of game-theoretic techniques to address power system security challenges allows operators to devise defense strategies that can better secure the grid, preserving its integrity and availability. To this end, the devised game-theoretic models should account for the following:

1. *Practicality constraints*: a defender's goal cannot solely be to protect the system against potential attacks. It must also meet power system performance requirements. In fact, to completely secure the system against possible cyber attacks, a complete disconnection of the cyber layer from the physical system might be the best solution. However, even though such a strategy would improve the security of the system, it would deprive it from all the operational advantages introduced by the underlying communication and data processing systems.
2. *Feasibility restrictions*: attackers and defenders both have limited resources that they can use for their attack or defense strategies. Such resources can correspond to monetary resources, skills, time, or computational abilities, among others. Thus, the attackers' and defenders' strategies need to abide by their corresponding resource limitations.

This section focuses on the security threats targeting the observability of the power system as well as its dynamic stability. First, observability attacks on the power system's state estimation are considered while focusing on modeling the strategic behavior of multiple data injection attackers, targeting the state estimator, and a system defender. Second, dynamic attacks targeting the power system are considered. In this regard, first, a general cyber-physical security analysis of the power system is provided based on which a robust and resilient controller is designed to alleviate physical disturbances and mitigate cyber attacks. Then, the focus is shed on the security of a main control scheme of interconnected power systems known as automatic generation control.

4.1 Power System State Estimator Security

The full observability of the state of operation of a power system can be obtained by a complete concurrent knowledge of the voltage magnitudes and voltage phase angles at every bus in the system. These quantities are, in fact, known as the system states, which are estimated using a state estimator. In this regard, distributed measurement units are spread across the power system to take various measurements such as power flows across transmission lines, power injections and voltage magnitudes at various buses, and, by the introduction of phasor measurement units (PMUs), synchronized voltage and current phase angles. These measurements are sent via communication links and fed to a state estimator which generates a real-time estimate of the system states. Hence, the observability of the power system directly relies on the integrity of the collected data. Various operational decisions are based on the monitoring ability provided by the state estimator. Thus, any perturbation to the state estimation process can cause misled decisions by the operator, whose effect can range from incorrect electricity pricing to false operational decisions which can destabilize the system. In this regard, data injection attacks have emerged as a highly malicious type of integrity attacks using which malicious adversaries compromise measurement units and send false data aiming at altering the state estimator's estimate of the real-time system states.

This subsection focuses on the problem of data injection attacks targeting the electric energy market. Using such attacks, the adversary aims at perturbing the estimation of the real-time state of operation leading to incorrect electricity pricing which benefits the adversary. First, the state estimation process is explained followed by the data injection attack model. Then, based on Sanjab and Saad (2016a), the effect of such attacks on the energy markets and electricity pricing is explored followed by a modeling of the strategic behavior of a number, M, of data injection attackers and a system defender.

In a linear power system state estimation model, measurements of the real power flows and real power injections are used to estimate the voltage phase angles at all the buses in the system.[11] The linear state estimator is based on what is known as a MW-only power flow model which was defined along with its underlying assumptions in Sect. 2. The power P_{ij} flowing from bus i to bus j over a transmission line of reactance X_{ij} is expressed in (27.6), where θ_k corresponds to the voltage phase angle at bus k; while the power injection at a certain bus i is as expressed in (27.7). Given that in the MW-only model the voltage magnitudes are assumed to be fixed, the states of the system that need to be estimated are the voltage phase angles. The power injections and power flows collected from the system are linearly related to the system states, following from (27.6) and (27.7). Thus, the measurement vector, $\mathbf{z}$, can be linearly expressed in terms of the vector of system states, $\boldsymbol{\theta}$, as follows:

$$\mathbf{z} = \mathbf{H}\boldsymbol{\theta} + \mathbf{e}, \tag{27.30}$$

where $\mathbf{H}$ can be built based on (27.6) and (27.7) and $\mathbf{e}$ is the vector of random measurement errors assumed to follow a normal distribution, $\mathrm{N}(0, \mathbf{R})$. Using a weighted least squares estimator, the estimated system states are given by

[11] Except for the reference bus whose phase angle is the reference angle and is hence assumed to be equal to 0 rad.

$$\hat{\theta} = (\mathbf{H}^T\mathbf{R}^{-1}\mathbf{H})^{-1}\mathbf{H}^T\mathbf{R}^{-1}\mathbf{z}. \tag{27.31}$$

The estimated measurements, $\hat{\mathbf{z}}$, and what is known as the measurement residuals, $\mathbf{r}$, can be computed as follows:

$$\hat{z} = \mathbf{H}\hat{\theta} = \mathbf{S}\mathbf{z}, \quad \mathbf{r} = \mathbf{z} - \hat{z} = (\mathbf{I}_\eta - \mathbf{S})\mathbf{z} = \mathbf{W}\mathbf{z}, \tag{27.32}$$

where I_η is the identity matrix of size $(\eta \times \eta)$ and η is the total number of collected measurements.

When additive data injection attacks are launched by M attackers in the set $\mathscr{M} = \{1, \ldots, M\}$, the measurements are altered via the addition of their attack vectors $\{\mathbf{z}^{(1)}, \mathbf{z}^{(2)}, \ldots, \mathbf{z}^{(M)}\}$. Hence, this leads to the following measurements and residuals:

$$\mathbf{z}^{\text{att}} = \mathbf{z} + \sum_{m=1}^{M} \mathbf{z}^{(m)}, \quad \mathbf{r}^{\text{att}} = \mathbf{r} + \mathbf{W}\sum_{m=1}^{M} \mathbf{z}^{(m)}. \tag{27.33}$$

Typically, identification of outliers is based on the calculated measurement residuals. In this respect, $\mathbf{W}\sum_{i=1}^{M} \mathbf{z}^{(m)}$ must be chosen to keep the magnitude of the residuals, $||\mathbf{r}^{\text{att}}||_2^2 = \sum_{i=1}^{\eta}(r_i^{\text{att}})^2$, relatively small to minimize the chance of detection of the attack.

To model the effect of data injection on an energy market, consider a competitive wholesale electric energy market architecture based on day-ahead and real-time markets. In the DA market, hourly DA LMPs, μ^{DA}, are issued by the operator for the next operating day based on the generation offers of the participating GENCOs and the solution of an OPF problem as discussed in Sect. 2 in (27.2), (27.3), (27.4), (27.5) and (27.14). However, electricity pricing is not only based on DA forecasts but also on the real-time state of operation of the system. Thus, in real-time, using the state estimator output, the real-time behavior of all market participants and system states is estimated. This estimation outcome is fed to an OPF problem to generate the real-time LMPs, μ^{RT}. An illustration of the DA-RT market architecture is presented in Fig. 27.3. Some market participants in the DA and RT markets do not possess any generation units nor serve any loads. Such participants are known as virtual bidders (VBs). Virtual bidding is a protocol using which VBs submit what is known as virtual supply offers and demand bids. However, since the VBs do not possess any physical means to meet their generation offers or demand bids, a VB that buys (sells) virtual power at a specific bus in DA is required to sell (buy) that same power at the same bus in RT. In this respect, VBs can make a profit from possible mismatch between DA and RT LMPs.

As such, using data injection attacks, a VB can successfully manipulate system measurements to change the real-time LMPs and intentionally create a beneficial mismatch between μ^{DA} and μ^{RT} at the buses over which it has placed virtual bids. On the other hand, to protect the system, the operator aims at securing some

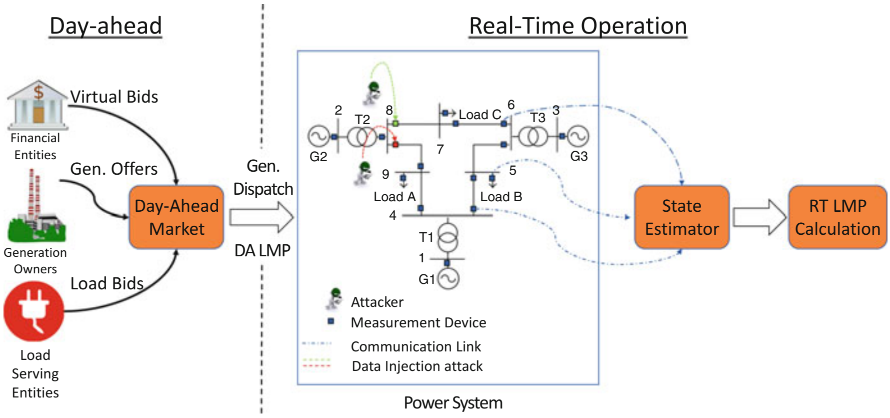


Fig. 27.3 Illustration of energy market operation based on DA and RT markets

measurement units such that they become immune to such attacks. Securing a measurement unit can be done by, for example, a disconnection from the Internet, encryption techniques, or replacing the unit by a more secure one. However, such security measures are preventive aiming at thwarting potential future attacks. To this end, a skilled attacker can have the ability to observe which measurements have been secured. In fact, installation of new measurement units can be physically noticeable by the attackers, while encrypted sensors' outputs can be observed by an adversary attempting to read the transmitted data. Hence, the strategic interaction between the defender and the M attackers is hierarchical in which the defender (i.e. the leader) chooses a set of measurements to protect while the attackers (i.e. the followers) observe which measurements have been protected and react, accordingly, by launching their optimal attack. This single leader multi-follower hierarchical security model presented in Sanjab and Saad (2016a) is detailed next.

Starting with the attackers (i.e. the followers game), consider a VB, m, which is also a data injection attacker and which places its virtual bids as follows. In the DA market, VB m buys and sells P_m MW of virtual power at buses i_m and j_m, respectively. In the RT market, VB m sells and buys the same amount of virtual power P_m MW at, respectively, buses i_m and j_m. As a result, the payoff, ζ_m, of VB m earned from this virtual bidding process can be expressed as follows:

$$\zeta_m = \left[\left(\mu_{i_m}^{RT} - \mu_{i_m}^{DA}\right) + \left(\mu_{j_m}^{DA} - \mu_{j_m}^{RT}\right)\right] P_m. \tag{27.34}$$

As a result, attacker m aims at choosing a data injection attack vector $\mathbf{z}^{(m)}$ from the possible set of attacks $\mathscr{Z}^{(m)}$ which solves the following optimization problem:

$$\max_{\mathbf{z}^{(m)} \in \mathscr{Z}^{(m)}} U_m(\mathbf{z}^{(m)}, \mathbf{z}^{-(m)}) = \zeta_m - c_m(\mathbf{z}^{(m)}). \tag{27.35}$$

subject to:

$$\|\mathbf{W}\mathbf{z}^{(m)}\|_2 + \sum_{l=1, l\neq m}^{M} \|\mathbf{W}\mathbf{z}^{(l)}\|_2 \leq \epsilon_m, \tag{27.36}$$

where $c_m(\mathbf{z}^{(m)})$ is the cost of launching attack $\mathbf{z}^{(m)}$ and $\mathbf{z}^{-(m)}$ corresponds to the attack vectors of all other attackers except m as well as on the defense vector (to be defined next). $\mathscr{Z}^{(m)}$ rules out the measurements that have been previously secured by the defender and hence is dependent on the defender's strategy. The dependence of U_m on the attacks launched by the other attackers as well as on the defense vector is obvious since these attack and defense strategies will affect the LMPs in RT. The limit on the residuals of the attacked measurements expressed in (27.36) is introduced to minimize the risk of being identified as outliers, where ϵ_m is a threshold value specified by m. As such, in response to a defense strategy by the operator, the attackers play a noncooperative game based on which each one chooses its optimal attack vector. However, since the attackers' constraints are coupled, their game formulation is known as a generalized Nash equilibrium problem (GNEP). The equilibrium concept of such games known as the generalized Nash equilibrium (GNE) is similar to the NE; however, it requires that the collection of the equilibrium strategies of all the attackers not violate the common constraints.

The system operator, on the other hand, chooses a defense vector $\mathbf{a}^{(0)} \in \mathscr{A}^{(0)}$ indicating which measurement units are to be secured and which takes into consideration the possible attack strategies that can be implemented by the attackers in response to its defense policy. The objective of the defender is to minimize the potential mismatch in LMPs between DA and RT which can be caused by data injection attacks. Thus, the defender's objective is as follows:

$$\min_{\mathbf{a}_0 \in \mathscr{A}_0} U_0(\mathbf{a}_0, \mathbf{a}_{-0}) = P_L \sqrt{\frac{1}{N} \sum_{i=1}^{N} (\mu_i^{RT} - \mu_i^{DA})^2} + c_0(\mathbf{a}_0), \tag{27.37}$$

subject to:

$$\|\mathbf{a}_0\|_0 \leq B_0, \tag{27.38}$$

where $c_0(\mathbf{a}_0)$ constitutes the cost of the implemented defense strategy, P_L is the total electric load of the system, N is the number of buses, and B_0 corresponds to the limit on the number of measurements that the defender can defend concurrently.[12]

[12]This feasibility constraint in (27.38) insures the implementability and practicality of the derived defense solutions. To this end, a defender with more available defense resources may be more likely to thwart potential attacks. For a game-theoretic modeling of the effects of the level of resources, skills, and computational abilities that the defenders and adversaries have on their optimal defense and attack policies in a power system setting, see Sanjab and Saad (2016b).

$\mathbf{a}_{-0}$ corresponds to the collection of attack vectors chosen by the M attackers, i.e., $\mathbf{a}_{-0} = [\mathbf{z}^{(1)}, \mathbf{z}^{(2)}, \ldots, \mathbf{z}^{(M)}]$.

Let $\mathscr{R}^{\text{att}}(\mathbf{a}_0)$ be the set of optimal responses of the attackers to a defense strategy $\mathbf{a}_0$. Namely, $\mathscr{R}^{\text{att}}(\mathbf{a}_0)$ is the set of GNEs of the attackers' game defined in (27.35) and (27.36). As such, the hierarchical equilibrium of the defender-attackers' game is one that satisfies:

$$\max_{\mathbf{a}_{-0} \in \mathscr{R}^{\text{att}}(\mathbf{a}_0^*)} U_0(\mathbf{a}_0^*, \mathbf{a}_{-0}) = \min_{\mathbf{a}_0 \in \mathscr{A}_0} \max_{\mathbf{a}_{-0} \in \mathscr{R}^{\text{att}}(\mathbf{a}^0)} U_0(\mathbf{a}_0, \mathbf{a}_{-0}). \tag{27.39}$$

4.1.1 Game Solution and Main Results

By transforming each attacker's problem in (27.35) and (27.36) to a convex optimization problem, the work in Sanjab and Saad (2016a) has proven the existence of at least one GNE for the attackers' game which can be obtained through a proposed solution algorithm. Moreover, the case in which the defender has limited information about the attackers has been treated using the concept of satisfaction equilibrium. In this regard, given that the defender may not be able to anticipate the reaction of the attackers, rather than aiming at minimizing the mismatch between DA and RT LMPs, the defender aims at finding a defense strategy $\mathbf{a}_0$ which keeps this mismatch within a certain bound. Thus, the defender aims at meeting a certain performance requirement. As in the hierarchical framework, the attackers respond to the defense strategy by playing a GNEP. The authors in Sanjab and Saad (2016a) introduced an equilibrium concept for this hybrid satisfaction equilibrium-generalized Nash equilibrium problem (denoted as hybrid hierarchical equilibrium) which is a state of the game in which the defender plays a strategy that meets its performance requirement and the attackers respond by playing a GNE.

Simulation results over a 30-bus test system in Sanjab and Saad (2016a) showed that the competitive behavior of multiple attackers can alleviate the aggregate effect of the attacks. Moreover, the results showed that protecting a small subset of measurements can significantly mitigate the effect of the attacks on the system.

This subsection has investigated observability attacks on the power system. In the next subsection, dynamic system attacks are studied.

4.2 Dynamic System Attacks on the Power System

Modern power systems are cyber-physical systems formed of an interconnection between a physical dynamic system, comprising physical components (such as synchronous generators and transformers) and the underlying control systems, and a cyber system comprising the communication, data processing, and network layers. Such a multilayered system is subject to a set of externalities ranging from disturbances to the physical system to cyber attacks. Thus, preserving the security of the power system requires an accurate multilayered cyber-physical modeling of the system with the goal to attenuate and reject physical disturbances as well as mitigate and thwart cyber attacks.

4.2.1 Cyber-Physical Security of the Power System

The power network's physical dynamic system is continuously subject to exogenous disturbances which affect the evolution of its states. As such, robust control policies are typically implemented to attenuate the effects of such disturbances. Moreover, the interconnection of the physical dynamic system to a cyber – information and communication – layer puts the system in danger of cyber attacks which can occur at random times and cause sudden changes to the state of operation of the system. Thus, defense policies should be implemented to enhance the resilience of the power system against such attacks.

To this end, a multilayered cyber-physical security model was proposed in Zhu and Başar (2011) using a layered stochastic-differential dynamic game. In this model, the physical system dynamics are modeled to be continuous in time and governed by controller inputs and external disturbances. The disturbances/uncertainties over the physical system are modeled as continuous and deterministic. A robust optimal control is designed, using a zero-sum differential game, to achieve robustness of the physical system facing such disturbances. In addition, cyber attacks are modeled to induce sudden changes to the system's structural states. As such, the state space of the cyber system is modeled to be discrete, and the transition between these states is modeled to be stochastic and dependent on the attack and defense strategies. The conflicting strategic behavior of the attacker and defender is modeled using a zero-sum stochastic game enabling the derivation of a defense strategy to mitigate the effect of the attacks. Such an intertwined security model allows the derivation of optimal control and security strategies to mitigate the effect of externalities and cyber attacks targeting the power system. The work in Zhu and Başar (2011) in this domain is detailed next; for a more detailed exposure to such dynamic games within the context of cyber-physical system security, see Zhu and Başar (2015).

Consider $x(t) \in \mathbb{R}^n$ to be the state of the physical system which evolves from the initial condition $x_0 = x(t_0)$ based on the state dynamic equation:

$$\dot{x}(t) = h(t, x, u, w; s(t, \alpha, \delta)), \tag{27.40}$$

where $u \in \mathbb{R}^r$ is the control input, $w \in \mathbb{R}^p$ are disturbances to the physical system, and s is the state of the cyber system (i.e. the structural states of the system) belonging to the state space $\mathscr{S} = \{1, 2, \ldots, S\}$. α and δ are attack and defense actions which will be explored in greater details next. The structural states denote the security state of the system. For example, a certain structural state s_i can denote that a transmission line has been disconnected due to a certain cyber attack.

The transition between the discrete structural states of the system can be represented using a Markov process with initial distribution π_0 and rate matrix $\Lambda = \{\lambda_{i,j}\}_{i,j \in \mathscr{S}}$. The transition between structural states depends on the implemented defense action $\delta \in \mathscr{D}$ and the attack action $\alpha \in \mathscr{A}$ carried out by the attacker.

Note, here, that the transitions between structural states happen on a different time scale than the evolution of the physical states. In fact, the physical states evolve in the order of seconds, while the time scale of cyber attacks is typically in the order of days. As such, at a time k at which a cyber event occurs, the physical dynamic system can be assumed to be in steady state. This is referred to as "time-scale separation."

An attack can lead to a sudden transition in the state of operation of the system, while a defense strategy can bring the system back to its nominal operating state. Consider the set of mixed defense and attack strategies to be, respectively, denoted by $\mathscr{D}_k^m = \{\mathbf{d}(k) \in [0,1]^{|\mathscr{D}|}\}$ and $\mathscr{A}_k^m = \{\mathbf{a}(k) \in [0,1]^{|\mathscr{A}|}\}$. The transition between the system states for a mixed defense strategy $\mathbf{d}(k)$ and mixed attack strategy $\mathbf{a}(k)$ can be modeled as follows:

$$\text{Prob}\{s(k+\Delta) = j | s(k) = i\} = \begin{cases} \lambda_{i,j}(\mathbf{d}(k), \mathbf{a}(k)), & j \neq i \\ \lambda_{i,i}(\mathbf{d}(k), \mathbf{a}(k)), & j = i, \end{cases} \tag{27.41}$$

where Δ is a time increment on the time scale of k. Denoting $\hat{\lambda}_{i,j}(\delta(k), \alpha(k))$ to be the transition rate between $i \in \mathscr{S}$ and $j \in \mathscr{S}$ under pure defense action $\delta(k)$ and attack action $\alpha(k)$, then $\lambda_{i,j}$ in (27.41) corresponds to the average transition rates between i and j and can be calculated as follows:

$$\lambda_{i,j}(\mathbf{d}(k), \mathbf{a}(k)) = \sum_{\delta \in \mathscr{D}} \sum_{\alpha \in \mathscr{A}} d_\delta(k) a_\alpha(k) \hat{\lambda}_{i,j}(\delta(k), \alpha(k)). \tag{27.42}$$

As such, (27.40) and (27.41) present a hybrid discrete-continuous security model of the power system. The goal is to design an optimal control strategy to mitigate the effect of physical disturbances and characterize a dynamic defense strategy to mitigate the effect of cyber attacks on the system. Due to the time-scale separation between the physical and cyber layers, the control and defense strategies can be studied separately but while considering the interdependence between the two.

For the optimal controller design, let $s_{[t_0,t]}$ be defined as $s_{[t_0,t]} := \{s(\tau), \tau \leq t\}$. Consider that the controller has access to $x_{[t_0,t]}$ and $s_{[t_0,t]}$ at time t. As such, denote the perfect state closed-loop control strategy by $u(t) = \mu(t, x_{[t_0,t]}; s_{[t_0,t]})$ in the class of all admissible control strategies $\mathscr{M}_{\text{CL}}$ and closed-loop disturbance by $v(t) = \nu(t, x_{[t_0,t]}; s_{[t_0,t]})$ in the set of all admissible disturbances $\mathscr{N}_{\text{CL}}$. In this respect, a cost function over $t \in [t_0, t_f]$ to be minimized by the controller is one of the form:

$$L(x, u, w; s) = q_0(x_0; s(t_0)) + q_f(x(t_f); s(t_f)) + \int_{t_0}^{t_f} g(t, x(t), u(t), w(t); s(t))dt, \tag{27.43}$$

where $g(.)$, $q_0(.)$, and $q_f(.)$ define the cost structure.

Based on this cost function, the performance index of the controller can be defined as

$$J(u, v) = \mathbb{E}_s\{L(x, u, w; \theta)\}. \tag{27.44}$$

The goal of the optimal controller is to devise a minimax closed-loop control strategy $\mu_{\text{CL}}^* \in \mathscr{M}_{\text{CL}}$ which achieves

$$\sup_{\nu \in \mathscr{N}_{\mathrm{CL}}} J(\mu^*_{\mathrm{CL}}, \nu) = \inf_{\mu \in \mathscr{M}_{\mathrm{CL}}} \sup_{\nu \in \mathscr{N}_{\mathrm{CL}}} J(\mu, \nu). \tag{27.45}$$

The expression in (27.45) along with the dynamics in (27.40) define a zero-sum differential game. A solution of the differential game will lead to an optimal control minimizing the effect of disturbances on the physical system. Next, characterizing the optimal defense policy against cyber attacks is investigated.

The interaction between an attacker and a defender can be modeled as a zero-sum stochastic game in which the discounted payoff criterion, $V_\beta(i, \mathbf{d}, \mathbf{a})$, with discount factor β is given by

$$V_\beta(i, \mathbf{d}(k), \mathbf{a}(k)) = \int_0^\infty e^{-\beta k} \mathbb{E}_i^{\mathbf{d}(k),\mathbf{a}(k)}[V^i(k, \mathbf{d}(k), \mathbf{a}(k))]dk, \tag{27.46}$$

where $V^i(k, \mathbf{d}(k), \mathbf{a}(k))$ is the value function of the zero-sum differential game at state $s = i$ with a starting time at k in the cost function in (27.43). In this game, the defender's goal is to minimize (27.46) while the goal of the attacker is to maximize it.

Let $\mathbf{d}_i^m \in \mathscr{D}_i$ and $\mathbf{a}_i^m \in \mathscr{A}_i$ be a class of mixed stationary strategies dependent on the current structural state $i \in \mathscr{S}$. The goal is to find a pair of stationary strategies $(\mathbf{D}^* = \{\mathbf{d}^i : i \in \mathscr{S}\}, \mathbf{A}^* = \{\mathbf{a}^i : i \in \mathscr{S}\})$ constituting a saddle point of the zero-sum stochastic game, i.e., satisfying:

$$V_\beta(\mathbf{D}^*, \mathbf{A}) \le V_\beta(\mathbf{D}^*, \mathbf{A}^*) \le V_\beta(\mathbf{D}, \mathbf{A}^*). \tag{27.47}$$

Game Solution and Main Results

A robust and resilient control of the cyber-physical power system described in (27.40) can be described as an intertwined set of optimal control policies, denoted by $\{(\mathbf{D}^*, \mathbf{A}^*), (\mu^*_{\mathrm{CL}}, \nu^*_{\mathrm{CL}})\}$, generating an equilibrium of the interconnected differential and stochastic games. As detailed in Zhu and Başar (2011), the optimal defense strategy $(\mathbf{D}^*, \mathbf{A}^*)$, which can be found using a value iteration algorithm described in Zhu and Başar (2011), should satisfy the following fixed point equation for all $i \in \mathscr{S}$:

$$\beta v^*_\beta(i) = V^i(\mathbf{D}^*, \mathbf{A}^*) + \sum_{j \in \mathscr{S}} \lambda_{i,j}(\mathbf{D}^*, \mathbf{A}^*) v^*_\beta(j), \tag{27.48}$$

while the optimal control policy $(\mu^*_{\mathrm{CL}}, \nu^*_{\mathrm{CL}})$ should satisfy the Hamilton-Jacobi-Isaacs equation given by

$$\begin{aligned}
&-V^i_t(t,x) = \\
&\inf_{u \in \mathbb{R}^r} \sup_{w \in \mathbb{R}^p} \left[V^i_x(t,x) f(t,x,u,w,i) + g(t,x,u,w,i) + \sum_{j \in \mathscr{S}} \lambda_{i,j} V^j(t,x) \right], \\
&V^i(t_f, x) = q_f(x(t_f); i), \text{for } i \in \mathscr{S}.
\end{aligned} \tag{27.49}$$

As a case analysis, the work in Zhu and Başar (2011) has studied the case of voltage regulation of a single generator. The simulation results first show the evolution of the angular frequency of the generator and its stabilization under normal operation and then the ability of the controller to bring back stable nominal system operation after the occurrence of a sudden failure. This highlights the robustness and resilience of the developed controller.

The presented security model in this subsection is one of a general cyber-physical power system. Next, the focus is on a specific control mechanism of interconnected power systems, namely, automatic generation control.

4.2.2 Automatic Generation Control Security

The power system is a dynamic system which operates at a fixed nominal frequency, i.e., synchronous frequency corresponding to 50 Hz in most parts of the world; except for North and South America, where the operating frequency in most countries is 60 Hz. To this end, control designs are applied to the power system to damp any frequency deviations and maintain this nominal frequency.

The rotor dynamics of every synchronous generator is affected by the electric load connected to it. In fact, in a synchronous generator, the shaft connecting the turbine to the synchronous machine's rotor is subject to a mechanical torque, T_m, generated by the turbine rotation and an electric torque, T_e, generated by the connected electric load. Based on Newton's second law, the acceleration of the machine is expressed as

$$I\frac{d\omega}{dt} = T_m - T_e, \tag{27.50}$$

where ω is the angular frequency of the rotor and I is the moment of inertia of the rotating mass. Thus, as can be seen from (27.50), for a fixed mechanical torque, an increase in the electric load leads to a deceleration of the machine's rotor while a decrease in the electric load leads to the acceleration of the rotor. Transforming the torques in (27.50) into powers generates what is known as the swing equation given by

$$M\frac{d\Delta\omega}{dt} = \Delta P_m - \Delta P_e, \tag{27.51}$$

where ΔP_m and ΔP_e correspond to a change to the mechanical and electric power, respectively, while M is a constant known as the inertia constant at synchronous speed.[13]

Hence, to maintain the change in angular frequency $\Delta\omega$ close to 0, a control design should constantly change the mechanical power of the machine to match changes in the electric loads of the system. Such a frequency control design typically follows a three-layer scheme. The first layer is known as the primary control layer. The primary control layer corresponds to a local proportional controller connected

[13] All the quantities in (27.51) are expressed in per unit based on the synchronous machine's rated complex power.

to the synchronous machine. The primary proportional controller can reduce the frequency deviation due to a change in load; however, it can leave a steady-state error. The elimination of this steady-state error is achieved by an integrator. In contrast to the local proportional control, the integral control is central to an area. This central integral control corresponds to the second control layer known as the secondary control. The third control layer is known as the tertiary control and is a supervisory control layer insuring the availability of spinning reserves (which are needed by the primary control) and the optimal dispatch of the units taking part in the secondary control.

Moreover, the power grid is composed of the interconnection of many systems. These various systems are interconnected via tie lines. A drop in the frequency of one system triggers a deviation in the power flow over the tie lines from its scheduled power. The change in the power flow of the tie line, ΔP_{TL}, can be expressed in the Laplace domain in terms of the frequency deviations in the two interconnected systems i and j as follows:

$$\Delta P_{TL}(s) = \frac{\omega_s}{sX_{TL}}(\Delta\omega_i(s) - \Delta\omega_j(s)), \tag{27.52}$$

where ω_s is the synchronous angular frequency and X_{TL} is the reactance of the tie line. To minimize such deviation in the tie line power flow, generation control in the two interconnected areas is needed. This control is known as the load-frequency control. This load-frequency control can be achieved through the combination of the primary and secondary controls. The automatic generation control (AGC) constitutes an optimized load-frequency control which is combined with economic dispatch based on which the variations of the generators' outputs dictated by the load-frequency controller follow a minimal cost law.

When these control designs fail to control the deviation from the nominal frequency, frequency relays take protection measures to prevent the loss of synchronism in the system, which can be caused by the large frequency deviations. In this respect, if the frequency deviation is measured to have positively surpassed a given threshold, over-frequency relays trigger the disconnection of generation units to reduce the frequency. Similarly, if the frequency deviation is measured to have negatively surpassed a given threshold, under-frequency relays trigger a load shedding mechanism to decrease the electric load, bringing back the frequency to its nominal value and preserving the safety of the system. However, these frequency relays can be subject to cyber attacks which can compromise them and feed them false data with the goal of triggering an unnecessary disconnection of generation or load shedding. This problem has been addressed in Law et al. (2015).

For example, a data injection attacker can target a frequency relay by manipulating its measured true frequency deviation, Δf, by multiplying it by a factor k such that the frequency relay's reading is $k\Delta f$. Such an attack is known as an overcompensation attack since the integral controller overcompensates this sensed false frequency deviation leading to unstable oscillations overpassing the frequency relay's under-frequency or over-frequency thresholds which triggers load shedding or generation disconnection. When the multiplication factor is negative, this attack is known as a negative compensation attack. Other types of frequency sensing manipulation attacks are constant injection and bias injection. Constant injection continuously feeds the frequency deviation measurement unit a constant mismatch Δf. Bias injection is an additive attack through which a constant additive deviation c is added to the real Δf so that the manipulated measurement is $\Delta_f + c$. Both of these attacks perturb the functionality of the integral controller and leads the system to converge to a frequency different from the nominal one. If this frequency deviation surpasses the frequency relays' thresholds, it leads to shedding of a fraction of the load or disconnection of generation units.

On the other hand, the defender can take defensive actions to mitigate and detect such attacks. To mitigate the effect of such attacks, two measures can be taken. First, saturation filters can be employed to limit the frequency deviation at the input of the integral controller. The cutoff frequencies of such saturation bandpass filters are, respectively, above the over-frequency threshold and below the under-frequency threshold. Thus, the frequency relays can still act to protect the system; however, large deviations do not get fed into the integral controller. Second, the use of redundancy can lead to a decrease in the probability of success of a data injection attack since, clearly, the probability of successfully attacking multiple frequency relays concurrently is smaller than the likelihood of successfully attacking only one of them. Using redundancy, the controller can in an alternating manner take a frequency reading sample from a different measuring unit at each time sample to reduce the risk of issuing a control action based on manipulated readings. In addition to mitigating an attack's effect on the system, detection techniques are essential to detect and eliminate the presence of attacks. In this regard, a detection algorithm can compare successive frequency deviation readings to sense the presence of an attack. Moreover, a clustering-based framework can also be used for data injection detection since, normally, frequency deviations should be clustered around 0. Hence, the presence of more than one cluster can be an indication of the presence of an attack.

The loss to the system, i.e. the loss to the defender which is equivalent to the gain of the attacker, due to such attacks on under-frequency relays can be quantified by the amount of load that has been shed due to that attack, which is denoted by P_{shed}. Various cost of load shed assessment techniques can be adopted. In this regard, Law et al. (2015) considers two approaches; one that quantifies the loss as the expected value of the load shed, $\mathbb{E}[P_{\text{shed}}]$, and the other that quantifies the loss using the risk assessment concept known as the conditional value at risk (CVaR). The value at risk (VaR), with significance level denoted by $0 < \gamma < 1$, corresponds to the minimum value of a loss limit ζ such that the probability that the incurred loss is greater than ζ is at most equal to γ. Hence, considering $F_{P_{\text{shed}}}$ to be the cumulative density function of P_{shed} that is assumed to be smooth and continuous, the VaR and CVaR are defined, respectively, as follows:

$$\text{VaR}_\gamma(P_{\text{shed}}) = \inf\{\zeta \mid F_{P_{\text{shed}}}(\zeta) \geq 1-\gamma\} = F^{-1}_{P_{\text{shed}}}(1-\gamma), \quad (27.53)$$

$$\text{CVaR}_\gamma(P_{\text{shed}}) = \mathbb{E}[P_{\text{shed}}|P_{\text{shed}} \geq \text{VaR}_\gamma(P_{\text{shed}})] = \frac{1}{\gamma}\int_{1-\gamma}^{1} F^{-1}_{P_{\text{shed}}}(\beta)d\beta. \quad (27.54)$$

Thus, the attacker and the defender choose their optimal attack and defense strategies aiming at, respectively, maximizing and minimizing the incurred cost of load shed. Due to the coupling in actions and payoffs of the attacker and the defender, their strategic interaction can be modeled using the tools of game theory. However, the game model needs to consider the dynamic evolution of the system affected by the actions taken by the attacker and the defender. To this end, Law et al. (2015) proposes the use of the framework of a stochastic game. This stochastic game is introduced next.

The authors in Law et al. (2015) consider under-frequency load shedding in two interconnected areas 1 and 2 where a load is shed in area i if the frequency deviation Δf_i in i is such that $\Delta f_i \leq -0.35\,\text{Hz}$. The frequency deviation in area i is measured using frequency sensor i. The work in Law et al. (2015) defines the stochastic game between the defender d and attacker a, Ξ, as a 7-tuple $\Xi = \langle \mathscr{I}, \mathscr{X}, \mathscr{S}^d, \mathscr{S}^a, \mathbf{M}, U^d, U^a \rangle$. $\mathscr{I} = \{d, a\}$ is the set of players. $\mathscr{X} = \{x_{00}, x_{01}, x_{10}, x_{11}\}$ is the set of states reflecting the frequency deviations at the two areas. A state $x \in \mathscr{X}$ can be defined as follows:

$$x = \begin{cases} x_{00} \text{ if } \Delta f_1 > -0.35 \text{ and } \Delta f_2 > -0.35, \\ x_{01} \text{ if } \Delta f_1 > -0.35 \text{ and } \Delta f_2 \leq -0.35, \\ x_{10} \text{ if } \Delta f_1 \leq -0.35 \text{ and } \Delta f_2 > -0.35, \\ x_{11} \text{ if } \Delta f_1 \leq -0.35 \text{ and } \Delta f_2 \leq -0.35. \end{cases}$$

$\mathscr{S}^d$ and $\mathscr{S}^a$ are the set of actions available to the attacker and defender, respectively, and are defined as follows. As a redundancy measure, the defender implements two frequency measuring devises and reads n consecutive samples from each. Each n consecutive samples constitute a session. Following the reading of these n samples, a clustering algorithm is run to detect attacks. The defender's action space reflects how sharp or fuzzy the used clusters are which is dictated by the size of the cross-correlation filter. In this regard, $\mathscr{S}^d = \{s_1^d, s_2^d\}$ where s_1^d sets the filter size to $n/4$ and s_2^d sets the filter size to $2n/5$. In both cases, the frequency measuring unit is considered under attack if at least two clusters are formed. In that case, the frequency sensor is disinfected through a firmware update which is assumed to take one session. When the attacker compromises a meter, it chooses between two strategies: (i) s_1^a which is to overcompensate half of the observed samples or (ii) s_2^a which corresponds to overcompensating all of the observed samples. Thus, $\mathscr{S}^a = \{s_1^a, s_2^a\}$. The attacker is assumed to require four sessions to compromise meter 1 and eight sessions to compromise meter 2. When the meter disinfection is done by the defender, this meter is again subject to attack when it is back in operation.

The transition matrix $\mathbf{M}(s^d, s^a) = [m_{x_i,x_j}(s^d, s^a)]_{(4\times4)}$, i.e. the probability of transitioning from state x_i to state x_j under attack s^a and defense s^d, can be inferred from previous collected samples and observed transitions.

The payoff of the defender, U^d, reflects the cost of the expected load shed and the expected cost of false positives. A false positive corresponds to falsely identifying a meter to be under attack and disinfecting it when it actually is not compromised. The expected cost of false positives can be calculated as $c_{\text{fp}} \times p_{\text{fp}}$ where c_{fp} is the cost associated with a false positive and p_{fp} is the probability of occurrence of this false positive. Thus, U^d, at time t, can be defined in two ways depending on the expected loss model used as follows:

$$U^d_{[\text{E}]}(s^d, s^a, x; t) = -\mathbb{E}[P_{\text{shed}}(s^d, s^a, x; t)] - c_{\text{fp}} p_{\text{fp}}(s^d, s^a, x; t), \tag{27.55}$$

$$U^d_{[\text{CVaR}]}(s^d, s^a, x; t) = -\text{CVaR}_\gamma(P_{\text{shed}}(s^d, s^a, x; t)) - c_{\text{fp}} p_{\text{fp}}(s^d, s^a, x; t). \tag{27.56}$$

This model assumes that the loss to the defender is a gain to the attacker. Hence, the stochastic game is a two-player zero-sum game, and $U^a_{[y]} = -U^d_{[y]}$.

Under this game framework, the objective of the defender (attacker) is to minimize (maximize) the discounted cost C, with discount factor β, over all time periods:

$$C = \sum_{t=0}^{\infty} \beta^t U^a(s^d, s^a, x; t). \tag{27.57}$$

Game Solution and Main Results

For the solution of the game, the Nash equilibrium solution concept is adopted. In this regard, since the considered stochastic game is a two-player zero-sum discounted game, it is proven to have a unique NE in stationary strategies. A stationary strategy is a strategy that is independent of time t. As a result, the game in Law et al. (2015) is solved using dynamic programming techniques by recursively solving a matrix game at each stage.

The authors in Law et al. (2015) provide simulation results considering a two-area AGC with a session size of 20 samples ($n = 20$), a simulation window of 200 minutes, and a sapling rate of 1 Hz. The results obtained show that both defense strategies s^d_1 and s^d_2 achieve a 100% rate of detection against the attacks s^a_1 and s^a_2, while s^d_2 shows a higher rate of false positives. The results obtained also aimed at comparing the use of different risk assessment models with regard to choosing the optimal defense and attack strategies. In fact, the results show that the optimal attack and defense strategies differ in between using the expected load shed loss model and the CVaR model. In addition, the results presented show that CVaR is highly adequate when considering losses whose probability of occurrence is low but whose associated magnitude is very high.

The game model adopted enables a dynamic analysis of attack and defense of a multi-area automatic generation control which provides essential tools for the defender to devise appropriate defense strategies against potential attacks.

5 Conclusion

This chapter has introduced and discussed the application of dynamic game theory to model various strategic decision making processes arising in modern power system applications including wholesale competitive electric energy markets, demand-side management, microgrid energy exchange, as well as power system security with applications to state estimation and dynamic stability.

In summary, dynamic game theory can play a vital role in enabling an accurate modeling, assessment, and prediction of the strategic behavior of the various interconnected entities in current and future power systems; which is indispensable to guiding the grid's evolution into a more efficient, economic, resilient, and secure system.

Acknowledgements This work was supported by the US National Science Foundation under Grants ECCS-1549894 and CNS-1446621.

References

Abur A, Exposito AG (2004) Power system state estimation: theory and implementation. Marcel Dekker, New York

Couillet R, Perlaza SM, Tembine H, Debbah M (2012) Electrical vehicles in the smart grid: a mean field game analysis. IEEE J Sel Areas Commun 30(6):1086–1096

El Rahi G, Sanjab A, Saad W, Mandayam NB, Poor HV (2016) Prospect theory for enhanced smart grid resilience using distributed energy storage. In: Proceedings of the 54th annual Allerton conference on communication, control, and computing, Sept 2016

Glover JD, Sarma MS, Overbye TJ (2012) Power system analysis and design, 5th edn. Stamford, Cengage Learning

Gomez-Exposito A, Conejo AJ, Canizares C (2009) Electric energy systems: analysis and operation. CRC Press, Boca Raton

Guan X, Ho Y-C, Pepyne DL (2001) Gaming and price spikes in electric power markets. IEEE Trans Power Syst 16(3):402–408

Horowitz SH, Phadke AG, Niemira JK (2013) Power system relaying, 4th edn. John Wiley & Sons Inc, Chichester, West Sussex

Law YW, Alpcan T, Palaniswami M (2015) Security games for risk minimization in automatic generation control. IEEE Trans Power Syst 30(1):223–232

Li T, Shahidehpour M (2005) Strategic bidding of transmission-constrained GENCOs with incomplete information. IEEE Trans Power Syst 20(1):437–447

Maharjan S, Zhu Q, Zhang Y, Gjessing S, Başar T (2013) Dependable demand response management in the smart grid: a Stackelberg game approach. IEEE Trans Smart Grid 4(1): 120–132

Maharjan S, Zhu Q, Zhang Y, Gjessing S, Başar T (2016) Demand response management in the smart grid in a large population regime. IEEE Trans Smart Grid 7(1):189–199

Nanduri V, Das T (2007) A reinforcement learning model to assess market power under auction-based energy pricing. IEEE Trans Power Syst 22(1):85–95

Saad W, Han Z, Poor HV (2011) Coalitional game theory for cooperative micro-grid distribution networks. In: IEEE international conference on communications workshops (ICC), June 2011, pp 1–5

Saad W, Han Z, Poor HV, Başar T (2012) Game-theoretic methods for the smart grid: an overview of microgrid systems, demand-side management, and smart grid communications. IEEE Signal Process Mag 29(5):86–105

Sanjab A, Saad W (2016a) Data injection attacks on smart grids with multiple adversaries: a game-theoretic perspective. IEEE Trans Smart Grid 7(4):2038–2049

Sanjab A, Saad W (2016b) On bounded rationality in cyber-physical systems security: game-theoretic analysis with application to smart grid protection. In: IEEE/ACM CPS week joint workshop on cyber-physical security and resilience in smart grids (CPSR-SG), Apr 2016, pp 1–6

Sauer PW, Pai MA (1998) Power system dynamics and stability. Prentice Hall, Upper Saddle River

Wood AJ, Wollenberg BF (2012) Power generation, operation, and control. John Wiley & Sons, New York

Zhu Q, Başar T (2011) Robust and resilient control design for cyber-physical systems with an application to power systems. In: 50th IEEE conference on decision and control and European control conference, Dec 2011, pp 4066–4071

Zhu Q, Başar T (2015) Game-theoretic methods for robustness, security, and resilience of cyberphysical control systems: games-in-games principle for optimal cross-layer resilient control systems. IEEE Control Syst Mag 35(1):46–65

Trends and Applications in Stackelberg Security Games

28

Debarun Kar, Thanh H. Nguyen, Fei Fang, Matthew Brown, Arunesh Sinha, Milind Tambe, and Albert Xin Jiang

Contents

D. Kar (✉) · T. H. Nguyen · F. Fang · M. Brown · A. Sinha · M. Tambe
University of Southern California, Los Angeles, CA, USA
e-mail: dkar@usc.edu; thanhhng@usc.edu; feifang@usc.edu; mattheab@usc.edu; aruneshs@usc.edu; tambe@usc.edu

A. X. Jiang
Trinity University, San Antonio, TX, USA
e-mail: xjiang@trinity.edu

T. Başar, G. Zaccour (eds.), *Handbook of Dynamic Game Theory*,
https://doi.org/10.1007/978-3-319-44374-4_27

Abstract

Security is a critical concern around the world, whether it is the challenge of protecting ports, airports, and other critical infrastructure; interdicting the illegal flow of drugs, weapons, and money; protecting endangered wildlife, forests, and fisheries; or suppressing urban crime or security in cyberspace. Unfortunately, limited security resources prevent full security coverage at all times; instead, we must optimize the use of limited security resources. To that end, we founded a new "security games" framework that has led to building of decision aids for security agencies around the world. Security games are a novel area of research that is based on computational and behavioral game theory while also incorporating elements of AI planning under uncertainty and machine learning. Today security-games-based decision aids for infrastructure security are deployed in the US and internationally; examples include deployments at ports and ferry traffic with the US Coast Guard, for security of air traffic with the US Federal Air Marshals, and for security of university campuses, airports, and metro trains with police agencies in the US and other countries. Moreover, recent work on "green security games" has led our decision aids to be deployed, assisting NGOs in protection of wildlife; and "opportunistic crime security games" have focused on suppressing urban crime. In cyber-security domain, the interaction between the defender and adversary is quite complicated with high degree of incomplete information and uncertainty. Recently, applications of game theory to provide quantitative and analytical tools to network administrators through defensive algorithm development and adversary behavior prediction to protect cyber infrastructures has also received significant attention. This chapter provides an overview of use-inspired research in security games including algorithms for scaling up security games to real-world sized problems, handling multiple types of uncertainty, and dealing with bounded rationality and bounded surveillance of human adversaries.

Keywords

Security games · Scalability · Uncertainty · Bounded rationality · Bounded surveillance · Adaptive adversary · Infrastructure security · Wildlife protection

1 Introduction

Security is a critical concern around the world that manifests in problems such as protecting our ports, airports, public transportation, and other critical national infrastructure from terrorists, in protecting our wildlife and forests from poachers and smugglers, and curtailing the illegal flow of weapons, drugs, and money across international borders. In all of these problems, we have limited security resources which prevents security coverage on all the targets at all times; instead, security resources must be deployed intelligently taking into account differences in the importance of targets, the responses of the attackers to the security posture, and potential uncertainty over the types, capabilities, knowledge, and priorities of attackers faced.

To address these challenges in adversarial reasoning and security resource allocation, a new "security games" framework has been developed (Tambe 2011); this framework has led to building of decision aids for security agencies around the world. Security games are based on computational and behavioral game theory while also incorporating elements of AI planning under uncertainty and machine learning. Security games algorithms have led to successes and advances over previous human-designed approaches in security scheduling and allocation by addressing the key weakness of predictability in human-designed schedules. These algorithms are now deployed in multiple applications. The first application was ARMOR, which was deployed at the Los Angeles International Airport (LAX) in 2007 to randomize checkpoints on the roadways entering the airport and canine patrol routes within the airport terminals (Jain et al. 2010b). Following that came several other applications: IRIS, a game-theoretic scheduler for randomized deployment of the US Federal Air Marshals (FAMS), has been in use since 2009 (Jain et al. 2010b); PROTECT, which schedules the US Coast Guard's randomized patrolling of ports, has been deployed in the port of Boston since April 2011 and is in use at the port of New York since February 2012 (Shieh et al. 2012) and has spread to other ports such as Los Angeles/Long Beach, Houston, and others; another application for deploying escort boats to protect ferries has been deployed by the US Coast Guard since April 2013 (Fang et al. 2013); and TRUSTS (Yin et al. 2012) which has been evaluated in field trials by the Los Angeles Sheriff's Department (LASD) in LA Metro system. Most recently, PAWS – another game-theoretic application – was tested by rangers in Uganda for protecting wildlife in Queen Elizabeth National Park in April 2014 (Yang et al. 2014); MIDAS was tested by the US Coast Guard for protecting fisheries (Haskell et al. 2014). These initial successes point the way to major future applications in a wide range of security domains.

Researchers have recently started to explore the use of such security game models in tackling security issues in the cyber world. In Vanek et al. (2012), the authors study the problem of optimal resource allocation for packet selection and inspection to detect potential threats in large computer networks with multiple computers of differing importance. In their paper, they study the application of security games to deep packet inspection as countermeasure to intrusion detection. In a recent paper (Durkota et al. 2015), the authors study the problem of optimal number of *honeypots* to be placed in a network using a security game framework. Another interesting work, called audit games (Blocki et al. 2013, 2015), enhances the security games model with choice of punishments in order to capture scenarios of security and privacy policy enforcement in large organizations (Blocki et al. 2013, 2015).

Given the many game-theoretic applications for solving real-world security problems, this chapter provides an overview of the models and algorithms, key research challenges, and a description of our successful deployments. Overall, the work in security games has produced numerous decision aids that are in daily use by security agencies to optimize their limited security resources. The implementation of these applications required addressing fundamental research challenges. We categorize the research challenges associated with security games into four broad categories: (1) addressing scalability across a number of dimensions of the game, (2) tackling different forms of uncertainty that be present in the game, (3) addressing human adversaries' bounded rationality and bounded surveillance (limited capabilities in surveillance), and (4) evaluation of the framework in the field. Given the success in providing solutions for many security domains involving the protection of critical infrastructure, the topic of security games has evolved and expanded to include new types of security domains, for example, for wildlife and environmental protection.

The rest of the chapter is organized as follows: Sect. 2 introduces the general security games model, Sect. 3 discusses three different types of security games, Sect. 4 describes the approaches used to tackle scalability issues, Sect. 5 describes the approaches to deal with uncertainty, Sect. 6 focuses on bounded rationality and bounded surveillance, and Sect. 7 provides details of field evaluation of the science of security games.

2 Stackelberg Security Games

Stackelberg games were first introduced to model leadership and commitment (von Stackelberg 1934). A Stackelberg game is a game played sequentially between two players: the first player is the leader who commits to a strategy first, and then the second player, called the follower, observes the strategy of the leader and then commits to his own strategy. The term Stackelberg security games (SSG) was first introduced by Kiekintveld et al. (2009) to describe specializations of a particular type of Stackelberg game for security as discussed below. This section provides details on this use of Stackelberg games for modeling security domains. We first give a generic description of security domains followed by *security games*, the model by which security domains are formulated in the Stackelberg game framework.[1]

2.1 Stackelberg Security Game

In Stackelberg security games, a defender must perpetually defend a set of targets T using a limited number of resources, whereas the attacker is able to surveil and learn the defender's strategy and attack after careful planning. An action, or *pure strategy*, for the defender represents deploying a set of resources R on patrols or checkpoints, e.g., scheduling checkpoints at the LAX airport or assigning federal air marshals to protect flight tours. The pure strategy for an attacker represents an attack at a target, e.g., a flight. The *mixed strategy* of the defender is a probability

[1] Note that *not* all security games in the literature are Stackelberg security games (see Alpcan and Başar 2010).

distribution over the pure strategies. Additionally, each target is also associated with a set of payoff values that define the utilities for both the defender and the attacker in case of a successful or a failed attack.

A key assumption of Stackelberg security games (we will sometimes refer to them as simply security games) is that the payoff of an outcome depends only on the target attacked and whether or not it is *covered* (protected) by the defender (Kiekintveld et al. 2009). The payoffs do *not* depend on the remaining aspects of the defender allocation. For example, if an adversary succeeds in attacking target t_1, the penalty for the defender is the same whether the defender was guarding target t_2 or not.

This allows us to compactly represent the payoffs of a security game. Specifically, a set of four payoffs is associated with each target. These four payoffs are the rewards and penalties to both the defender and the attacker in case of a successful or an unsuccessful attack and are sufficient to define the utilities for both players for all possible outcomes in the security domain. More formally, if target t is attacked, the defender's utility is $U_d^c(t)$ if t is covered or $U_d^u(t)$ if t is not covered. The attacker's utility is $U_a^c(t)$ if t is covered or $U_a^u(t)$ if t is not covered. Table 28.1 shows an example security game with two targets, t_1 and t_2. In this example game, if the defender was covering target t_1 and the attacker attacked t_1, the defender would get 10 units of reward, whereas the attacker would receive -1 units. We make the assumption that in a security game, it is always better for the defender to cover a target as compared to leaving it uncovered, whereas it is always better for the attacker to attack an uncovered target. This assumption is consistent with the payoff trends in the real world. A special case is *zero-sum games*, in which for each outcome the sum of utilities for the defender and attacker is zero, although general security games are not necessarily zero sum.

2.2 Solution Concept: Strong Stackelberg Equillibrium

The solution to a security game is a *mixed* strategy[2] for the defender that maximizes the expected utility of the defender, given that the attacker learns the mixed strategy of the defender and chooses a best response for himself. The defender's mixed

Table 28.1 Example of a security game with two targets

	Defender		Attacker	
Target	Covered	Uncovered	Covered	Uncovered
t_1	10	0	-1	1
t_2	0	-10	-1	1

[2]Note that mixed strategy solutions apply beyond Stackelberg games.

strategy is a probability distribution over all pure strategies, where a pure strategy is an assignment of the defender's limited security resources to targets. This solution concept is known as a Stackelberg equilibrium (Leitmann 1978).

The most commonly adopted version of this concept in related literature is called Strong Stackelberg Equilibrium (SSE) (Breton et al. 1988; Conitzer and Sandholm 2006; Paruchuri et al. 2008; von Stengel and Zamir 2004). In security games, the mixed strategy of the defender is equivalent to the probabilities that each target t is covered by the defender, denoted by $C = \{c_t\}$ (Korzhyk et al. 2010). Furthermore, it is enough to consider a pure strategy of the rational adversary (Conitzer and Sandholm 2006), which is to attack a target t. The expected utility for defender for a strategy profile (C, t) is defined as $U_d(t, C) = c_t U_d^c(t) + (1 - c_t) U_d^u(t)$ and a similar form for the adversary. An SSE for the basic security games (non-Bayesian, rational adversary) is defined as follows:

Definition 1. A pair of strategies (C^*, t^*) form a *Strong Stackelberg Equilibrium* (SSE) if they satisfy the following:

1. The defender plays a best response: $U_d(t^*, C^*) \geq U_d(t(C), C)$ for all defender's strategy C where $t(C)$ is the attacker's response against the defender strategy C.
2. The attacker plays a best-response: $U_a(t^*, C^*) \geq U_a(t, C^*)$ for all target t.
3. The attacker breaks ties in favor of the defender: $U_d(t^*, C^*) \geq U_d(t', C^*)$ for all target t' such that $t' = \text{argmax}_t\, U_a(t, C^*)$.

The assumption that the follower will always break ties in favor of the leader in cases of indifference is reasonable because in most cases the leader can induce the favorable strong equilibrium by selecting a strategy arbitrarily close to the equilibrium that causes the follower to strictly prefer the desired strategy (von Stengel and Zamir 2004). Furthermore an SSE exists in all Stackelberg games, which makes it an attractive solution concept compared to versions of Stackelberg equilibrium with other tie-breaking rules. Finally, although initial applications relied on the SSE solution concept, we have since proposed new solution concepts that are more robust against various uncertainties in the model (An et al. 2011; Pita et al. 2012; Yin et al. 2011) and have used these robust solution concepts in some of the later applications.

For simple examples of security games, such as the one shown above, the Strong Stackelberg Equilibrium can be calculated by hand. However, as the size of the game increases, hand calculation is no longer feasible, and an algorithmic approach for generating the SSE becomes necessary. Conitzer and Sandholm (Conitzer and Sandholm 2006) provided the first complexity results and algorithms for computing optimal commitment strategies in Stackelberg games, including both pure- and mixed-strategy commitments. An improved algorithm for solving Stackelberg games, DOBSS (Paruchuri et al. 2008), is central to the fielded application ARMOR that was in use at the Los Angeles International Airport (Jain et al. 2010b).

Decomposed Optimal Bayesian Stackelberg Solver (DOBSS): We now describe the DOBSS[3] in detail as it provides a starting point for the algorithms we develop in the next section. We first present DOBSS in its most intuitive form as a mixed-integer quadratic program (MIQP); we then present a linearized equivalent mixed-integer linear program (MILP). The DOBSS model explicitly represents the actions by the leader and the *optimal* actions for the follower in the problem solved by the leader. Note that we need to consider only the reward-maximizing pure strategies of the follower, since for a given fixed mixed strategy x of the leader, each follower faces a problem with fixed linear rewards. If a mixed strategy is optimal for the follower, then so are all the pure strategies in support of that mixed strategy.

Thus, we denote by x the leader's policy, which consists of a probability distribution over the leader's pure strategies $\sigma_i \in \Sigma_\Theta$, where Σ_Θ is the set of all pure strategies of the leader. Hence, the value x_i is the proportion of times in which pure strategy $\sigma_i \in \Sigma_\Theta$ is used in the policy. Similarly, q_j is the probability of taking strategy $\sigma_j \in \Sigma_\Psi$ for the follower, where Σ_Ψ is the set of all pure strategies for the follower. We denote by X and Q the index sets of the leader and follower pure strategies, respectively. We also index the payoff matrices of the leader and the follower by the matrices R and C where R_{ij} and C_{ij} are the rewards obtained if the leader takes strategy $\sigma_i \in \Sigma_\Theta$ and the follower takes strategy $\sigma_j \in \Sigma_\Psi$. Let M be a large positive number; constraint 3 in the MIQP below requires that the variable a be set to the maximum reward a follower can obtain given the current policy x taken by the leader. The leader then solves the following:

$$\max_{x,q,a} \quad \sum_{i\in X}\sum_{j\in Q} R_{ij}x_i q_j \tag{28.1}$$

$$\text{s.t.} \quad \textstyle\sum_{i\in X} x_i = 1 \tag{28.2}$$

$$\textstyle\sum_{j\in Q} q_j = 1 \tag{28.3}$$

$$0 \le (a - \textstyle\sum_{i\in X} C_{ij}x_i) \le (1-q_j)M \quad \forall j \in Q \tag{28.4}$$

$$x_i \in [0\ldots 1] \quad \forall i \in X \tag{28.5}$$

$$q_j \in \{0,1\} \quad \forall j \in Q \tag{28.6}$$

$$a \in \Re \tag{28.7}$$

Here, for a leader strategy x and a strategy q for the follower, the objective (Line 1) represents the expected reward for the leader. The first (Line 2) and the fourth (Line 5) constraints define the set of feasible solutions $x \in X$ as a probability distribution over the set of strategies $\sigma_i \in \Sigma_\Theta$. The second (Line 3) and third (Line 6) constraints limit the vector of strategies, q, to be a pure strategy over the set Q (that is each q has exactly one coordinate equal to one and the rest equal

[3]DOBSS addresses Bayesian Stackelberg games with multiple follower types, but for simplicity we do not introduce Bayesian Stackelberg games here.

to zero). The two inequalities in the third constraint (Line 4) ensure that $q_j = 1$ only for a strategy j that is optimal for the follower. Indeed this is a linearized form of the optimality conditions for the linear programming problem solved by each follower. We explain the third constraint (Line 4) as follows: this constraint enforces dual feasibility of the follower's problem (leftmost inequality) and the complementary slackness constraint for an optimal pure strategy q for the follower (rightmost inequality). Note that the leftmost inequality ensures that $\forall j \in Q$, $a \geq \sum_{i \in X} C_{ij} x_i$. This means that given the leader's policy x, a is an upper bound on follower's reward for any strategy. The rightmost inequality is inactive for every strategy where $q_j = 0$, since M is a large positive quantity. In fact, since only one pure strategy can be selected by the follower, say some $q_j = 1$, for the strategy that has $q_j = 1$, this inequality states $a \leq \sum_{i \in X} C_{ij} x_i$, which combined with the left inequality enforces $a = \sum_{i \in X} C_{ij} x_i$, thereby imposing no additional constraint for all other pure strategies which have $q_j = 0$ and showing that this strategy must be optimal for the follower.

We can linearize the quadratic programming problem (Lines 28.1 to 7) through the change of variables $z_{ij} = x_i q_j$ to obtain a mixed integer linear programming problem as shown in Paruchuri et al. (2008).

$$\max_{q,z,a} \quad \sum_{i \in X} \sum_{j \in Q} pR_{ij} z_{ij} \tag{28.8}$$

$$\text{s.t.} \quad \sum_{i \in X} \sum_{j \in Q} z_{ij} = 1 \tag{28.9}$$

$$\sum_{j \in Q} z_{ij} \leq 1 \qquad \forall i \in X \tag{28.10}$$

$$q_j \leq \sum_{i \in X} z_{ij} \leq 1 \qquad \forall j \in Q \tag{28.11}$$

$$\sum_{j \in Q} q_j = 1 \tag{28.12}$$

$$0 \leq (a - \sum_{i \in X} C_{ij} (\sum_{h \in Q} z_{ih})) \leq (1 - q_j) M \quad \forall j \in Q \tag{28.13}$$

$$\sum_{j \in Q} z_{ij} = \sum_{j \in Q} z^1_{ij} \qquad \forall i \in X \tag{28.14}$$

$$z_{ij} \in [0 \ldots 1] \qquad \forall i \in X, j \in Q \tag{28.15}$$

$$q_j \in \{0, 1\} \qquad \forall j \in Q \tag{28.16}$$

$$a \in \Re \tag{28.17}$$

DOBSS solves this resulting mixed integer linear program using efficient integer programming packages. The MILP was shown to be equivalent to the MIQP (Lines 28.1 to 7) and the equivalent Harsanyi transformed Stackelberg game (Paruchuri et al. 2008). For a more in-depth explanation of DOBSS, please see Paruchuri et al. (2008).

3 Categorizing Security Games

With progress in the security games research and the expanding set of applications, it is valuable to consider categorizing this work into three separate areas. These categories are driven by applications, but they also impact the types of games (e.g., single shot vs repeated games) considered and the research issues that arise. Specifically, the three categories are (i) infrastructure security games, (ii) green security games, and (iii) opportunistic crime security games. We discuss each category below.

3.1 Infrastructure Security Games

These types of games and their applications are where the original research on security games was initiated. Key characteristics of these games include the following:

– *Application characteristics*: These games are focused on applications of protecting infrastructure, such as ports, airports, trains, flights, and so on; the goal is often assisting agencies engaged in counterterrorism. Notice that the infrastructure being protected tends to be static, and little changes in a few months, e.g., an airport being protected, may have new construction once in 2–3 years. The activities in the infrastructure are regulated by well-established schedules of movement of people or goods. Furthermore, the targets being protected often have a discrete structure, e.g., terminals at an airport, individual flights, individual trains, etc.
– *Overall characteristics of the defender and adversary play*: These games are single-shot games. The defender does play her strategy repeatedly, i.e., the defender commits to a mixed strategy in this security game. This mixed strategy may get played for months at a time. However, a single attack by an adversary ends the game. The game could potentially restart after such an attack, but it is not set up as a repeated game as in the game categories described below.
– *Adversary characteristics*: The games assume that the adversaries are highly strategic, who may attack after careful planning and surveillance. These carefully planned attacks have high consequences. Furthermore, since these attacks are a result of careful planning with the anticipation of high consequences, attackers commit to these plans of attacks and are not considered to opportunistically move from target to target.
– *Defender characteristics*: The defender does not repeatedly update her strategies. In these domains, there may be just a few attacks that may occur, but these tend to be rare; they are not a very large number of attacks that occur repeatedly. As a result, traditionally, no machine learning is used in this work for the defender to update her strategies over time.

3.2 Green Security Games

These types of games and their applications are focused on trying to protect the environment; and we adopt the term from “green criminology.”[4]

- *Application characteristics*: These games are focused on applications of protecting the environment, including forests, fish, and wildlife. The goal is thus often to assist security agencies against poachers, illegal fishermen, or those illegally cutting trees in national parks in countries around the world. Unlike infrastructure security games, animals or fish being protected may move around in geographical space, introducing new dimensions of complexity. Finally, the targets being protected are spread out over vast open geographical spaces, e.g., large forest regions protect trees from illegal cutting.
- *Overall characteristics of the defender and adversary play*: These games are **not** single-shot games. Unfortunately, the adversaries often conduct multiple repeated “attacks,” e.g., poaching animals repeatedly. Thus, a single illegal activity does not end the game. Instead, usually, after obtaining reports, e.g., over a month, of illegal activities, the defender often replans her security activities. In other words, these are repeated security games where the defender plays a mixed strategy while the attacker attacks multiple times, and then the defender replans and plays a new mixed strategy and the cycle repeats. Notice also that the illegal activities of concern here may be conducted by multiple individuals, and thus there are multiple adversaries that are active at any one point.
- *Adversary characteristics*: As mentioned earlier, the adversaries are engaged in repeated illegal activities; and the consequences of failure or success are not as severe as in the case of counterterrorism. As a result, every single attack (illegal action) cannot be carried out with the most detailed surveillance and planning; the adversaries will hence exhibit more of a bounded rationality and bounded surveillance in these domains.

  Nonetheless, these domains are not ones where illegal activities can be conducted opportunistically (as in the opportunistic crime security games discussed below). This is because in these green security games, the adversaries often have to act in extremely dangerous places (e.g., deep in forests, protecting themselves from wild animals), and thus given the risks involved, they cannot take an entirely opportunistic approach.

[4]We use the term green security games also to avoid any confusion that may come about given that terms related to the environment and security have been adopted for other uses. For example, the term “environmental security” broadly speaking refers to threats posed to humans due to environmental issues, e.g., climate change or shortage of food. The term “environmental criminology” on the other hand refers to analysis and understanding of how different environments affect crime.

– *Defender characteristics*: Since this is a repeated game setting, the defender repeatedly updates her strategies. Machine learning can now be used in this work for the defender to update her strategies over time, given that attack data is available over time. The presence of large amounts of such attack data is very unfortunate in that very large numbers of crimes against the environment are recorded in real life, but the silver lining is that the defender can improve her strategy exploiting this data.

3.3 Opportunistic Crime Security Games

These types of games and their applications are focused on trying to combat opportunistic crime. Such opportunistic crime may include criminals engaged in thefts such as snatching of cell phones in metros or stealing student laptops from libraries.

– *Application characteristics*: These games focused on applications involving protecting the public against opportunistic crime. The goal is thus often to assist security agencies in protecting public's property such as cell phones, laptops, or other valuables. Here, human crowds may move around based on scheduled activities, e.g., office hours in downtown settings or class timings on a university campus, and thus the focus of what needs to be protected may shift on a regular schedule. At least in urban settings, these games focus on specific limited geographical areas as opposed to vast open spaces as involved in "green security games."
– *Overall characteristics of the defender and adversary play*: While these games are not explicitly formulated as repeated games, the adversary may conduct or attempt to conduct multiple "attacks" (thefts) in any one round of the game. Thus, the defender commits to a mixed strategy, but a single attack by a single attacker does not end the game. Instead multiple attackers may be active at a time, conducting multiple thefts while the defender attempts to stop these thefts from taking place.
– *Adversary characteristics*: Once again, the adversaries are engaged in repeated illegal activities; and the consequences of failure or success are not as severe as in the case of counterterrorism. As a result, once again, given that every single attack (illegal action) cannot be carried out with the most detailed surveillance and planning, the adversaries may thus act even less strategically and exhibit more of a bounded rationality and bounded surveillance in these domains. Furthermore, the adversaries are not as committed to detailed plans and are flexible in their execution of their plans, as targets of opportunity present themselves.
– *Defender characteristics*: How to update defender strategies in these games from crime data is still an open research challenge.

3.4 Cybersecurity Games

These types of games and their applications are focused on trying to combat cyber crimes. Such crimes include, attackers compromising network infrastructures to disrupt normal system operations, launch a physical attack, steal valuable digital data, conducting social engineering attacks such as phishing attacks, etc.

– *Application characteristics*: These games are focused on applications involving protecting network assets against cyber attacks. The goal is thus often to assist network administrators in protecting computer systems such as data servers, switches, etc., from data theft or damage to hardware, software, or information, as well as preventing disruption of services.
– *Overall characteristics of the defender and adversary play*: Depending on the problem at hand, the attacker (or the intruder) may want to gain control over (or to disable) a valuable computer in the network by scanning the network, compromising a more vulnerable system, and/or gaining access to further devices on the computer network. The ultimate goal could be to use the compromised systems to launch further attacks or to steal data, etc. The broader goal of the defender (a human network administrator, or a detection system) could be formulated as preventing the adversary from gaining control over systems in the network by detecting malicious attacks.
– *Adversary characteristics*: The adversary's characteristics vary from one application domain to another. In some application scenarios, the intruder may simply want to gain control over (or to disable) a valuable computer in the network to launch other attacks, by scanning the network and thus compromising a more vulnerable system and/or gaining access to further devices on the computer network. The actions of the attacker can therefore be seen as sending malicious packets from a controlled computer (termed source) to a single or multiple vulnerable computers (termed targets). In other scenarios, the attacker may be interested in stealing valuable information from a particular data server and therefore takes necessary actions to compromise the desired system, possibly through a series of disruptions as studied in the advanced persistent threat (APT) literature.
– *Defender characteristics*: Although this is a new and open problem, there has been recent literature that studies the problem of optimal defender resource allocation for packet selection and inspection to detect potential threats in large computer networks with multiple computers of differing importance. Therefore, the objective of the defender in such problems is to prevent the intruder from succeeding by selecting the packets for inspection, identifying the attacker, and subsequently thwarting the attack.

Even though we have categorized the research and applications of security games in these three categories, not everything is very cleanly divided in this fashion.

Further research may reveal other categories of need to generate subcategories of the above three categories.

In the rest of this chapter, we will concentrate only on infrastructure security games and green security games. In the following sections, we first present three key challenges in solving real-world security problems which are summarized in Fig. 28.1: (1) scaling up to real-world sized security problems, (2) handling multiple uncertainties in security games, and (3) dealing with bounded rationality and bounded surveillance of human adversaries. While Fig. 28.1 does not provide an exhaustive overview of all research in SSG, it provides a general overview of the areas of research and a road map to the rest of the book chapter. In each case, we will use a domain example to motivate the specific challenge and then outline the key algorithmic innovation needed to address the challenge.

	Challenge	**Domain Example**	**Algorithmic Solution**
Scalability	Large defender strategy space	Federal Air Marshals Service	ASPEN: strategy generation approach
	Large defender & attacker strategy spaces	Road Network Security	RUGGED: double oracle approach
	Mobile resources & moving targets	Ferry Protection	CASS: compact representation of strategy
	Multiple boundedly rational attackers	Fishery Protection	MIDAS: cutting plane approach
	Incorporating fine-grained spatial information	Green Security Domains: wildlife/fishery protection	Hierarchical modeling approach
Uncertainty	Unifications of uncertainties	Security in LAX Airport	URAC: multi-dimensional reduction & divide-and-conquer approach
	Dynamic execution uncertainty	Security in Transit System	Markov Decision Processes approach
Attacker Bounded Rationality and Bounded Surveillance	Learning attacker behaviors	Green Security Domains: wildlife/fishery protection	Behavioral models & Human subject experiments

Fig. 28.1 Summary of Real-world Security Challenges

4 Addressing Scalability in Real-World Problems

The early works in Stackelberg security games such as DOBSS (Paruchuri et al. 2008) required that the full set of pure strategies for both players be considered when modeling and solving Stackelberg security games. However, many real-world problems feature billions of pure strategies for either the defender and/or the attacker. Such large problem instances cannot even be represented in modern computers, let alone solved using previous techniques.

In addition to large strategy spaces, there are other scalability challenges presented by different real-world security domains. There are domains where, rather than being static, the targets are moving and thus the security resources need to be mobile and move in a continuous space to provide protection. There are also domains where the attacker may not conduct the careful surveillance and planning that is assumed for a Strong Stackelberg Equilibrium, and thus it is important to model the bounded rationality and bounded surveillance of the attacker in order to predict their behavior. In the former case, both the defender and attacker's strategy spaces are infinite. In the latter case, computing the optimal strategy for the defender given attacker behavioral (bounded rationality and/or bounded surveillance) model is computationally expensive. Furthermore, in certain domains, it is important to incorporate fine-grained topographical information to generate realistic patrol strategies for the defender. However, in doing so, existing techniques lead to a significant challenge in scalability especially when scheduling constraints need to be satisfied. In this section, we thus highlight the critical scalability challenges faced to bring Stackelberg security games to the real world and the research contributions that served to address these challenges.

4.1 Scale Up with Large Defender Strategy Spaces

This section provides an example of a research challenge in security games where the number of defender strategies is too enormous to be enumerated in computer memory. In this section, as in others that will follow, we will first provide a domain example motivating the challenge and then the algorithmic solution for the challenge.

Domain Example – IRIS for US Federal Air Marshals Service. The US Federal Air Marshals Service (FAMS) allocates air marshals to flights departing from and arriving in the USA to dissuade potential aggressors and prevent an attack should one occur. Flights are of different importance based on a variety of factors such as the numbers of passengers, the population of source and destination cities, and international flights from different countries. Security resource allocation in this domain is significantly more challenging than for ARMOR: a limited number of air marshals need to be scheduled to cover thousands of commercial flights each day. Furthermore, these air marshals must be scheduled on tours of flights that obey various constraints (e.g., the time required to board, fly, and disembark). Simply finding schedules for the marshals that meet all of these constraints is a computational challenge. For an example scenario with 1000 flights and 20 marshals, there are over 10^{41} possible schedules that could be considered. Yet there are currently tens of thousands of commercial flights flying each day, and public estimates state that there are thousands of air marshals that are scheduled daily by the FAMS (Keteyian 2010). Air marshals must be scheduled on tours of flights that obey logistical constraints (e.g., the time required to board, fly, and disembark). An example of a schedule is an air marshal assigned to a round trip from New York to London and back.

Against this background, the IRIS system (Intelligent Randomization In Scheduling) has been developed and deployed by FAMS since 2009 to randomize schedules of air marshals on international flights. In IRIS, the targets are the set of n flights and the attacker could potentially choose to attack one of these flights. The FAMS can assign $m < n$ air marshals that may be assigned to protect these flights.

Since the number of possible schedules exponentially increases with the number of flights and resources, DOBSS is no longer applicable to the FAMS domain. Instead, IRIS uses the much faster ASPEN algorithm (Jain et al. 2010a) to generate the schedule for thousands of commercial flights per day.

Algorithmic Solution-Incremental Strategy Generation (ASPEN). In this section, we describe one particular algorithm, ASPEN, that computes strong Stackelberg equilibria (SSE) in domains with a *very large* number of pure strategies (up to billions of actions) for the defender (Jain et al. 2010a). ASPEN builds on the insight that in many real-world security problems, there exist solutions with *small support sizes*, which are mixed strategies in which only a small set of pure strategies are played with positive probability (Lipton et al. 2003). ASPEN exploits this by using a *incremental strategy generation* approach for the defender, in which defender pure strategies are iteratively generated and added to the optimization formulation.

In ASPEN's security game, the attacker can choose any of the flights to attack, and each air marshal can cover one schedule. Each schedule here is a feasible set of targets that can be covered together; for the FAMS, each schedule would represent a flight tour which satisfies all the logistical constraints that an air marshal could fly. For example, $\{t_1, t_2\}$ would be a flight schedule, where t_1 is an outbound flight and t_2 is an inbound flight for one air marshal. A *joint schedule* then would assign every air marshal to a flight tour, and there could be exponentially many joint schedules in the domain. A pure strategy for the defender in this security game is a joint schedule. Thus, for example, if there are two air marshals, one possible joint schedule would be $\{\{t_1, t_2\}, \{t_3, t_4\}\}$, where the first air marshal covers flights t_1 and t_2 and the second covers flights t_3 and t_4. As mentioned previously, ASPEN employs incremental strategy generation since all the defender pure strategies cannot be enumerated for such a massive problem. ASPEN decomposes the problem into a *master* problem and a *slave* problem, which are then solved iteratively. Given a number of pure strategies, the master solves for the defender and the attacker optimization constraints, while the slave is used to generate a new pure strategy

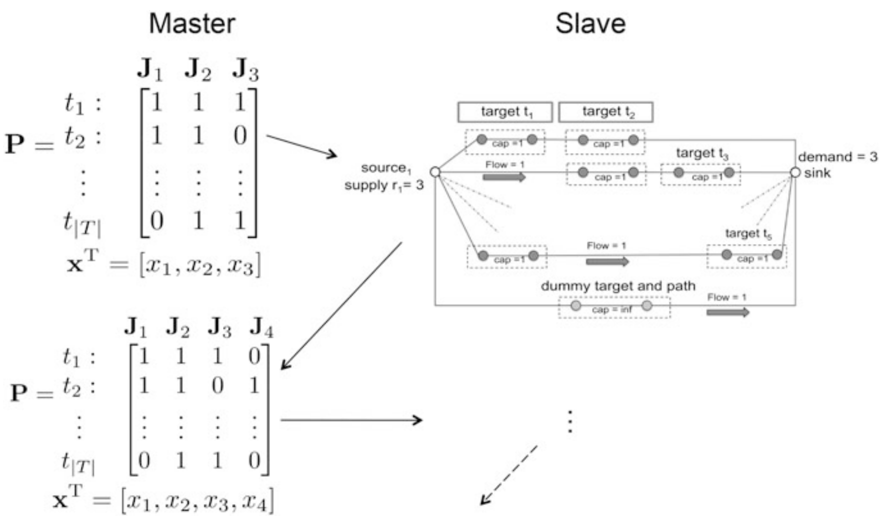


Fig. 28.2 Strategy generation employed in ASPEN: The schedules for a defender are generated iteratively. The *slave* problem is a novel minimum-cost integer flow formulation that computes the new pure strategy to be added to **P**; $\mathbf{J}_4$ is computed and added in this example

for the defender in every iteration. *This incremental, iterative strategy generation process allows ASPEN to avoid generation of the entire set of pure strategies*. In other words, by exploiting the small support size mentioned above, only a few pure strategies get generated via the iterative process; and yet we are guaranteed to reach the optimal solution.

The iterative process is graphically depicted in Fig. 28.2. The master operates on the pure strategies (joint schedules) generated thus far, which are represented using the matrix **P**. Each column of **P**, $\mathbf{J}_j$, is one pure strategy (or joint schedule). An entry P_{ij} in the matrix **P** is 1 if a target t_i is covered by joint-schedule $\mathbf{J}_j$, and 0 otherwise. For example, in Fig. 28.2, the joint schedule J_3 covers target t_1 but not target t_2. The objective of the master problem is to compute **x**, the optimal mixed strategy of the defender over the pure strategies in **P**. The objective function for the slave is updated based on the solution of the master, and the slave is solved to identify the best new column to add to the master problem, using reduced costs (explained later). If no column can improve the solution, the algorithm terminates. Therefore, in terms of our example, the objective of the slave problem is to generate the best joint schedule to add to **P**. The best joint schedule is identified using the concept of *reduced costs*, which captures the total change in the defender payoff if a candidate column is added to **P**, i.e., it measures if a pure strategy can potentially increase the defender's expected utility. The candidate column with minimum reduced cost improves the objective value the most. The details of the approach are provided in Jain et al. (2010a)). While a naïve approach would be to iterate overall possible pure strategies to identify the pure strategy with the maximum potential, ASPEN uses a novel minimum-cost integer flow problem to efficiently identify the best pure strategy to add. ASPEN always converges on the optimal mixed strategy for the defender.

Employing incremental strategy generation for large optimization problems is not an "out-of-the-box" approach; the problem has to be formulated in a way that allows for domain properties to be exploited. The novel contribution of ASPEN is to provide a linear formulation for the master and a minimum-cost integer flow formulation for the slave, which enables the application of strategy generation techniques.

4.2 Scale Up with Large Defender and Attacker Strategy Spaces

Whereas the previous section focused on domains where only the defender's strategy was difficult to enumerate, we now turn to domains where both defender and attacker strategies are difficult to enumerate. Once again we provide a domain example and then an algorithmic solution.

Domain Example – Road Network Security One area of great importance is securing urban city networks, transportation networks, computer networks, and other network-centric security domains. For example, after the terrorist attacks in Mumbai of 2008 (Chandran and Beitchman 2008), the Mumbai police started setting up vehicular checkpoints on roads. We can model the problem faced by the Mumbai police as a security game between the Mumbai police and an attacker. In this urban security game, the pure strategies of the defender correspond to allocations of resources to edges in the network – for example, an allocation of police checkpoints to roads in the city. The pure strategies of the attacker correspond to paths from any *source* node to any *target* node – for example, a path from a landing spot on the coast to the airport.

The strategy space of the defender grows exponentially with the number of available resources, whereas the strategy space of the attacker grows exponentially with the size of the network. For example, in a fully connected graph with 20 nodes and 190 edges, the number of defender pure strategies for only five defender resources is $\binom{190}{5}$ or almost 2 billion, while the number of attacker pure strategies (i.e., paths without cycles) is on the order of 10^{18}. Real-world networks are significantly larger, e.g., the entire road network of the city of Mumbai has 9,503 nodes (intersections) and 20,416 edges (streets), and the security forces can deploy dozens (but not as many as number of edges) of resources. In addressing this computational challenge, novel algorithms based on incremental strategy generation have been able to generate randomized defender strategies that scale up to the entire road network of Mumbai (Jain et al. 2013).

Algorithmic Solution-Double Oracle Incremental Strategy Generation (RUGGED) In domains such as the urban network security setting, the number of pure strategies of both the defender and the attacker is exponentially large. In

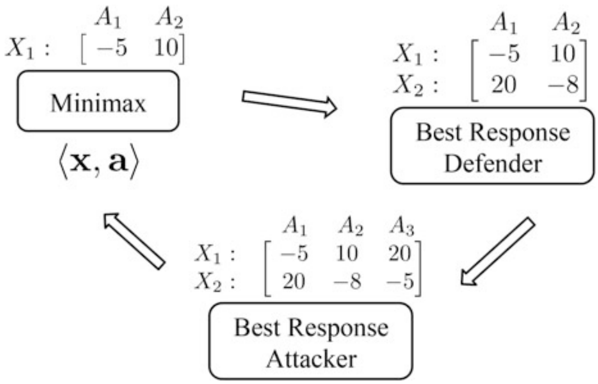


Fig. 28.3 Strategy Generation employed in RUGGED: The pure strategies for both the defender and the attacker are generated iteratively

this section, we describe the RUGGED algorithm (Jain et al. 2011), which generates pure strategies for both the defender and the attacker.

RUGGED models the domain as a zero-sum game and computes the minimax equilibrium, since the minimax strategy is equivalent to the SSE in zero-sum games. Figure 28.3 shows the working of RUGGED: at each iteration, the minimax module generates the optimal mixed strategies $\langle \mathbf{x}, \mathbf{a} \rangle$ for the two players for the current payoff matrix, the Best Response Defender module generates a new strategy for the defender that is a best response against the attacker's current strategy $\mathbf{a}$, and the Best Response Attacker module generates a new strategy for the attacker that is a best response against the defender's current strategy $\mathbf{x}$. The rows X_i in the figure are the pure strategies for the defender; they would correspond to an allocation of checkpoints in the urban road network domain. Similarly, the columns A_j are the pure strategies for the attacker; they represent the attack paths in the urban road network domain. The values in the matrix represent the payoffs to the defender. For example, in Fig. 28.3, the row denoted by X_1 indicates that there was one checkpoint setup, and it provides a defender payoff of -5 against attacker strategy (path) A_1 and a payoff of 10 against attacker strategy (path) A_2.

In Fig. 28.3, we show that RUGGED iterates over two oracles: the defender best response and the attacker best response oracles. In this case, the defender best response oracle has added a strategy X_2, and the attacker best response oracle then adds a strategy A_3. The algorithm stops when neither of the generated best responses improve on the current minimax strategies.

The contribution of RUGGED is to provide the mixed integer formulations for the best response modules which enable the application of such a strategy generation approach. The key once again is that RUGGED is able to converge to the optimal solution without enumerating the entire space of defender and attacker strategies. However, originally RUGGED could only compute the optimal solution for deploying up to *four* resources in real-city network with 250 nodes within a time frame of 10 h (the complexity of this problem can be estimated by observing that both the best response problems are NP hard themselves (Jain et al. 2011)).

More recent work Jain et al. (2013) builds on RUGGED and proposes SNARES, which allows scale-up to the entire city of Mumbai, with 10–15 checkpoints.

4.3 Scale-Up with Mobile Resources and Moving Targets

Whereas the previous two sections focused on incremental strategy generation as an approach for scale-up, this section introduces another approach: use of compact marginal probability representations. This alternative approach is shown in use in the context of a new application of protecting ferries.

Domain Example – Ferry Protection for the US Coast Guard The US Coast Guard is responsible for protecting domestic ferries, including the Staten Island Ferry in New York, from potential terrorist attacks. Here are a number of ferries carrying hundreds of passengers in many waterside cities. These ferries are attractive targets for an attacker who can approach the ferries with a small boat packed with explosives at any time; this attacker's boat may only be detected when it comes close to the ferries. Small, fast, and well-armed patrol boats can provide protection to such ferries by detecting the attacker within a certain distance and stop him from attacking with the armed weapons. However, the number of patrol boats is often limited; thus, the defender cannot protect the ferries at all times and locations. We thus developed a game-theoretic system for scheduling escort boat patrols to protect ferries, and this has been deployed at the Staten Island Ferry since 2013 (Fang et al. 2013).

The key research challenge is the fact that the ferries are continuously moving in a continuous domain, and the attacker could attack at any moment in time. This type of moving targets domain leads to game-theoretic models with continuous strategy spaces, which presents computational challenges. Our theoretical work showed that while it is "safe" to discretize the defender's strategy space (in the sense that the solution quality provided by our work provides a lower bound), discretizing the attacker's strategy space would result in loss of utility (in the sense that this would provide only an upper bound, and thus an unreliable guarantee of true solution quality). We developed a novel algorithm that uses a compact representation for the defender's mixed strategy space while being able to exactly model the attacker's continuous strategy space. The implemented algorithm, running on a laptop, is able to generate daily schedules for escort boats with guaranteed expected utility values (Fig. 28.4).

Algorithmic Solution – Compact Strategy Representation (CASS). In this section, we describe the CASS (Solver for Continuous Attacker Strategy) algorithm (Fang et al. 2013) for solving security problems where the defender has mobile patrollers to protect a set of mobile targets against the attacker who can attack these moving targets at any time during their movement. In these security problems, the sets of pure strategies for both the defender and attacker are continuous w.r.t, the continuous spatial and time components of the problem domain. The CASS

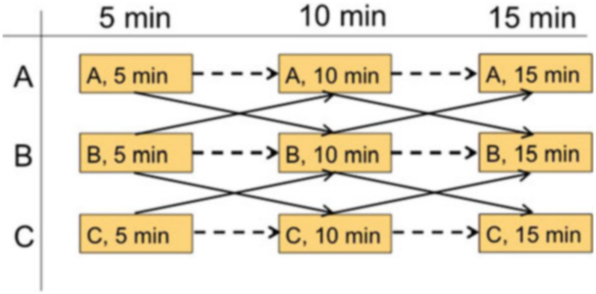
Fig. 28.4 Escort boats protecting the Staten Island Ferry use strategies generated by our system

algorithm attempts to compute the optimal mixed strategy for the defender without discretizing the attacker's continuous strategy set; it exactly models this set using subinterval analysis which exploits the piecewise-linear structure of the attacker's expected utility function. The insight of CASS is to compactly represent the defender's mixed strategies as a *marginal* probability distribution, overcoming the shortcoming of an exponential number of pure strategies for the defender.

CASS casts problems such as the ferry protection problem mentioned above as a *zero-sum* security game in which targets move along a *one-dimensional* domain, i.e., a straight line segment connecting two terminal points. This *one-dimensional* assumption is valid as in real-world domains such as ferry protection, ferries normally move back-and-forth in a straight line between two terminals (i.e., ports) around the world. Although the targets' locations vary w.r.t time changes, these targets have a fixed daily schedule, meaning that determining the locations of the targets at a certain time is straightforward. The defender has mobile patrollers (i.e., boats) that can move along between two terminals to protect the targets. While the defender is trying to protect the targets, the attacker will decide to attack a certain target at a certain time. The probability that the attacker successfully attacks depends on the positions of the patroller at that time. Specifically, each patroller possesses a protective circle of radius within which she can detect and try to intercept any attack, whereas she is incapable of detecting the attacker prior to that radius.

In CASS, the defender's strategy space is discretized, and her mixed strategy is compactly represented using flow distributions. Figure 28.5 shows an example of

Fig. 28.5 An example of a ferry transition graph



a ferry transition graph in which each node of the graph indicates a particular pair of (location, time step) for the target. Here, there are three location points namely A, B, and C on a straight line where B lies between A and C. Initially, the target is at one of these location points at the 5-min time step. Then the target moves to the next location point which is determined based on the connectivity between these points at the 10-min time step and so on. For example, if the target is at the location point A at the 5-min time step, denoted by (A, 5 min) in the transition graph, it can move to the location point B or stay at location point A at the 10-min time step. The defender follows this transition graph to protect the target.

A pure strategy for the defender is defined as a trajectory of this graph, e.g., the trajectory including (A, 5 min), (B, 10 min), and (C, 15 min) indicates a pure strategy for the defender. One key challenge of this representation for the defender's pure strategies is that the transition graph consists of an exponential number of trajectories, i.e., $O(N^T)$ where N is the number of location points and T is the number of time steps. To address this challenge, CASS proposes a compact representation of the defender's mixed strategy. Instead of directly computing a probability distribution over pure strategies for the defender, CASS attempts to compute the marginal probability that the defender will follow a certain edge of the transition graph, e.g., the probability of being at the node (A, 5 min) and moving to the node (B, 10 min). We show that given a discretized strategy space for the defender, *any strategy in full representation can be mapped into a compact representation* as well as *compact representation does not lead to any loss in solution quality* compared to the full representation (see Theorem 1 in **?**). This compact representation allows CASS to reformulate the resource allocation problem as computing the optimal *marginal* coverage of the defender over a number of $O(NT)$, the edges of the transition graph.

4.4 Scale-Up with Boundedly Rational Attackers

One key challenge of real-world security problems is that the attacker is boundedly rational; the attacker's target choice is nonoptimal. In SSGs, attacker bounded rationality is often modeled via behavior models such as quantal response (QR) (McFadden 1972; McKelvey and Palfrey 1995). In general, QR attempts to predict the probability the attacker will choose each target with the intuition is that the higher the expected utility at a target is, the more likely that the adversary will attack that target. Another behavioral model that was recently shown to provide higher prediction accuracy in predicting the attacker's behavior than QR is subjective utility quantal response (SUQR) (Nguyen et al. 2013). SUQR is motivated by the lens model which suggested that evaluation of adversaries over targets is based on a linear combination of multiple observable features (Brunswik 1952). We provide a detailed discussion on modeling and learning the attacker's behavioral model in Sect. Addressing Bounded Rationality and Bounded Surveillance in Real-World Problems. However, even when the attacker's bounded rationality is modeled and those models are learned efficiently, handling multiple attackers with these behavioral models in the context of large defender's strategy space is computational challenge. Therefore in this section, we mainly focus on handling the scalability problem given behavioral models of the attacker.

To handle the problem of large defender's strategy space given behavioral models of attackers, we introduce yet another technique of scaling up, which is similar to the incremental strategy generation. Instead, here we use incremental marginal space refinement. We use the compact marginal representation, discussed earlier, but refine that space incrementally if the solution produces violates the necessary constraints.

Domain Example – Fishery Protection for US Coast Guard Fisheries are a vital natural resource from both an ecological and economic standpoint. However, fish stocks around the world are threatened with collapse due to illegal, unreported, and unregulated (IUU) fishing. The US Coast Guard (USCG) is tasked with the responsibility of protecting and maintaining the nation's fisheries. To this end, the USCG deploys resources (both air and surface assets) to conduct patrols over fishery areas in order to deter and mitigate IUU fishing. Due to the large size of these patrol areas and the limited patrolling resources available, it is impossible to protect an entire fishery from IUU fishing at all times. Thus, an intelligent allocation of patrolling resources is critical for security agencies like the USCG.

Natural resource conservation domains such as fishery protection raise a number of new research challenges. In stark contrast to counterterrorism settings, there is frequent interaction between the defender and attacker in these resource conservation domains. This distinction is important for three reasons. First, due to the comparatively low stakes of the interactions, rather than a handful of persons or groups, the defender must protect against numerous adversaries (potentially hundreds or even more), each of which may behave differently. Second, frequent interactions make it possible to collect data on the actions of the adversary actions over time. Third, the adversaries are less strategic given the short planning windows between actions.

Algorithmic Solution – Incremental Constraint Generation (MIDAS). Generating effective strategies for domains such as fishery protection requires an algorithmic approach which is both *scalable* and *robust*. For scalability, the defender is responsible for protecting a large patrol area and therefore must consider a large strategy space. Even if the patrol area is discretized into a grid or graph structure, the defender must still reason over an exponential number of patrol strategies. For robustness, the defender must protect against *multiple* boundedly rational adversaries. Bounded rationality models, such as the quantal response (QR) model (McKelvey and Palfrey 1995) and the subjective utility quantal response (SUQR) model (Nguyen et al. 2013), introduce stochastic actions, relaxing the strong assumption in classical game theory that all players are perfectly rational and utility maximizing. These models are able to better predict the actions of human adversaries and thus lead the defender to choose strategies that perform better in practice. However, both QR and SUQR are nonlinear models resulting in a computationally difficult optimization problem for the defender. Combining these factors, MIDAS models a population of boundedly rational adversaries and utilizes available data to learn the behavior models of the adversaries using the subjective utility quantal response (SUQR) model in order to improve the way the defender allocates its patrolling resources.

Previous work on boundedly rational adversaries has considered the challenges of scalability and robustness separately, by Yang et al. (2012, 2013a) and Yang et al. (2014), Haskell et al. (2014), respectively. The MIDAS algorithm was introduced to merge these two research threads for the first time by addressing scalability and robustness simultaneously. Figure 28.6 provides a visual overview of how MIDAS operates as an iterative process. Similar to the ASPEN algorithm described earlier, given the sheer complexity of the game being solved, the problem is decomposed using a master-slave formulation. The master utilizes multiple simplifications to create a relaxed version of the original problem which is more efficient to solve. First, a piecewise-linear approximation of the security game is taken to make the optimization problem both linear and convex. Second, the complex spatiotemporal constraints associated with patrols are initially ignored and then incrementally added

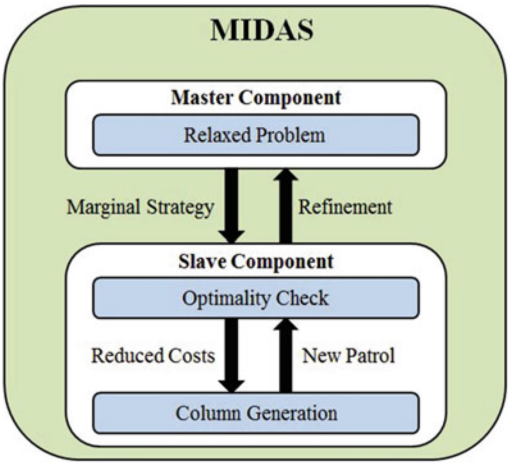


Fig. 28.6 Overview of the multiple iterative process within the MIDAS algorithm

back using cut generation. In other words, we ignore the spatiotemporal constraint that a patroller cannot simply appear and disappear at different locations instantaneously and that a patroller must pass through regions connecting two different regions if the patroller is going from one region to another. This significantly simplifies the master problem.

Due to the relaxations, solving the master produces a marginal strategy **x** which is a probability distribution over targets. However, the defender ultimately needs a probability distribution over patrols. Additionally, since not all of the spatiotemporal constraints are considered in the master, the relaxed solution **x** may not be a feasible solution to the original problem. Therefore, the slave checks if the marginal strategy **x** can expressed as a linear combination, i.e., probability distribution, of patrols. Otherwise, the marginal distribution is infeasible for the original problem. However, given the exponential number of patrol strategies, even performing this optimality check is intractable. Thus, column generation is used *within* the slave where only a small set of patrols is considered initially in the optimality check and the set is expanded over time. Much like previous examples of column generation in security games, e.g., (Jain et al. 2010a), new patrols are added by solving a minimum cost network flow problem using reduced cost information from the optimality check. If the optimality check fails, then the slave generates a cut which is returned to refine and constrain the master, incrementally bringing it closer to the original problem. The entire process is repeated until an optimal solution is found. Finally, MIDAS has been successfully deployed and evaluated by the USCG in the Gulf of Mexico.

4.5 Scale-Up with Fine-Grained Spatial Information

Discretization is a standard way to convert a continuous problem to a discrete problem. Therefore, a grid map is often used to describe a large area. However, when fine-grained spatial information needs to be considered, each cell in the grid map should be of small size, and the total number of cells is large, which leads to a significant challenge in scalability in security games especially when scheduling constraints need to be satisfied. In this section, we introduce a hierarchical modeling approach for problems with fine-grained spatial information, which is used in the context of designing foot patrols in area with complex terrain (Fang et al. 2016).

Domain Example – Wildlife Protection for Area with Complex Terrain There is an urgent need to protect wildlife from poaching. Indeed, poaching can lead to extinction of species and destruction of ecosystems. For example, poaching is considered a major driver (Chapron et al. 2008) of why tigers are now found in less than 7% of their historical range (Sanderson et al. 2006), with three out of nine tiger subspecies already extinct (IUCN 2015). As a result, efforts have been made by law enforcement agencies in many countries to protect endangered animals; the most commonly used approach is conducting foot patrols. However, given their limited human resources, improving the efficiency of patrols to combat poaching remains a major challenge.

Fig. 28.7 Patrols through a forest in Malaysia. (**a**) A foot-patrol (Malaysia). (**b**) Illegal camp sign

While game-theoretic framework can be used to address this challenge, the complex terrain of the patrolled area introduces additional complexity. In many conservation areas, high changes in elevation and the existence of large water bodies may result in a big difference in the effort needed for patrollers' movement. These factors also have a direct effect on poachers' movement. Therefore, when designing defender strategies, it is important to incorporate such topographic information. Figure 28.7a shows a sample foot patrol through a forest in Malaysia and the difficulty of conducting these patrols due to topographical constraints. Figure 28.7b shows illegal camping signs observed during those foot patrols. To generate patrol routes that contain detailed information for the patrollers and are compatible with the terrain, a fine-grained discretization of the area is necessary, leading to a large grid map of the area. On the other hand, the number of feasible routes is exponential to the number of discretized cells in the grid map due to the practical scheduling constraints such as patrol time limit and starting and ending at the base camp. Therefore, computing the optimal patrolling strategy is exceptionally computationally challenging.

Algorithmic Solution – Hierarchical Modeling Approach The hierarchical modeling approach allows us to attain a good compromise between scaling up and providing detailed guidance. This approach would be applicable in many other domains for large open area patrolling where security games are applicable, not only other green security games applications, but others including patrolling of large warehouse areas or large open campuses via robots or UAVs.

We leverage insights from hierarchical abstraction for heuristic search such as path planning (Botea et al. 2004) and apply two levels of discretization to the area of interest. We first discretize the area into large-sized *grid cells* and treat every grid cell as a target. We further discretize the grid cells into small-sized *raster pieces* and describe the spatial information for each raster piece. The defender actions are patrol routes defined over a virtual "street map" – which is built in the terms of raster pieces while aided by the grid cells in this abstraction as described below. With this hierarchical modeling, the model keeps a small number of targets and reduces the number of patrol routes while allowing for details at a fine-grained scale. The street map is a graph consisting of nodes and edges, where the set of nodes is a small subset of the raster pieces and edges are sequences of raster pieces linking the nodes. We denote nodes as key access points (KAPs) and edges as route segments. While designing foot patrols in areas with complex terrain, the street map not only helps scalability but also allows us to focus patrolling on preferred terrain features such as ridgelines which patrollers find easier to move around and are important conduits for certain mammal species such as tigers.

The street map is built in three steps: (i) determine the accessibility type for each raster piece, (ii) define KAPs, and (iii) find route segments to link the KAPs. In the first step, we check the accessibility type of every raster piece. In the example domain, raster pieces in a lake are inaccessible, whereas raster pieces on ridge lines or previous patrol tracks are easily accessible. In other domains, the accessibility of a raster piece can be defined differently. The second step is to define a set of KAPs, via which patrols will be routed. We want to build the street map in such a way that each grid cell can be reached. So we first choose raster pieces which can serve as entries and exits for the grid cells as KAPs, i.e., the ones that are on the boundary of grid cells and are easily accessible. In addition, we consider existing base camps and mountain tops as KAPs as they are key points in planning the patroller's route. We choose additional KAPs to ensure KAPs on the boundary of adjacent cells are paired. Figure 28.8 shows identified KAPs and easily accessible pieces (black and grey raster pieces respectively). The last step is to find route segments to connect the KAPs. Instead of inefficiently finding route segments to connect each pair of KAPs on the map globally, we find route segments locally for each pair of KAPs within the same grid cell, which is sufficient to connect all the KAPs. When finding the route segment, we design a distance measure which estimates the actual patrol effort according to the accessibility type of the raster pieces. Given the distance measure,

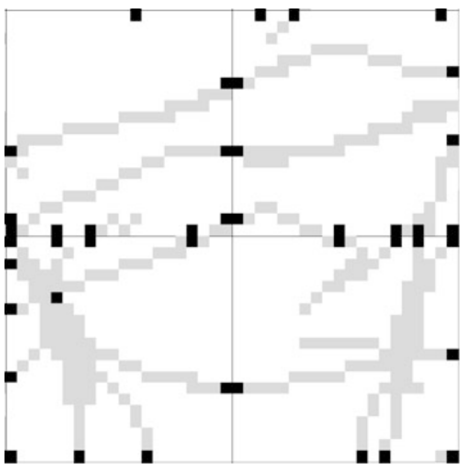


Fig. 28.8 KAPs (*black*) for 2 by 2 grid cells

the route segment is defined as the shortest distance path linking two KAPs within the grid cell.

The defender's pure strategy is defined as a patrol route on the street map, starting from the base node, walking along route segments, and ending with the base node, with its total distance satisfying the patrol distance limit. The defender's goal is to find an optimal mixed patrol strategy – a probability distribution over patrol routes. Based on the street map concept, we use a cutting-plane approach (Yang et al. 2013b) that is similar to MIDAS; specifically, in the master component, we use ARROW (Nguyen et al. 2015) algorithm to handle payoff uncertainty using the concept of minimax regret and in the slave component, we also use optimality check and column generation, and in generating new column (new patrol), we use a random selection approach over the street map. This framework is the core of the PAWS (Protection Assistant for Wildlife Security) application. Collaborating with two NGOs (Panthera and Rimba), PAWS has been deployed in Malaysia for tiger conservation.

5 Addressing Uncertainty in Real-World Problems

The standard security game model features a number of strong assumptions including that the defender has perfect information about the game payoff matrix as well as the attacker's behavorial model. Additionally, the defender is assumed to be capable of exactly executing the computed patrolling strategy. However, uncertainty is endemic in real-world security domains and thus is may be impossible or impractical for the defender to the accurately estimate various aspects of the game. Also, there are any number of practicalities and unforeseen events that may force the defender to change their patrolling strategy. These types of uncertainty can significantly deteriorate the effectiveness of the defender's strategy and thus addressing uncertainty when generating strategies is a key challenge of solving real-world security problems. This section describes several approaches for dealing with various types of uncertainties in SSGs.

We first summarize the major types of uncertainties in SSGs as a three-dimensional uncertainty space with the following three dimensions (Fig. 28.9): (1) uncertainty in the adversary's payoffs, (2) uncertainty related to the defender's strategy (including uncertainty in the defender's execution and the attacker's observation), and (3) uncertainty in the adversary's rationality. These dimensions refer to three key attributes which affect both players' utilities. The origin of the uncertainty space corresponds to the case with no uncertainty. Figure 28.9 also shows existing algorithms for addressing uncertainty in SSGs which follow the two different approaches:

(1) applying robust optimization techniques using uncertainty intervals to represent uncertainty in SSGs. For example, BRASS (Pita et al. 2009b) is a robust algorithm that only addresses attacker-payoff uncertainty, RECON (Yin et al. 2011) is another robust algorithm that focuses on addressing defender-strategy uncertainty, and monotonic maximin (Jiang et al. 2013b) is to handle the uncertainty in the attacker's bounded rationality. Finally,

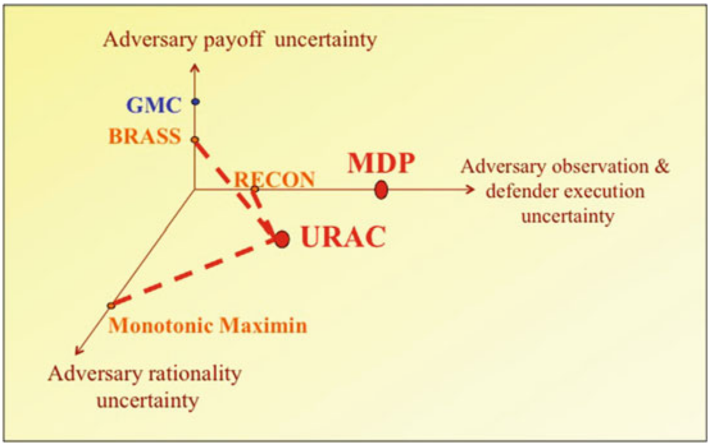


Fig. 28.9 Uncertainty space and algorithms

URAC (Nguyen et al. 2014) is a unified robust algorithm that handles all types of uncertainty; and (2) following Bayesian Stackelberg game model with dynamic execution uncertainty in which the uncertainty is represented using Markov decision process (MDP) where the time factor is incorporated.

In the following, we present two algorithmic solutions which are the representatives of these two approaches: URAC – a unified robust algorithm to handle all types of uncertainty with uncertainty intervals – and the MDP-based algorithm to handle execution uncertainty with an MDP representation of uncertainty.

5.1 Security Patrolling with Unified Uncertainty Space

Domain Example – Security in Los Angeles International Airport. Los Angeles International Airport (LAX) is the largest destination airport in the US and serves 60–70 million passengers per year. The LAX police use diverse measures to protect the airport, which include vehicular checkpoints, police units patrolling the roads to the terminals, patrolling inside the terminals (with canines), and security screening and bag checks for passengers. The application of our game-theoretic approach is focused on two of these measures: (1) placing vehicle checkpoints on inbound roads that service the LAX terminals, including both location and timing, and (2) scheduling patrols for bomb-sniffing canine units at the different LAX terminals. The eight different terminals at LAX have very different characteristics, like physical size, passenger loads, international versus domestic flights, etc. These factors contribute to the differing risk assessments of these eight terminals. Furthermore, the numbers of available vehicle checkpoints and canine units are limited by resource constraints. Thus, it is challenging to optimally allocate these resources to improve their effectiveness while avoiding patterns in the scheduled deployments (Fig. 28.10).

Fig. 28.10 LAX checkpoints are deployed using ARMOR

The ARMOR system (Assistant for Randomized Monitoring over Routes) focuses on two of the security measures at LAX (checkpoints and canine patrols) and optimizes security resource allocation using Bayesian Stackelberg games. Take the vehicle checkpoints model as an example. Assuming that there are n roads, the police's strategy is placing $m < n$ checkpoints on these roads where m is the maximum number of checkpoints. ARMOR randomizes allocation of checkpoints to roads. The adversary may conduct surveillance of this mixed strategy and may potentially choose to attack through one of these roads. ARMOR models different types of attackers with different payoff functions, representing different capabilities and preferences for the attacker. ARMOR has been successfully deployed since August 2007 at LAX (Jain et al. 2010b).

Although standard SSG-based solutions (i.e., DOBSS) have been demonstrated to improve the defender's patrolling effectiveness significantly, there remains potential improvements that can be made to further enhance the quality of such solutions such as taking uncertainties in payoff values, in the attacker's rationality, and in defender's execution into account. Therefore, we propose the unified robust algorithm, URAC, to handle these types of uncertainties by maximizing the defender's utility against the worst-case scenario resulting from these uncertainties.

Algorithmic Solution – Uncertainty Dimension Reduction (URAC). In this section, we present the robust URAC (Unified Robust Algorithmic framework for addressing unCertainties) algorithm for addressing a combination of all uncertainty types (Nguyen et al. 2014). Consider an SSG where there is uncertainty in the attacker's payoff, the defender's strategy (including the defender's execution and the attacker's observation), and the attacker's behavior, URAC represents all these uncertainty types (except for the attacker's behaviors) using uncertainty intervals. Instead of knowing exactly values of these game attributes, the defender only has prior information w.r.t the upper bounds and lower bounds of these attributes. For example, the attacker's reward if successfully attacking a target t is known to lie within the interval $[1, 3]$. Furthermore, URAC assumes the attacker monotonically responds to the defender's strategy. In other words, the higher the expected utility of a target, the more likely that the attacker will attack that target; however, the precise attacking probability is unknown for the defender. This monotonicity assumption is motivated by the quantal response model – a well-known human behavioral model for capturing the attacker's decision-making (McKelvey and Palfrey 1995).

Based on these uncertainty assumptions, URAC attempts to compute the optimal strategy for the defender by maximizing her utility against the worst-case scenario of uncertainty. The key challenge of this optimization problem is that it involves several types of uncertainty, resulting in multiple minimization steps for determining the worst-case scenario. Nevertheless, URAC introduces a unified representation of all these uncertainty types as an uncertainty set of attacker's responses. Intuitively, despite of any type of uncertainty mentioned above, what finally affects the defender's utility is the attacker's response, which is unknown to the defender due to uncertainty. As a result, URAC can represent the robust optimization problem as a single maximin problem.

However, the infinite uncertainty set of the attacker's responses depends on the planned mixed strategy for the defender, making this maximin problem difficult to solve if directly applying the traditional method (i.e., taking the dual maximization of the inner minimization of maximin and merging it with the outer maximization – maximin now can be represented a single maximization problem). Therefore, URAC proposes a divide-and-conquer method in which the defender's strategy set is divided into subsets such that the uncertainty set of the attacker's responses is the same for every defender strategy within each subset. This division leads to multiple sub-maximin problems which can be solved by using the traditional method. The optimal solution of the original maximin problem is now can be computed as a maximum over all the sub-maximin problems.

5.2 Security Patrolling with Dynamic Execution Uncertainty

Domain Example – TRUSTS for Security in Transit Systems. Urban transit systems face multiple security challenges, including deterring fare evasion, suppressing crime and counterterrorism. In particular, in some urban transit systems, including the Los Angeles Metro Rail system, passengers are legally required to purchase tickets before entering but are not physically forced to do so (Fig. 28.11). Instead, security personnel are dynamically deployed throughout the transit system, randomly inspecting passenger tickets. This proof-of-payment fare collection method is typically chosen as a more cost-effective alternative to direct fare collection, i.e.,

Fig. 28.11 TRUSTS for transit systems. (**a**) Los Angeles Metro. (**b**) Barrier-free entrance to transit system

when the revenue lost to fare evasion is believed to be less than what it would cost to directly preclude it. In the case of Los Angeles Metro, with approximately 300,000 riders daily, this revenue loss can be significant; the annual cost has been estimated at $5.6 million (Hamilton 2007). The Los Angeles Sheriff's Department (LASD) deploys uniformed patrols on board trains and at stations for fare checking (and for other purposes such as crime prevention). The LASD's current approach relies on humans for scheduling the patrols, which places a tremendous cognitive burden on the human schedulers who must take into account all of the scheduling complexities (e.g., train timings, switching time between trains, and schedule lengths).

The TRUSTS system (Tactical Randomization for Urban Security in Transit Systems) models the patrolling problem as a leader-follower Stackelberg game (Yin et al. 2012). The leader (LASD) precommits to a mixed strategy patrol (a probability distribution over all pure strategies), and riders observe this mixed strategy before deciding whether to buy the ticket or not. Both ticket sales and fines issued for fare evasion translate into revenue for the government. Therefore, the utility for the leader is the total revenue (total ticket sales plus penalties). The main computational challenge is the exponentially many possible patrol strategies, each subject to both the spatial and temporal constraints of travel within the transit network under consideration. To overcome this challenge, TRUSTS uses a compact representation of the strategy space which captures the spatiotemporal structure of the domain.

The LASD conducted field tests of this TRUSTS system in the LA Metro in 2012, and one of the feedback comments from the officers was that patrols are often interrupted due to execution uncertainty such as emergencies and arrests.

Algorithmic Solution – Marginal MDP Strategy Representation Utilizing techniques from planning under uncertainty (in particular Markov decision processes), we proposed a general approach to dynamic patrolling games in uncertain environments, which provides patrol strategies with contingency plans (Jiang et al. 2013a). This led to schedules now being loaded onto smartphones and given to officers. If interruptions occur, the schedules are then automatically updated on the smartphone app. The LASD has conducted successful field evaluations using the smartphone app, and the TSA is currently evaluating it toward nationwide deployment. We now describe the solution approach in more detail. Note that the targets, e.g., trains normally follow predetermined schedules; thus, timing is an important aspect which determines the effectiveness of the defender's patrolling schedules (the defender needs to be at the right location at a specific time in order to protect these moving targets). However, as a result of execution uncertainty (e.g., emergencies or errors), the defender could not carry out her planned patrolling schedule in later time steps. For example, in real-world trials for TRUSTS carried out by Los Angeles Sheriff's Department (LASD), there is interruption (due to writing citations, felony arrests, and handling emergencies) in a significant fraction of the executions, causing the officers to miss the train they are supposed to catch as following the pre-generated patrolling schedule.

In this section, we present the Bayesian Stackelberg game model for security patrolling with dynamic execution uncertainty introduced by Jiang et al. (2013a) in which the uncertainty is represented using Markov decision processes (MDP). The key advantage of this game-theoretic model is that patrol schedules which are computed based on Stackelberg equilibrium have contingency plans to deal with interruptions and are robust against execution uncertainty. Specifically, the security problem with execution uncertainty is represented as a two-player Bayesian Stackelberg game between the defender and the attacker. The defender has multiple patrol units, while there are also multiple types of attackers which are unknown to the defender. A (naive) patrol schedule consists of a set of sequenced commands in the following form: at time t, the patrol unit should be at location l and execute patrol action a. This patrol action a will take the unit to the next location and time if successfully executed. However, due to execution uncertainty, the patrol unit may end up at a different location and time. Figure 28.12 shows an example of execution uncertainty in a transition graph where if the patrol unit is currently at location A at the 5-min time step, she is supposed to take the on-train action to move to location B in the next time step. However, unlike CASS for ferry protection in which the defender's action is deterministic, there is a 10% chance that she will still stay at

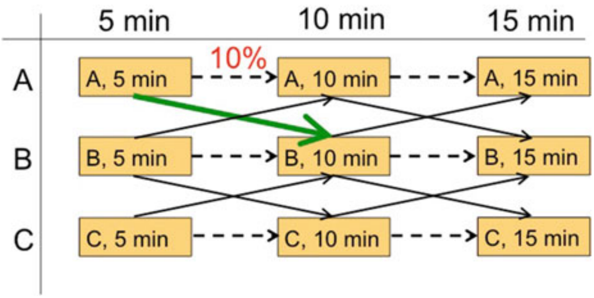


Fig. 28.12 An example of execution uncertainty in a transition graph

location A due to execution uncertainty. This interaction of the defender with the environment when executing patrol can be represented as an MDP.

In essence, the transition graph as represented above is augmented to indicate the possibility that there are multiple uncertain outcomes possible from a given state. Solving this transition graph results in marginals over MDP policies. When a sample MDP policy is obtained and loaded on to a smartphone, it provides a patroller not only the current action but contingency actions should the current action fail or succeed. So the MDP policy provides options for the patroller, allowing the system to handle execution uncertainty. A key challenge of computing the SSE for this type of security problem is that the dimension of the space of mixed strategies for the defender is exponential in the number of states in terms of the defender's times and locations. Therefore, instead of directly computing the mixed strategy, the defender attempts to compute the marginal probabilities of each patrolling unit reaching a state $s = (t, l)$ and taking action a which have dimensions polynomial in the sizes of the MDPs (the details of this approach are provided in Jiang et al. 2013a).

6 Addressing Bounded Rationality and Bounded Surveillance in Real-World Problems

Game theory models the strategic interactions between multiple players who are often assumed to be perfectly rational, i.e., they will always select the optimal strategy available to them. This assumption may be applicable for high-stakes security domains such as infrastructure protection where presumably the adversary will conduct careful surveillance and planning before attacking. However, there are other security domains where the adversary may not be perfectly rational due to short planning windows or because the adversary is less strategic due to lower stakes associated with attacking. Security strategies generated under the assumption of a perfectly rational adversary are not necessarily as effective as would be feasible against a less-than-optimal response.

In addition to bounded rationality, attackers' bounded surveillance also needs to be considered in real-world domains. In previous sections, a one-shot Stackelberg security game model is used, and it is assumed that the adversaries will conduct extensive surveillance to get a perfect understanding of the defender's strategy before an attack. However, this assumption does not apply to real-world domains involving frequent and repeated attacks. In carrying out frequent attacks, the attackers generally do not conduct extensive surveillance before performing an attack, and therefore the attackers' understanding of the defender strategy may not be up-to-date. As will be shown later in this section, if the bounded surveillance of attackers is known to the defender, the defender can exploit it to improve her average expected utility by carefully planning changes in her strategy. The improvement may depend on the level of bounded surveillance and the defender's correct understanding of the bounded surveillance. Therefore, addressing the human adversaries' boundedly rationality and bounded surveillance is a fundamental challenge for applying security games to a wide variety of domains.

Domain Example – Green Security Domains. A number of our newer applications are focused on resource conservation, through suppression of environmental crime. One area is protecting forests (Johnson et al. 2012), where we must protect a continuous forest area from extractors by patrols through the forest that seek to deter such extraction activity. With limited resources for performing such patrols, a patrol strategy will seek to distribute the patrols throughout the forest, in space and time, in order to minimize the resulting amount of extraction that occurs or maximize the degree of forest protection. This problem can be formulated as a Stackelberg game, and the focus is on computing optimal allocations of patrol density (Johnson et al. 2012) (Fig. 28.13).

As mentioned earlier, endangered species poaching is reaching critical levels as the populations of these species plummet to unsustainable numbers. The global tiger population, for example, has dropped over 95% from the start of the 1900s and has resulted in three out of nine species extinctions. Depending on the area and animals poached, motivations for poaching range from profit to sustenance, with the former being more common when profitable species such as tigers, elephants, and rhinos are the targets. To counter poaching efforts and to rebuild the species' populations, countries have set up protected wildlife reserves and conservation agencies tasked with defending these large reserves. Because of the size of the reserves and the common lack of law enforcement resources, conservation agencies are at a significant disadvantage when it comes to deterring and capturing poachers. Agencies use patrolling as a primary method of securing the park. Due to their limited resources, however, patrol managers must carefully create patrols that account for many different variables (e.g., limited patrol units to send out, multiple locations that poachers can attack at varying distances to the outpost).

6.1 Bounded Rationality Modeling and Learning

Recently, we have conducted some research on applying ideas from behavioral game theory (e.g., prospect theory (Kahneman and Tversky 1979) and quantal response (McFadden 1976)) within security game algorithms. One line of

Fig. 28.13 Examples of illegal activities in green security domains. (**a**) An illegal trapping tool. (**b**) Illegally cutting trees

approaches is based on the quantal response model to predict the behaviors of the human adversary and then to compute optimal defender strategies against such behavior of the adversary. These include BRQR (Yang et al. 2011) which follows the logit quantal response (QR) (McFadden 1976) model and subsequent work on subjective-utility quantal response (SUQR) models (Nguyen et al. 2013). The parameters of these models are estimated by experimental tuning. Data from a large set of participants on the Amazon Mechanical Turk (AMT) were collected and used to learn the parameters of the behavioral models to predict future attacks. In real-world domains like fisheries protection or wildlife crime, there are repeated interactions between the defender and the adversary, where the game progresses in "rounds." We call this a repeated Stackelberg security game (RSSG) wherein each round the defender would play a particular strategy and the adversary would observe that strategy and act accordingly. In order to simulate this scenario and conduct experiments to identify adversary behavior in such repeated settings, an online RSSG game was developed (shown in Fig. 28.14) and deployed.

Wildlife Poaching Game: In our game, human subjects play the role of poachers looking to place a snare to hunt a hippopotamus in a protected wildlife park. The portion of the park shown in the map is actually a Google Maps view of a portion of the Queen Elizabeth National Park (QENP) in Uganda. The region shown is divided into a 5*5 grid, i.e., 25 distinct cells. Overlaid on the Google Maps view of the park is a heat map, which represents the rangers' mixed strategy x – a cell i

Fig. 28.14 Interface of the Wildlife Poaching game to simulate an RSSG

with higher coverage probability x_i is shown more in red, while a cell with lower coverage probability is shown more in green. As the subjects play the game and click on a particular region on the map, they were given detailed information about the poacher's reward, penalty, and coverage probability at that region: R_i^a, P_i^a, and x_i for each target i. However, the participants are unaware of the exact location of the rangers while playing the game, i.e., they do not know the pure strategy that will be played by the rangers, which is drawn randomly from mixed strategy x shown on the game interface. Thus, we model the real-world situation that poachers have knowledge of past pattern of ranger deployment but not the exact location of ranger patrols when they set out to lay snares. In our game, there were nine rangers protecting this park, with each ranger protecting one grid cell. Therefore, at any point in time, only 9 out of the 25 distinct regions in the park are protected. A player succeeds if he places a snare in a region which is not protected by a ranger, else he is unsuccessful.

Similar to Nguyen et al. (2013), here also we recruited human subjects on AMT and asked them to play this game repeatedly for a set of rounds with the defender strategy changing per round based on the behavioral model being used to learn the adversary's behavior. Before we discuss more about the experiments conducted, we first give a brief overview of the bounded rationality models used in our experiments to learn adversary behavior.

Bounded Rationality Models: Subjective utility quantal response (SUQR) (Nguyen et al. 2013) is a behavioral model which builds upon prior work on quantal response (QR) (McFadden 1976) according to which rather than strictly maximizing utility, an adversary stochastically chooses to attack targets, i.e., the adversary attacks a target with higher expected utility with a higher probability. SUQR proposes a new utility function called subjective utility, which is a linear combination of key features that are considered to be the most important in each adversary decision-making step. Nguyen et al. (2013) experimented with three features: defender's coverage probability, adversary's reward and penalty (x_i, R_i^a, P_i^a) at each target i. According to this model, the probability that the adversary will attack target $i \in \mathbb{T}$ is given by:

$$q_i(\omega|x) = \frac{e^{SU_i^a(x)}}{\sum_{j \in \mathbb{T}} e^{SU_j^a(x)}} \tag{28.18}$$

where $SU_i^a(x)$ is the subjective utility of an adversary for attacking target i when defender employs strategy x and is given by:

$$SU_i^a(x) = \omega_1 x_i + \omega_2 R_i^a + \omega_3 P_i^a \tag{28.19}$$

The vector $\omega = (\omega_1, \omega_2, \omega_3)$ encodes information about the adversary's behavior, and each component of ω indicates the relative importance the adversary gives to each attribute in the decision-making process. The weights are computed by performing maximum likelihood estimation (MLE) on available attack data.

While behavioral models like QR (McFadden 1976) and SUQR (Nguyen et al. 2013) assume that there is a homogeneous population of adversaries, in the real world, we face heterogeneous populations of adversaries. Therefore Bayesian SUQR was proposed to learn the behavioral model for each attack (Yang et al. 2014). Protection Assistant for Wildlife Security (PAWS) is an application which was originally created using Bayesian SUQR. However, in real-world security domains, we may have very limited data or may only have some limited information on the biases displayed by adversaries. An alternative approach is based on robust optimization: instead of assuming a particular model of human decision-making, try to achieve good defender expected utility against a range of possible models. One instance of this approach is MATCH (Pita et al. 2012), which guarantees a bound for the loss of the defender to be within a constant factor of the adversary loss if the adversary responds nonoptimally. Another robust solution concept is monotonic maximin (Jiang et al. 2013b), which tries to optimize defender utility against the worst-case monotonic adversary behavior, where monotonicity is the property that actions with higher expected utility is played with higher probability. Recently, there has been attempts to combine such robust-optimization approaches with available behavior data (Haskell et al. 2014) for RSSGs, resulting in a new human behavior model called Robust SUQR. However, one question of research is how these proposed models and algorithms will fare against human subjects in RSSGs. This has been explored in recent research (Kar et al. 2015) in the "first-of-its-kind" human subjects experiments in RSSGs over a period of 46 weeks with the "Wildlife Poaching" game. A brief description of our experimental observations from the RSSG human subject experiments is presented below.

Results in RSSG Experiments – An Overview: In our human subject experiments in RSSGs, we observe that (i) existing approaches (QR, SUQR, Bayesian SUQR) (Haskell et al. 2014; Nguyen et al. 2013; Yang et al. 2014) perform poorly in initial rounds, while Bayesian SUQR which is the basis for PAWS (Yang et al. 2014) performs poorly throughout all rounds; and (ii) surprisingly, simpler models like SUQR which were originally proposed for single-shot games performed better than recent advances like Bayesian SUQR and Robust SUQR which are geared specifically toward addressing repeated SSGs. These results are shown in Fig. 28.16a–d. Therefore, we proposed a new model called SHARP (Stochastic Human behavior model with AttRactiveness and Probability weighting) (Kar et al. 2015) which is specifically suited for dynamic settings such as RSSGs. SHARP addresses the limitations of the existing models in the following way: (i) modeling the adversary's adaptive decision-making process in repeated SSGs, SHARP reasons based on success, or failure of the adversary's past actions on exposed portions of the attack surface, where attack surface is defined as the n-dimensional space of the features used to model adversary behavior; (ii) addressing limited exposure to significant portions of the attack surface in initial rounds, SHARP reasons about similarity between exposed and unexposed areas of the attack surface, and also incorporates a discounting parameter to mitigate adversary's lack of exposure to enough of the attack surface; (iii) addressing the limitation that existing models do not account for the adversary's weighting of probabilities, we incorporate a two parameter probability weighting function. We discuss these three modeling aspects of SHARP.

SHARP – Probability Weighting: SHARP has three key novelties, of which we discuss probability weighting first. The need for probability weighting became apparent when we observed based on our initial experiments with existing models (Haskell et al. 2014; Nguyen et al. 2013; Yang et al. 2014) that the weight on coverage probability was positive for experiments. That is, counterintuitively humans were modeled as being attracted to cells with high coverage probability, even though they were *not* attacking targets with very high coverage, but they were going after targets with moderate to very low coverage probability. We hypothesize that this counterintuitive result of a model with $\omega_1 > 0$ may be because the SUQR model may not be considering people's *actual* weighting of probability. SUQR assumes that people weigh probabilities of events in a linear fashion, while existing work on probability weighting (Kahneman and Tversky 1979; Tversky and Kahneman 1992) suggests otherwise. To address this issue, we augment the subjective utility function with a two-parameter probability weighting function (Eq. 28.20) proposed by Gonzalez and Wu (1999) that can be either inverse S-shaped (concave near probability zero and convex near probability one) or S-shaped.

$$f(p) = \frac{\delta p^{\gamma}}{\delta p^{\gamma} + (1-p)^{\gamma}} \tag{28.20}$$

The SU of an adversary denoted by "a" can then be computed as:

$$SU_i^a(x) = \omega_1 f(x_i) + \omega_2 R_i^a + \omega_3 P_i^a \tag{28.21}$$

where $f(x_i)$ for coverage probability x_i is computed as per Eq. 28.20.

One of our key findings is that the curve representing human weights for probability is *S-shaped in nature and not inverse S-shaped* as prospect theory suggests. The S-shaped curve indicates that people would overweight high probabilities and underweight low to medium probabilities. An example of learned curves on our data over several rounds of the RSSG experiment is shown in Fig. 28.15. Recent studies (Alarie and Dionne 2001; Etchart-Vincent 2009; Humphrey and Verschoor 2004) have also found S-shaped probability curves which contradict the inverse S-shaped observation of prospect theory. Given S-shaped probability weighting functions, the learned ω_1 was negative as it accurately captured the trend that a significantly higher number of people were attacking targets with low to medium coverage probabilities and *not* attacking high coverage targets.

SHARP – Adaptive Utility Function: A second major innovation in SHARP is the adaptive nature of the adversary and addressing the issue of attack surface exposure where ***attack surface*** α is defined as the n-dimensional space of the features used to model adversary behavior. A ***target profile*** $\beta_k \in \alpha$ is defined as a point on the attack surface α and can be associated with a target. Exposing the adversary to a lot of different target profiles would therefore mean exposing the

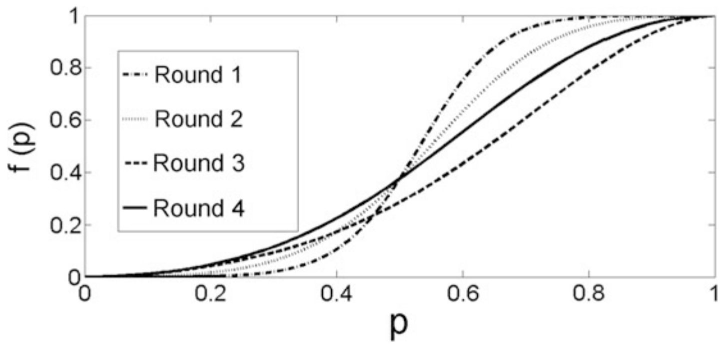


Fig. 28.15 Probability curves from rounds 1 to 4

adversary to more of the attack surface and gathering valuable information about their behavior. While a particular target location, defined as a distinct cell in the 2D space, can only be associated with one target profile in a particular round, more than one target may be associated with the same target profile in the same round. β_k^i denotes that target profile β_k is associated with target i in a particular round. Below is an observation from our human subjects data that reveal interesting trends in attacker behavior in RSSGs.

Observation 1. *Consider two sets of adversaries: (i) those who have succeeded in attacking a target associated with a particular target profile in one round and (ii) those who have failed in attacking a target associated with a particular target profile in the same round. In the subsequent round, the first set of adversaries are significantly more likely than the second set of adversaries to attack a target with a target profile which is "similar" to the one they attacked in the earlier round.*

Now, existing models only consider the adversary's actions from round $(r-1)$ to predict their actions in round r. However, based on our observation (Obs. 1), it is clear that the adversary's actions in a particular round are dependent on his past successes and failures. The *adaptive* probability weighted subjective utility function proposed in Eq. 28.22 captures this adaptive nature of the adversary's behavior in such dynamic settings by capturing the shifting trends in attractiveness of various target profiles over rounds.

$$ASU_{\beta_i}^R = (1 - d * A_{\beta_i}^R)\omega_1 f(x_{\beta_i}) + (1 + d * A_{\beta_i}^R)\omega_2 \phi_{\beta_i} + (1 + d * A_{\beta_i}^R)\omega_3 P_{\beta_i}^a + (1 - d * A_{\beta_i}^R)\omega_4 D_{\beta_i} \tag{28.22}$$

Here, $A_{\beta_i}^R$ denotes the ***attractiveness*** of a target profile β_i at the end of round R and models the attacker's current affinity toward targets he attacked in the past based on his past successes and failures. The parameter d $(0 \leq d \leq 1)$ in Eq. 28.22 is a discounting parameter which is based on a measure of the amount of attack

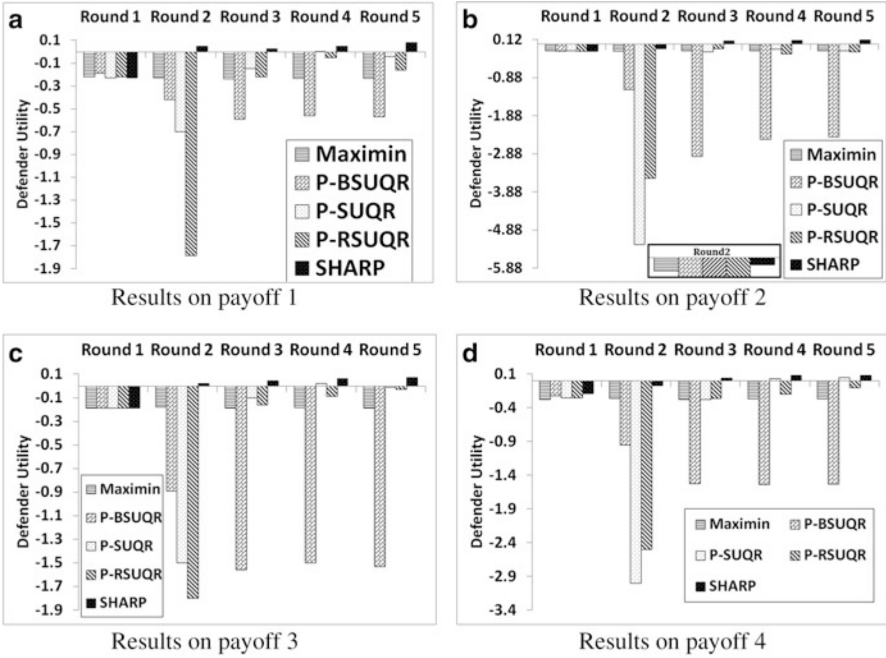


Fig. 28.16 (**a**), (**b**), (**c**), and (**d**): Defender utilities for various models on four payoff structures, respectively

surface exposed and mitigates this attack surface exposure problem. Therefore, there are three main parts to SHARP's adaptive utility computation: (i) adapting the subjective utility based on past successes and failures on exposed parts of the attack surface, (ii) discounting to handle situations where not enough attack surface has been exposed, and (ii) reasoning about similarity of unexposed portions of the attack surface based on other exposed parts of the attack surface (see Kar et al. 2015 for details).

Based on our human subjects experiments with SHARP and other models on four different payoff structures, we observe in Fig. 28.16a–d that SHARP completely outperforms existing approaches consistently over all rounds, most notably in initial rounds (refer to Kar et al. 2015 for more details about the experimental results and observations).

6.2 Bounded Surveillance Modeling and Planning

We have discussed above some of the bounded rationality models applied to RSSGs. However, sometimes the adversaries may be bounded by their surveillance capabilities. Therefore, to account for adversaries' bounded surveillance, more recent work has generalized the perfect Stackelberg assumption, and they assume that the adversaries' understanding of the defender strategy may not be up to date and can be instead approximated as a convex combination of the defender strategies used in recent rounds (Fang et al. 2015). The RSSG framework, which assumes that the attackers always have up-to-date information, can be seen as a special case of this more generalized Green Security Games (GSG) model.

More specifically, a GSG model considers a repeated game between a defender and multiple attackers. Each round corresponds to a period of time, which can be a time interval (e.g., a month) after which the defender (e.g., warden) communicate with local guards to assign them a new strategy. In each round, the defender chooses a mixed strategy at the beginning of the round. Different from RSSG, an attacker in GSG is characterized by his memory length and weights on recent rounds in addition to his SUQR model parameters. The attacker is assumed to respond to a weighted sum of the defender strategies used in recent rounds (within his memory length). The defender aims to maximize her total expected utility over all the rounds.

Due to the bounded surveillance of attackers, the defender can potentially improve her average expected utility by carefully planning changes in her strategy from round to round in a GSG. Based on the GSG model, we provide two algorithms that plan ahead – the generalization of the Stackelberg assumption introduces a need to plan ahead and take into account the effect of defender strategy on future attacker decisions. While the first algorithm plans a fixed number of steps ahead, the second one designs a short sequence of strategies for repeated execution.

For clarity of exposition, we first focus on the case where the attackers have one round memory and have no information about the defender strategy in the current round, i.e., the attackers respond to the defender strategy in the last round. To maximize her average expected utility, the defender could optimize over all rounds simultaneously. However, this approach is computationally expensive when the game has many rounds: it needs to solve a non-convex optimization problem with at least NT variables where N is the number of targets considered and T is the length of the game. An alternative is the myopic strategy, i.e., the defender can always protect the targets with the highest expected utility in the current round. However, this myopic choice may lead to significant quality degradation as it ignores the impact of current strategy in the future round.

Therefore, we propose an algorithm named PlanAhead-M (or PA-M) in Fang et al. (2015) that looks ahead a few steps. PA-M finds an optimal strategy for the current round as if it is the M^{th} last round of the game. If $M = 2$, the defender chooses a strategy assuming she will play a myopic strategy in the next round and end the game. PA-T corresponds to the optimal solution and PA-1 is the myopic strategy. Choosing $1 < M < T$ can balance the solution quality and the computation complexity.

While PA-M presents an effective way to design sequential defender strategies, we provide another algorithm called FixedSequence-M (FS-M) for GSGs in (Fang et al. 2015). FS-M not only has provable theoretical guarantees but may also ease the implementation in practice. The idea of FS-M is to find a short sequence of strategies with fixed length M and require the defender to execute this sequence repeatedly. If $M = 2$, the defender will alternate between two strategies, and she can exploit the attackers' delayed response. It can be easier to communicate with local guards to implement FS-M in green security domains as the guards only need to alternate between several types of maneuvers.

7 Addressing Field Evaluation in Real-World Problems

Evidence showing the benefits of the algorithms discussed in the previous sections is definitely an important issue that is necessary for us to answer. Unlike conceptual ideas, where we can run thousands of careful simulations under controlled conditions, it is not possible to conduct such experiments in the real world with our deployed applications. Nor is it possible to provide a proof of 100% security – there is no such thing.

Instead, we focus on the specific question of are our game-theoretic algorithms presented better at security resource optimization or security allocation than how they were allocated previously, which was typically relying on human schedulers or a simple dice roll for security scheduling (simple dice roll is often the other "automation" that is used or offered as an alternative to our methods). We have used the following methods to illustrate these ideas. These methods range from simulations to actual field tests.

1. **Simulations (including using a "machine learning" attacker)**: We provide simulations of security schedules, e.g., randomized patrols, assignments, comparing our approach to earlier approaches based on techniques used by human schedulers. We have a machine learning-based attacker who learns any patterns and then chooses to attack the facility being protected. Game-theoretic schedulers are seen to perform significantly better in providing higher levels of protections (Jain et al. 2010b; Pita et al. 2008). This is also shown in Fig. 28.17.
2. **Human adversaries in the lab**: We have worked with a large number of human subjects and security experts (security officials) to have them get through randomized security schedules, where some are schedules generated by our algorithms and some are baseline approaches for comparison. Human subjects are paid money based on the reward they collect by successfully intruding through our security schedules; again our game-theoretic schedulers perform significantly better (Pita et al. 2009a).
3. **Actual security schedules before and after**: For some security applications, we have data on how scheduling was done by humans (before our algorithms were deployed) and how schedules are generated after deployment of our algorithms. For measures of interest to security agencies, e.g., predictability in schedules, it is possible to compare the actual human-generated schedules vs our algorithmic schedules. Again, game-theoretic schedulers are seen to perform significantly better by avoiding predictability and yet ensuring that more important targets are covered with higher frequency of patrols. Some of this data is published (Shieh et al. 2012) and is also shown in Fig. 28.18.

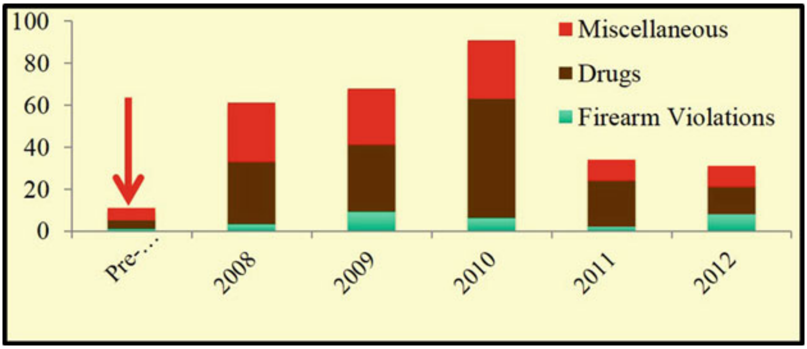


Fig. 28.17 ARMOR evaluation results

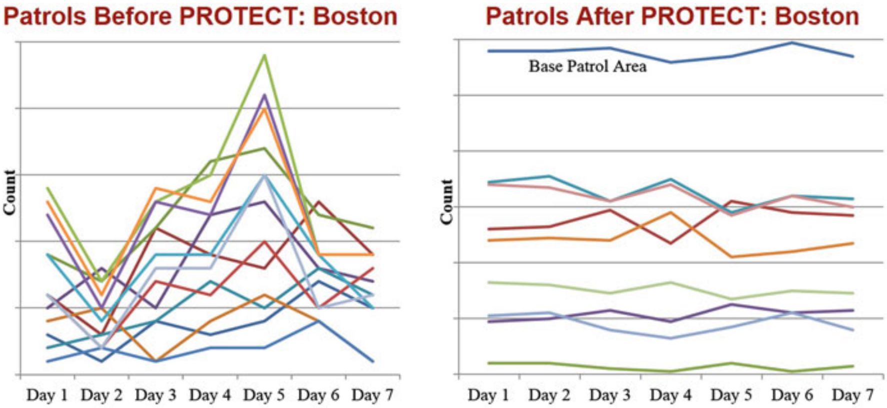


Fig. 28.18 PROTECT evaluation results: pre deployment (*left*) and post deployment patrols (*right*)

4. **"Adversary" teams simulate attack**: In some cases, security agencies have deployed adversary perspective teams or "mock attacker teams" that will attempt to conduct surveillance to plan attacks; this is done before and after our algorithms have been deployed to check which security deployments worked better. This was done by the US Coast Guard indicating that the game-theoretic scheduler provided higher levels of deterrence (Shieh et al. 2012).
5. **Real-time comparison: human vs algorithm**: This is a test we ran on the metro trains in Los Angeles. For a day of patrol scheduling, we provided head-to-head comparison of human schedulers trying to schedule 90 officers on patrols vs an automated game-theoretic scheduler. External evaluators then provided an evaluation of these patrols; the evaluators did not know who had generated each of the schedules. The results show that while human schedulers required significant effort even for generating one schedule (almost a day) and

the game-theoretic scheduler ran quickly, the external evaluators rated the game-theoretic schedulers higher (with statistical significance) (Fave et al. 2014a).

6. **Actual data from deployment**: This is another test run on the metro trains in LA. We had a comparison of game-theoretic scheduler vs an alternative (in this case a uniform random scheduler augmented with real time human intelligence) to check fare evaders. In 21 days of patrols, the game-theoretic scheduler led to significantly higher numbers of fare evaders captured than the alternative (Fave et al. 2014a,b).
7. **Domain expert evaluation (internal and external)**: There have been of course significant numbers of evaluations done by domain experts comparing their own scheduling method with game-theoretic schedulers, and repeatedly the game-theoretic schedulers have come out ahead. The fact that our software is now in use for several years at several different important airports, ports, air traffic, and so on is an indicator to us that the domain experts must consider this software of some value.

8 Conclusions

Security is recognized as a worldwide challenge, and game theory is an increasingly important paradigm for reasoning about complex security resource allocation. We have shown that the general model of security games is applicable (with appropriate variations) to varied security scenarios. There are applications deployed in the real world that have led to a measurable improvement in security. We presented approaches to address four significant challenges: scalability, uncertainty, bounded rationality, and field evaluation in security games.

In short, we introduced specific techniques to handle each of these challenges. For scalability, we introduced three approaches: (i) incremental strategy generation for addressing the problem of large defender strategy spaces, (ii) double oracle incremental strategy generation w.r.t large defender and attacker strategy spaces, (iii) compact representation of strategies for the case of mobile resources and moving targets, (iv) cutting plane (incremental constraint generation) for handling multiple boundedly rational attacker, and (v) a hierarchical approach for incorporating fine-grained spatial information. For handling uncertainty we introduced two approaches: (i) dimensionality reduction in uncertainty space for addressing a unification of uncertainties and (ii) Markov Decision Process with marginal strategy representation w.r.t dynamic execution uncertainty. In terms of handling attacker bounded rationality and bounded surveillance, we propose different behavioral models to capture the attackers' behaviors and introduce human subject experiments with game simulation to learn such behavioral models. Finally, for addressing field evaluation in real-world problems, we discussed two approaches: (i) data from deployment and (ii) mock attacker team.

While the deployed game-theoretic applications have provided a promising start, significant amount of research remains to be done. These are large-scale interdisciplinary research challenges that call upon multiagent researchers to work with researchers in other disciplines, be “on the ground” with domain experts and examine real-world constraints and challenges that cannot be abstracted away.

References

Alarie Y, Dionne G (2001) Lottery decisions and probability weighting function. J Risk Uncertain 22(1):21–33

Alpcan T, Başar T (2010) Network security: a decision and game-theoretic approach. Cambridge University Press, Cambridge/New York

An B, Tambe M, Ordonez F, Shieh E, Kiekintveld C (2011) Refinement of strong Stackelberg equilibria in security games. In: Proceedings of the 25th conference on artificial intelligence, pp 587–593

Blocki J, Christin N, Datta A, Procaccia AD, Sinha A (2013) Audit games. In: Proceedings of the 23rd international joint conference on artificial intelligence

Blocki J, Christin N, Datta A, Procaccia AD, Sinha A (2015) Audit games with multiple defender resources. In: AAAI conference on artificial intelligence (AAAI)

Botea A, Müller M, Schaeffer J (2004) Near optimal hierarchical path-finding. J Game Dev 1:7–28

Breton M, Alg A, Haurie A (1988) Sequential Stackelberg equilibria in two-person games. Optim Theory Appl 59(1):71–97

Brunswik E (1952) The conceptual framework of psychology, vol 1. University of Chicago Press, Chicago/London

Chandran R, Beitchman G (2008) Battle for Mumbai ends, death toll rises to 195. Times of India. http://articles.timesofindia.indiatimes.com/2008-11-29/india/27930171_1_taj-hotel-three-terrorists-nariman-house

Chapron G, Miquelle DG, Lambert A, Goodrich JM, Legendre S, Clobert J (2008) The impact on tigers of poaching versus prey depletion. J Appl Ecol 45:1667–1674

Conitzer V, Sandholm T (2006) Computing the optimal strategy to commit to. In: Proceedings of the ACM conference on electronic commerce (ACM-EC), pp 82–90

Durkota K, Lisy V, Kiekintveld C, Bosansky B (2015) Game-theoretic algorithms for optimal network security hardening using attack graphs. In: Proceedings of the 2015 intemational conference on autonomous agents and multiagent systems (AAMAS’15)

Etchart-Vincent N (2009) Probability weighting and the level and spacing of outcomes: an experimental study over losses. J Risk Uncertain 39(1):45–63

Fang F, Jiang AX, Tambe M (2013) Protecting moving targets with multiple mobile resources. J Artif Intell Res 48:583–634

Fang F, Stone P, Tambe M (2015) When security games go green: designing defender strategies to prevent poaching and illegal fishing. In: International joint conference on artificial intelligence (IJCAI)

Fang F, Nguyen TH, Pickles R, Lam WY, Clements GR, An B, Singh A, Tambe M, Lemieux A (2016) Deploying paws: field optimization of the protection assistant for wildlife security. In: Proceedings of the twenty-eighth innovative applications of artificial intelligence conference (IAAI 2016)

Fave FMD, Brown M, Zhang C, Shieh E, Jiang AX, Rosoff H, Tambe M, Sullivan J (2014a) Security games in the field: an initial study on a transit system (extended abstract). In: International conference on autonomous agents and multiagent systems (AAMAS) [Short paper]

Fave FMD, Jiang AX, Yin Z, Zhang C, Tambe M, Kraus S, Sullivan J (2014b) Game-theoretic security patrolling with dynamic execution uncertainty and a case study on a real transit system. J Artif Intell Res 50:321–367

Gonzalez R, Wu G (1999) On the shape of the probability weighting function. Cogn Psychol 38:129–166

Hamilton BA (2007) Faregating analysis. Report commissioned by the LA Metro. http://boardarchives.metro.net/Items/2007/11_November/20071115EMACItem27.pdf
Haskell WB, Kar D, Fang F, Tambe M, Cheung S, Denicola LE (2014) Robust protection of fisheries with compass. In: Innovative applications of artificial intelligence (IAAI)
Humphrey SJ, Verschoor A (2004) The probability weighting function: experimental evidence from Uganda, India and Ethiopia. Econ Lett 84(3):419–425
IUCN (2015) IUCN red list of threatened species. version 2015.2. http://www.iucnredlist.org
Jain M, Kardes E, Kiekintveld C, Ordonez F, Tambe M (2010a) Security games with arbitrary schedules: a branch and price approach. In: Proceedings of the 24th AAAI conference on artificial intelligence, pp 792–797
Jain M, Tsai J, Pita J, Kiekintveld C, Rathi S, Tambe M, Ordonez F (2010b) Software assistants for randomized patrol planning for the LAX airport police and the federal air marshal service. Interfaces 40:267–290
Jain M, Korzhyk D, Vanek O, Pechoucek M, Conitzer V, Tambe M (2011) A double oracle algorithm for zero-sum security games on graphs. In: Proceedings of the 10th international conference on autonomous agents and multiagent systems (AAMAS)
Jain M, Tambe M, Conitzer V (2013) Security scheduling for real-world networks. In: AAMAS
Jiang A, Yin Z, Kraus S, Zhang C, Tambe M (2013a) Game-theoretic randomization for security patrolling with dynamic execution uncertainty. In: AAMAS
Jiang AX, Nguyen TH, Tambe M, Procaccia AD (2013b) Monotonic maximin: a robust Stackelberg solution against boundedly rational followers. In: Conference on decision and game theory for security (GameSec)
Johnson M, Fang F, Yang R, Tambe M, Albers H (2012) Patrolling to maximize pristine forest area. In: Proceedings of the AAAI spring symposium on game theory for security, sustainability and health
Kahneman D, Tversky A (1979) Prospect theory: an analysis of decision under risk. Econometrica 47(2):263–291
Kar D, Fang F, Fave FD, Sintov N, Tambe M (2015) "a game of thrones": when human behavior models compete in repeated Stackelberg security games. In: International conference on autonomous agents and multiagent systems (AAMAS 2015)
Keteyian A (2010) TSA: federal air marshals. http://www.cbsnews.com/stories/2010/02/01/earlyshow/main6162291.shtml. Retrieved 1 Feb 2011
Kiekintveld C, Jain M, Tsai J, Pita J, Tambe M, Ordonez F (2009) Computing optimal randomized resource allocations for massive security games. In: Proceedings of the 8th international conference on autonomous agents and multiagent systems (AAMAS), pp 689–696
Korzhyk D, Conitzer V, Parr R (2010) Complexity of computing optimal Stackelberg strategies in security resource allocation games. In: Proceedings of the 24th AAAI conference on artificial intelligence, pp 805–810
Leitmann G (1978) On generalized Stackelberg strategies. Optim Theory Appl 26(4):637–643
Lipton R, Markakis E, Mehta A (2003) Playing large games using simple strategies. In: EC: proceedings of the ACM conference on electronic commerce. ACM, New York, pp 36–41
McFadden D (1972) Conditional logit analysis of qualitative choice behavior. Technical report
McFadden D (1976) Quantal choice analysis: a survey. Ann Econ Soc Meas 5(4):363–390
McKelvey RD, Palfrey TR (1995) Quantal response equilibria for normal form games. Games Econ Behav 10(1):6–38
Nguyen TH, Yang R, Azaria A, Kraus S, Tambe M (2013) Analyzing the effectiveness of adversary modeling in security games. In: Conference on artificial intelligence (AAAI)
Nguyen T, Jiang A, Tambe M (2014) Stop the compartmentalization: Unified robust algorithms for handling uncertainties in security games. In: International conference on autonomous agents and multiagent systems (AAMAS)
Nguyen TH, Fave FMD, Kar D, Lakshminarayanan AS, Yadav A, Tambe M, Agmon N, Plumptre AJ, Driciru M, Wanyama F, Rwetsiba A (2015) Making the most of our regrets: regret-based solutions to handle payoff uncertainty and elicitation in green security games. In: Conference on decision and game theory for security

Paruchuri P, Pearce JP, Marecki J, Tambe M, Ordonez F, Kraus S (2008) Playing games with security: an efficient exact algorithm for Bayesian Stackelberg games. In: Proceedings of the 7th international conference on autonomous agents and multiagent systems (AAMAS), pp 895–902

Pita J, Jain M, Western C, Portway C, Tambe M, Ordonez F, Kraus S, Parachuri P (2008) Deployed ARMOR protection: the application of a game-theoretic model for security at the Los Angeles International Airport. In: Proceedings of the 7th international conference on autonomous agents and multiagent systems (AAMAS), pp 125–132

Pita J, Bellamane H, Jain M, Kiekintveld C, Tsai J, Ordóñez F, Tambe M (2009a) Security applications: lessons of real-world deployment. ACM SIGecom Exchanges 8(2):5

Pita J, Jain M, Ordóñez F, Tambe M, Kraus S, Magori-Cohen R (2009b) Effective solutions for real-world Stackelberg games: when agents must deal with human uncertainties. In: The eighth international conference on autonomous agents and multiagent systems

Pita J, John R, Maheswaran R, Tambe M, Kraus S (2012) A robust approach to addressing human adversaries in security games. In: European conference on artificial intelligence (ECAI)

Sanderson E, Forrest J, Loucks C, Ginsberg J, Dinerstein E, Seidensticker J, Leimgruber P, Songer M, Heydlauff A, O'Brien T, Bryja G, Klenzendorf S, Wikramanayake E (2006) Setting priorities for the conservation and recovery of wild tigers: 2005–2015. The technical assessment. Technical report, WCS, WWF, Smithsonian, and NFWF-STF, New York/Washington, DC

Shieh E, An B, Yang R, Tambe M, Baldwin C, DiRenzo J, Maule B, Meyer G (2012) PROTECT: a deployed game theoretic system to protect the ports of the United States. In: Proceedings of the 11th international conference on autonomous agents and multiagent systems (AAMAS)

von Stackelberg H (1934) Marktform und Gleichgewicht. Springer, Vienna

von Stengel B, Zamir S (2004) Leadership with commitment to mixed strategies. Technical report, LSE-CDAM-2004-01, CDAM research report

Tambe M (2011) Security and game theory: algorithms, deployed systems, lessons learned. Cambridge University Press, Cambridge

Tversky A, Kahneman D (1992) Advances in prospect theory: cumulative representation of uncertainty. J Risk Uncertain 5(4):297–323

Vanek O, Yin Z, Jain M, Bosansky B, Tambe M, Pechoucek M (2012) Game-theoretic resource allocation for malicious packet detection in computer networks. In: Proceedings of the 11th international conference on autonomous agents and multiagent systems (AAMAS)

Yang R, Kiekintveld C, Ordonez F, Tambe M, John R (2011) Improving resource allocation strategy against human adversaries in security games. In: IJCAI

Yang R, Ordonez F, Tambe M (2012) Computing optimal strategy against quantal response in security games. In: Proceedings of the 11th international conference on autonomous agents and multiagent systems (AAMAS)

Yang R, Jiang AX, Tambe M, Ordóñez F (2013a) Scaling-up security games with boundedly rational adversaries: a cutting-plane approach. In: Proceedings of the twenty-third international joint conference on artificial intelligence. AAAI Press, pp 404–410

Yang R, Jiang AX, Tambe M, Ordóñez F (2013b) Scaling-up security games with boundedly rational adversaries: a cutting-plane approach. In: IJCAI

Yang R, Ford B, Tambe M, Lemieux A (2014) Adaptive resource allocation for wildlife protection against illegal poachers. In: International conference on autonomous agents and multiagent systems (AAMAS)

Yin Z, Jain M, Tambe M, Ordonez F (2011) Risk-averse strategies for security games with execution and observational uncertainty. In: Proceedings of the 25th AAAI conference on artificial intelligence (AAAI), pp 758–763

Yin Z, Jiang A, Johnson M, Tambe M, Kiekintveld C, Leyton-Brown K, Sandholm T, Sullivan J (2012) TRUSTS: scheduling randomized patrols for Fare Inspection in Transit Systems. In: Proceedings of the 24th conference on innovative applications of artificial intelligence (IAAI)

Index

T. Başar, G. Zaccour (eds.), *Handbook of Dynamic Game Theory*,
https://doi.org/10.1007/978-3-319-44374-4

D

E

P

T