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

$$\begin{aligned}\int_X \omega(y)q(dy|x,a,b) &= \int_0^\infty \omega((x-a)(1+\rho_0)+s)b(ds)\\ &\le \omega(x(1+\rho_0)+\bar{s}) = (r + x(1+\rho_0)+\bar{s})^\sigma.\end{aligned}$$

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 Ξ. Define

$$\overline{u}_n(\xi) = \frac{1}{n}\sum_{k=1}^{n} u_k(\xi)$$

and

$$v_\beta = \inf_{\xi\in\Xi}(1-\beta)\sum_{k=1}^{\infty}\beta^{k-1}u_k(\xi) \quad \text{for} \quad \beta \in (0,1), \quad v_n := \inf_{\xi\in\Xi}\overline{u}_n(\xi).$$

Proposition 6. *Assume that $(u_n(\xi))_{n\in\mathbb{N}} \in S_D$ for each $\xi \in \Xi$. 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 $\xi \in \Xi$. 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* ($\tilde{\mathrm{C}}1$), ($\tilde{\mathrm{W}}1$), (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:

$$u(x,a,b) + \int_X w(y) q(dy|x,a,b) \le w(x) \quad \text{for all} \quad (x,a,b) \in K. \tag{5.21}$$

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:

$$\hat{j}(x,\pi,\gamma) = \liminf_{t\to\infty} \frac{E_x^{\pi\gamma}\left(\sum_{n=1}^{N(t)} u(x_n,a_n,b_n) + (T_{N(t)+1}-t)u_2(x_{N(t)},a_{N(t)},b_{N(t)})\right)}{t}.$$

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

$$\tau(x,a,b) \leq d\omega(x), \qquad |u(x,a,b)| \leq d\omega(x),$$

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

$$\int_X \omega(y)q(dy|x,a,b) \leq \hat{\lambda}\omega(x) + \eta 1_C(x),$$

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

$$q(D|x,a,b) \geq \delta\mu(D),$$

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

$$H(\kappa|x, a, b) \le \xi,$$

for all $x \in C$, $a \in A(x)$ and $b \in B(x)$. Moreover, $\tau(x, a, b) \le d$ for all $(x, a, b) \in K$.

(R2) There exists a decreasing function α such that $\alpha(0) \le d$, $\alpha(\infty) = 0$ and

$$\int_t^\infty sH(ds|x, a, b) \le \alpha(t)$$

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

$$\left| \int_X \phi(y) q(dy|x, f, g) - \int_X \phi(y) \pi_{fg}(dy) \right| \le \|\phi\|_\omega \theta \omega(x) \hat{\alpha}^n \tag{5.26}$$

for every $\phi \in \mathcal{B}_\omega(X)$ and $x \in X$, $n \ge 1$. 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 $n \ge 2$ 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*

$$h^*(x) = \operatorname{val}\left[u(x,\cdot,\cdot) - v\tau(x,\cdot,\cdot) + \int_X h^*(y)q(dy|x,\cdot,\cdot)\right] \tag{5.28}$$

$$= \sup_{\nu\in\Pr(A(x))}\inf_{\rho\in\Pr(B(x))}\left[u(x,\nu,\rho) - v\tau(x,\nu,\rho) + \int_X h^*(y)q(dy|x,\nu,\rho)\right]$$

$$= \inf_{\rho\in\Pr(B(x))}\sup_{\nu\in\Pr(A(x))}\left[u(x,\nu,\rho) - v\tau(x,\nu,\rho) + \int_X h^*(y)q(dy|x,\nu,\rho)\right]$$

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 $k \rightarrow \Phi^h(k)$ 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 $(x,a,b) \rightarrow q(D|x,a,b)$ 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 $h : X \rightarrow \mathbb{R}$ 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 $i \neq j$. Further, we have $q(1|1,1,1) = q(1|1,2,2) = 1$ and $q(2|1,i,j) = 1$ for $i \neq j$. 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

$$q_\delta^0(y,x,a,b) \geq 0 \;\text{ for } y \neq x, \quad q_\delta^0(y,y,a,b) \geq -1 \;\text{ and }\; \sum_{y \in X} q_\delta^0(y,x,a,b) = 0.$$

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