Indian Statistical Institute Series

S. K. Neogy
Ravindra B. Bapat
Dipti Dubey *Editors*

Mathematical Programming and Game Theory

STATISTICAL INDIAN INSTITUTE UNITY IN DIVERSITY

Springer

Indian Statistical Institute Series

The *Indian Statistical Institute Series* publishes high-quality content in the domain of mathematical sciences, bio-mathematics, financial mathematics, pure and applied mathematics, operations research, applied statistics and computer science and applications with primary focus on mathematics and statistics. Editorial board comprises of active researchers from major centres of Indian Statistical Institutes. Launched at the 125th birth Anniversary of P.C. Mahalanobis, the series will publish textbooks, monographs, lecture notes and contributed volumes. Literature in this series will appeal to a wide audience of students, researchers, educators, and professionals across mathematics, statistics and computer science disciplines.

More information about this series at http://www.springer.com/series/15910

S. K. Neogy · Ravindra B. Bapat
Dipti Dubey
Editors

Mathematical Programming and Game Theory

Springer

Editors
S. K. Neogy
Indian Statistical Institute
New Delhi, India

Dipti Dubey
Indian Statistical Institute
New Delhi, India

Ravindra B. Bapat
Indian Statistical Institute
New Delhi, India

ISSN 2523-3114 ISSN 2523-3122 (electronic)
Indian Statistical Institute Series
ISBN 978-981-13-3058-2 ISBN 978-981-13-3059-9 (eBook)
https://doi.org/10.1007/978-981-13-3059-9

Library of Congress Control Number: 2018959267

Cover photo: Reprography & Photography Unit, Indian Statistical Institute, Kolkata

This Springer imprint is published by the registered company Springer Nature Singapore Pte Ltd.
The registered company address is: 152 Beach Road, #21-01/04 Gateway East, Singapore 189721, Singapore

Preface

Mathematical programming and game theory models are applied frequently in management, business, and social studies. This volume deals with certain topics of fundamental importance in mathematical programming, game theory, and other related sciences that are presented in the form of 12 chapters. It is a peer-reviewed volume under Indian Statistical Institute Series with a primary focus on recent topics that discuss new challenges from theory and practice. Some pioneers in the field and some prominent young researchers have contributed chapters to this volume. This volume presents an integration of mathematical programming and game theory models that use different methodologies to improve the decision making associated with the new challenges of the present and future problems.

The linear complementarity problem (LCP) is normally identified as a problem of mathematical programming, and it provides a unifying framework for several optimization problems like linear programming, linear fractional programming, convex quadratic programming, and bimatrix game problem. More specifically, LCP models the optimality conditions of these problems. Chapter 1 by D. Dubey and S. K. Neogy starts with the presentation of various mathematical programming problems and bimatrix game problem as the linear complementarity problem. Rest of the chapter is devoted to a study of the properties of some matrix classes in the linear complementarity theory and its usefulness for solving LCP by Lemke's algorithm. Under what conditions a linear complementarity problem can be solved as a linear programming problem is also discussed. Finally, various generalizations that appear in various applications in engineering, management science, and game theory are also discussed.

Chapters 2–4 deal with mathematical programming problems that arise in graph theory. Chapter 2 by R. B. Bapat considers two problems, namely the problems of maximizing the spectral radius and the number of spanning trees in a class of bipartite graphs with certain degree constraints, and the optimal graph for both the problems is conjectured to be a Ferrers graph. Several necessary and sufficient conditions under which the removal of an edge in a graph does not affect the resistance distance between the end-vertices of another edge are presented in this chapter. A brief survey of the problem and references to the literature containing

results and open problems are also given. A new proof of the formula for the number of spanning trees in a Ferrers graph is presented, which is different from the proof of Ehrenborg and van Willigenburg that uses electrical networks and resistances. Chapter 3 by Masahiro Hachimori considers optimization problem on orientations of a given graph, where the values of the objective functions are determined by the out-degrees of the resulting directed graph and the constraints contain acyclicity of the orientations. A survey of the applications of such optimization problems in polytope theory, shellability of simplicial complexes, and acyclic partitions are also discussed. Another interesting problem is to look for a nontrivial class of graphs for which optimization problems that are presented in this chapter can be solved in a polynomial time. Chapter 4 deals with the Max-Flow-Min-Cut property and total dual integrality. A matrix inequality $A\mathbf{x} \geq \mathbf{b}$ (resp. to $A\mathbf{x} \leq \mathbf{b}$) is called *totally dual integral* if the linear program $\min\{\langle \mathbf{w}, \mathbf{x}\rangle | A\mathbf{x} \geq \mathbf{b}\}$ (resp. to $\max\{\langle \mathbf{w}, \mathbf{x}\rangle | A\mathbf{x} \leq \mathbf{b}\}$) has an integral optimal dual solution $\mathbf{y}$ for every integral cost vector $\mathbf{w}$ for which the above linear program has a finite optimum. Motivated by the pluperfect and (weak) perfect graph theorems for the set covering problem by Fulkerson and Lovász, Seymour introduced the concept of the so-called Max-Flow-Min-Cut property (the MFMC property) of clutters, which is the packing counterpart of the totally dual integrality built in the perfection. A clutter $\mathcal{C}$ has the MFMC property if, for its clutter matrix $M(\mathcal{C})$, the linear system $M(\mathcal{C})\mathbf{x} \geqq \mathbf{1}, \mathbf{x} \geqq \mathbf{0}$ is totally dual integral. Conforti and Cornuéjols conjectured that a clutter has the packing property if and only if it has the MFMC property (Conjecture 1). Cornuéjols, Guenin, and Margot conjectured that the blocking number of every ideal minimally nonpacking clutter is 2. Furthermore, they proved that Conjecture 1 implies Conjecture 2. In this chapter, K. Kashiwabara and T. Sakuma provide a framework to attack Conjecture 2.

Chapter 5 deals with an important combinatorial optimization problem, namely travelling salesman problem (TSP). The objective of TSP is to find an optimal tour that visits every node in a finite set of nodes and returns to the origin node on a graph, given the matrix of distances between any two nodes. In this chapter, Tiru Arthanari and Kun Qian study TSP, followed by some preliminaries in graph theory. The authors then compare the Dantzig, Fulkerson, and Johnson (DFJ) formulation, Carr's cycle-shrink relaxation (an LP formulation), and multi-stage insertion (MI) formulation given by Arthanari. Various advantages of the MI formulation are discussed. With the same LP relaxation values as the classic DFJ formulation, the MI formulation has only n^3 variables and n^2 constraints, compared to DFJ with $n(n-1)$ variables and $2^{n-1}+n-1$ constraints. Using CPLEX, a commercial LP solver, the MI formulation has been shown to be competitive compared to other formulations of TSP. An interpretation of the MI formulation as a hypergraph minimum cost flow problem and some theoretical computational complexity results on the algorithms involved in solving the hypergraph minimum cost flow problem, namely the flow and potential algorithm, are also presented.

Chapter 6 by D. Aussel, J. Dutta, and T. Pandit discusses the links between equilibrium problems and variational inequalities. Under the most natural assumption, the equilibrium problem is shown to be equivalent to an associated variational

inequality and the existence results for equilibrium problems can be obtained from the existence results for variational inequality problems and vice versa. The authors also study a problem of existence of Nash equilibrium in an oligopolistic market and show that it is equivalent to a variational inequality under the most natural economic assumption. Further, the relation between the quasi-equilibrium problem and quasi-variational inequality is also studied.

Chapter 7 by Y. Kimura presents approximation techniques as the solution to convex minimization problems by using iterative sequences with resolvent operators and proposes an iterative scheme for an approximation of the solution to a common minimization problem for a finite family of convex functions.

Chapter 8 by Sushmita Gupta, Sanjukta Roy, Saket Saurabh, and Meirav Zehavi deals with an emerging area of research within algorithmic game theory: multivariate analysis of games. This chapter presents a survey of the landscape of work on various stable marriage problems and the use of parametrized complexity as a toolbox to study computationally hard variants of these problems. The entire survey is divided into three broad topics, namely strategic manipulation, maximum(minimum) sized matching in the presence of ties, and notions of fair or equitable stable matchings.

Chapter 9 by M. Kaneko deals with quasi-linear utility functions that are widely used in economics and game theory as convenient tools. The author makes an explicit connection between approximate quasi-linearity and expected utility theory and presents two applications of their results to the theories of cooperative games with side payments and of Lindahl-ratio equilibrium for a public goods economy with quasi-linearity.

Chapter 10 by L. Mallozzi and A. Sacco presents a cooperative game theoretical model for a multi-commodity network flow problem. In this game, each player receives a return for shipping his commodity and considers the possibility to have uncertainty on the costs. A cooperative game under interval uncertainty is presented for the model, and the existence of core solutions is also investigated.

Chapter 11 by Andrey Garnaev and Wade Trappe discusses an interesting topic on pricing competition between cell phone carriers in a growing market of customers. A game theoretical model for the competition between service providers, such as cell phone carriers, in a market of customers that is growing, was investigated. Solving this game helps to show how the loyalty factor associated with the carriers might impact the prices and relative market share between the carriers.

Chapter 12 by Reinoud Joosten and Robin Meijboom presents and analyzes a stochastic game in which transition probabilities between states are not fixed as in standard stochastic games, but depend on the history of the play, i.e., the players' past action choices. For the limiting average reward criterion, the authors determine the set of jointly convergent pure-strategy rewards which can be supported by equilibria involving threats. Further, for expository purposes, a stylized fishery game is analyzed. In each period, two agents choose between catching with restraint and catching without restraint. The resource is in either of two states, *high* or *low*. Restraint is harmless to the fish, but it is a dominated action at each stage. The lesser the restraint shown during the play, the higher the probabilities that the system

moves to or stays in *low*. The latter state may even become 'absorbing temporarily'; i.e., transition probabilities to *high* temporarily become zero, while transition probabilities to *low* remain nonzero. Future research should combine various modifications and extensions of the original Small Fish Wars with the innovation presented here.

It is hoped that the results presented in this research monograph will inspire young researchers for further contributions to the fields of mathematical programming, game theory, and graph theory, especially in the form of novel applications and development of computational techniques.

New Delhi, India
July 2018

S. K. Neogy
Ravindra B. Bapat
Dipti Dubey

Acknowledgements

The editors are thankful to the following referees who have helped in reviewing the chapters of this research monograph.

- Jeffrey Kline, University of Queensland, Australia.
- Satoru Takahashi, National University of Singapore, Singapore.
- Yaokun Wu, Shanghai Jiao Tong University, China.
- H. V. Zhao, University of Alberta, Canada.
- Adam N. Letchford, Lancaster University, UK.
- Fumiaki Kohsaka, Tokai University, Japan.
- Sivaramakrishnan Sivasubramanian, Indian Institute of Technology Bombay, Mumbai, India.
- Woong Kook, Seoul National University, Seoul, Korea.
- Antonino Maugeri, University of Catania, Italy.
- Gerhard-Wilhelm Weber, Middle East Technical University, Turkey.
- Kazuo Iwama, Kyoto University, Japan.
- Kimmo Berg, Aalto University, Finland.
- Mitsunobu Miyake, Tohoku University, Japan.

We are grateful to our authors who contributed chapters to this research monograph. Finally, we thank Springer for their cooperation at all stages in publishing this volume.

S. K. Neogy
Ravindra B. Bapat
Dipti Dubey

Contents

About the Editors

S. K. Neogy is Professor at Indian Statistical Institute, New Delhi. He obtained his Ph.D. from the same institute, and his primary areas of research are mathematical programming and game theory. He is the co-editor of the following books: *Modeling, Computation and Optimization and Mathematical Programming and Game Theory for Decision Making* (both from World Scientific). He has also been a co-editor of the special issue of several journals: *Annals of Operations Research, entitled Optimization Models with Economic and Game Theoretic Applications* (2016), *International Game Theory Review, Entitled Operations Research and Game Theory* (2001), and *Applied Optimization and Game-Theoretic Models*, Parts I and II (2015). He has published widely in several international journals of repute like *Mathematical Programming, Linear Algebra and its Applications,* OR *Spektrum, SIAM Journal on Matrix Analysis and Applications, SIAM Journal on Optimization, International Journal of Game Theory, Dynamic Games and Applications, Annals of Operations Research,* and *Mathematical Analysis and Applications*. He is a reviewer of zbMATH and Mathematical Reviews.

Ravindra B. Bapat obtained his Ph.D. from the University of Illinois at Chicago and is Professor at the Stat-Math Unit, Indian Statistical Institute, New Delhi. He was earlier associated with Northern Illinois University in DeKalb, Illinois, and the University of Mumbai, India, before joining Indian Statistical Institute, New Delhi, in 1983. He held visiting positions at various universities in the USA and visited several institutes in countries including France, Holland, Canada, China, and Taiwan for collaborative research and seminars. His main areas of research are nonnegative matrices, matrix inequalities, matrices in graph theory, and generalized inverses. He has published over 140 research papers in these areas in journals of repute and guided several Ph.D. students. He is the author of several books on linear algebra including *Linear Algebra and Linear Models* and *Graphs and Matrices* (both published by Springer). He also wrote a book on Mathematics for

the General Reader, in Marathi, which won the State Government Award for 2004 for the Best Literature in Science. In 2009, he was awarded the J.C. Bose Fellowship. He has been on the editorial boards of several journals: *Linear and Multilinear Algebra, Electronic Journal of Linear Algebra, Indian Journal of Pure and Applied Mathematics*, and *Kerala Mathematical Association Bulletin*. He has been elected Fellow of the Indian Academy of Sciences, Bangalore, and the Indian National Science Academy, New Delhi. He has served as President of the Indian Mathematical Society during its centennial year 2007–2008. For the past several years, he has been actively involved with the Mathematics Olympiad Program in India as the national coordinator for the program. He has also served as Head, Indian Statistical Institute, New Delhi, during 2007–2011.

Dipti Dubey is Postdoctoral Fellow at Indian Statistical Institute, New Delhi. A Ph.D. from the Indian Institute of Technology Delhi, her primary area of research is mathematical programming and game theory.

She has published widely in several international journals of repute like *Linear Algebra and its Applications, Linear and Multilinear Algebra, Annals of Operations Research, Dynamic Games and Applications, Operations Research Letters, and Fuzzy Sets and Systems*. She is a reviewer of many international journals on optimization and Mathematical Reviews.

Chapter 1
A Unified Framework for a Class of Mathematical Programming Problems

Dipti Dubey and S. K. Neogy

1.1 Introduction

The *linear complementarity problem (LCP)* appears in the literature as one of the fundamental problems in mathematical programming and it is a combination of linear and nonlinear system of inequalities and equations. LCP includes a large class of mathematical programming and game problems and it is always an extremely demanding and interesting topic to researchers on optimization. The novelty of the problem is that it unifies several mathematical programming problems like linear programming, linear fractional programming, convex quadratic programming, and the bimatrix game problem. The problem is studied for more than 50 years in the lliterature and it is stated as follows.

Given a matrix $M \in \mathbb{R}^{n\times n}$ and a vector $q \in \mathbb{R}^n$, find $z \in \mathbb{R}^n$ such that $Mz + q \geq 0,\ z \geq 0$ and $z^T(Mz + q) = 0$ (or prove that such a z does not exist).

Alternatively, the problem may be restated as follows: For a given matrix $M \in \mathbb{R}^{n\times n}$ and a vector $q \in \mathbb{R}^n$, the linear complementarity problem (denoted by LCP(q, M)) is to find vectors $w,\ z \in \mathbb{R}^n$ such that

$$w - Mz = q,\quad w \geq 0,\ z \geq 0 \tag{1.1}$$

$$w^T z = 0. \tag{1.2}$$

This work was supported by SERB Grant.

D. Dubey (✉) · S. K. Neogy
Indian Statistical Institute, 7 S. J. S Sansanwal Marg, New Delhi 110016, India
e-mail: diptidubey@isid.ac.in

S. K. Neogy
e-mail: skn@isid.ac.in

S. K. Neogy et al. (eds.), *Mathematical Programming and Game Theory*,
Indian Statistical Institute Series, https://doi.org/10.1007/978-981-13-3059-9_1

A pair (w, z) of vectors satisfying (1.1) and (1.2) is called a solution to the $LCP(q, M)$. We denote the feasible set by $F(q, M) = \{z : Mz + q \geq 0,\ z \geq 0\}$ and the solution set by $S(q, M) = \{z : z \in F(q, M),\ z^T(Mz + q) = 0\}$. LCP is normally identified as a part of optimization theory and equilibrium problems. Eaves [11] noted that the linear complementarity problem may be thought of a specialized quadratic program (QP) and it is basically the problem of finding an optimal solution (w, z) of the QP

$$\text{minimize } w^T z = z^T M z + z^T q \text{ subject to } \ Iw - Mz = q, w \geq 0, z \geq 0,$$

if the optimal objective value is zero. The algorithm presented by Lemke and Howson [24] to compute an equilibrium pair of strategies to a bimatrix game, later extended by Lemke [22] to solve an LCP(q, M) contributed significantly to the development of the linear complementarity theory and brought the LCP into the limelight. Ever since, the subject has been making great strides and has been a fertile field for practitioners and researchers. It also arises in a number of applications in operations research, control theory, mathematical economics, geometry, and engineering. For further details on this problem and its applications see [8, 13, 39].

1.2 Preliminaries

We consider matrices and vectors with real entries. Any vector $x \in \mathbb{R}^n$ is a column vector unless otherwise specified and x^T denotes the row transpose of x. $I_{\cdot j}$ denotes the vector whose jth coordinate is 1 and whose other coordinates are 0s. If $x = (x_1, \ldots, x_r)^T$ and $y = (y_1, \ldots, y_r)^T$ are two vectors, we write $x < y$ if $x_i < y_i,\ \forall\ 1 \leq i \leq r$ and $x \leq y$ if $x_i \leq y_i,\ \forall\ 1 \leq i \leq r$. For any vector $x \in \mathbb{R}^n$, x^+ and x^- are the vectors whose components are x_i^+ $(= \max\{x_i, 0\})$ and x_i^- $(= \max\{-x_i, 0\})$, respectively, for all i. If $x \in \mathbb{R}^n$ and $y \in \mathbb{R}^n$ are two vectors, the symbol $x \wedge y$ denotes the vector $u \in \mathbb{R}^n$ whose ith coordinate u_i is given by $u_i = \min(x_i, y_i)$. By writing $A \in \mathbb{R}^{m \times n}$, we denote that A is a matrix of real entries with m rows and n columns. For any matrix $A \in \mathbb{R}^{m \times n}$, a_{ij} denotes its ith row and jth column entry. $A_{\cdot j}$ denotes the jth column and $A_{i\cdot}$, the ith row of A. If A is a matrix of order $m \times n$, $\alpha \subseteq \{1, 2, \ldots, m\}$ and $\beta \subseteq \{1, 2, \ldots, n\}$ then $A_{\alpha\beta}$ denotes the submatrix of A consisting of only the rows and columns of A, whose indices are in α and β, respectively. If $\alpha = \beta$ then the submatrix $A_{\alpha\alpha}$ is called the *principal submatrix* of A and $\det(A_{\alpha\alpha})$ is called the *principal minor* of A. For a given integer k $(1 \leq p \leq n)$, the principal submatrix $A_{\alpha\alpha}$ where $\alpha = \{1, \ldots, p\}$ is called a *leading principal submatrix* of A. Given a symmetric matrix $S \in \mathbb{R}^{n \times n}$, its *inertia* is the triple $(\nu_+(S), \nu_-(S), \nu_0(S))$ where $\nu_+(S)$, $\nu_-(S)$, $\nu_0(S)$ denote the number of positive, negative and zero eigenvalues of S respectively. $A_{\alpha\cdot}$ denotes the submatrix formed by the rows of A, whose indices are in α. Similarly, $A_{\cdot\alpha}$ denotes the submatrix formed by the columns of the matrix A, whose indices are in α. For any set β, $|\beta|$ denotes its cardinality. For any set $\alpha \subseteq \{1, 2, \ldots, n\}$, $\bar{\alpha}$ denotes its complement

in $\{1, 2, \ldots, n\}$. Pos(A) denotes the cone generated by columns of A. A probability vector is a vector $x \in \mathbb{R}^n$ such that all the coordinates of x are nonnegative and $\sum_{i=1}^{n} x_i = 1$, where x_i is the ith coordinate of x. Tucker introduced the concept of principal pivot transforms (PPTs). The *principal pivot transform* of M with respect to $\alpha \subseteq \{1, \ldots, n\}$ is defined as the matrix given by

$$M' = \begin{bmatrix} M'_{\alpha\alpha} & M'_{\alpha\bar{\alpha}} \\ M'_{\bar{\alpha}\alpha} & M'_{\bar{\alpha}\bar{\alpha}} \end{bmatrix}$$

where $M'_{\alpha\alpha} = (M_{\alpha\alpha})^{-1}$, $M'_{\alpha\bar{\alpha}} = -(M_{\alpha\alpha})^{-1} M_{\alpha\bar{\alpha}}$, $M'_{\bar{\alpha}\alpha} = M_{\bar{\alpha}\alpha}(M_{\alpha\alpha})^{-1}$, $M'_{\bar{\alpha}\bar{\alpha}} = M_{\bar{\alpha}\bar{\alpha}} - M_{\bar{\alpha}\alpha}(M_{\alpha\alpha})^{-1} M_{\alpha\bar{\alpha}}$. The expression $M_{\bar{\alpha}\bar{\alpha}} - M_{\bar{\alpha}\alpha}(M_{\alpha\alpha})^{-1} M_{\alpha\bar{\alpha}}$ is the *Schur complement* of $M_{\alpha\alpha}$ in M and is denoted as $(M/M_{\alpha\alpha})$. The PPT of LCP (q, M) with respect to α (obtained by pivoting on $M_{\alpha\alpha}$) is given by LCP (q', M'), where $q'_\alpha = -M_{\alpha\alpha}^{-1} q_\alpha$ and $q'_{\bar{\alpha}} = q_{\bar{\alpha}} - M_{\bar{\alpha}\alpha} M_{\alpha\alpha}^{-1} q_\alpha$. We use the notation $\wp_\alpha(M) (= M')$ for PPT of M with respect to $\alpha \subseteq \{1, \ldots, n\}$. Note that PPT is only defined with respect to those α for which $\det M_{\alpha\alpha} \neq 0$. By a *legitimate principal pivot transform*, we mean the PPT obtained from M by performing a principal pivot on a nonsingular principal submatrix. When $\alpha = \emptyset$, by convention $\det M_{\alpha\alpha} = 1$ and $M = \wp_\alpha(M)$. For further details on principal pivot transform, see [3] and references therein.

1.3 A Class of Mathematical Programming Problems in Complementarity Framework

In this section, we consider a class of mathematical programming problems, namely linear programming problem, quadratic programming problem, linear fractional programming problem, etc., which lead to linear complementarity problems.

1.3.1 Linear Programming

Let $A \in \mathbb{R}^{m \times n}$, $b \in \mathbb{R}^m$, and $c \in \mathbb{R}^n$. Consider the primal linear program (P): minimize $c^T x$ subject to $Ax \geq b, x \geq 0$ and its dual (D): maximize $b^T y$ subject to $A^T y \leq c, y \geq 0$.

An important aspect of the primal–dual relationship is the complementary slackness principle which is the following:

If x is feasible to (P) and y is feasible to (D) then x, y are optimal to the respective problem if and only if

$$y^T(Ax - b) + x^T(c - A^T y) = 0.$$

Using the above result, we can associate to the problems, (P) and (D) as a complementarity problem. Indeed, adding slack variables $u \in \mathbb{R}^n$, $v \in \mathbb{R}^m$ such that $u = c - A^T y \geq 0$, $v = -b + Ax \geq 0$ and $u^T x = 0, v^T y = 0$ and denoting $M = \begin{bmatrix} 0 & -A^T \\ A & 0 \end{bmatrix}$, $q = \begin{bmatrix} c \\ -b \end{bmatrix}$, $z = \begin{bmatrix} x \\ y \end{bmatrix}$, and $w = \begin{bmatrix} u \\ v \end{bmatrix}$, we obtain LCP$(q, M)$.

Remark 1 The complementary slackness principle holds not only for the linear programming problem; it also holds for more general programming problem. In particular, this principle is useful for developing algorithms for the convex quadratic programming problems, in which the objective function is convex and quadratic and the constraints are linear. It is also useful for minimizing a linear fractional function, in which the denominator does not vanish for any feasible x, and the constraints are linear.

The complementary slackness principle for the more general programming problems is based on the Karush–Kuhn–Tucker conditions of optimality. A statement of these conditions for a programming problem with linear constraints in nonnegative variables is as follows.

Let $f : \mathbb{R}^n_+ \to \mathbb{R}$ be a convex function. Let $A \in \mathbb{R}^{m \times n}$ be a given matrix and $b \in \mathbb{R}^m$ be a given vector. Consider the problem: minimize $f(x)$ subject to $Ax \leq b$, $x \geq 0$. Let $S = \{x \mid x \geq 0, Ax \leq b\}$. The Karush–Kuhn–Tucker condition of optimality states that $\bar{x}$ is an optimal solution to the above problem if and only if there exist $\bar{u} \in \mathbb{R}^m$, $\bar{v} \in \mathbb{R}^n$ such that

$$\nabla f(\bar{x}) + A^T \bar{u} - \bar{v} = 0,$$

$$A\bar{x} \leq b, \bar{x} \geq 0,$$

$$\bar{u} \geq 0, \bar{v} \geq 0,$$

$$\bar{v}^T \bar{x} = 0,$$

$$\bar{u}^T (b - A\bar{x}) = 0.$$

Note that $\bar{u}^T (b - A\bar{x}) = 0$, $\bar{v}^T \bar{x} = 0$ is the complementary slackness property here.

1.3.2 Quadratic Programming

Quadratic programming problems have a number of applications in Economics. Hence through quadratic and linear programming problem, complementary slackness principle is also highly useful in the economic theory and models and has been recognized as an equilibrium condition. Consider the following quadratic programming problem:

$$\text{minimize } f(x) = c^T x + \frac{1}{2} x^T Q x$$

$$\text{subject to } Ax \geq b, x \geq 0$$

where $Q \in \mathbb{R}^{n\times n}$ is symmetric, $A \in \mathbb{R}^{m\times n}$, $b \in \mathbb{R}^m$, and $c \in \mathbb{R}^n$. Here we have assumed without loss of generality that Q is symmetric. The function $c^T x + \frac{1}{2}x^T Qx$ is convex if and only if Q is positive semidefinite (PSD). In this case, the Karush–Kuhn–Tucker conditions are necessary and sufficient for a given $\bar{x}$ in the set of feasible solutions $S = \{x| - Ax + b \leq 0,\ x \geq 0\}$ to be a solution. The Karush–Kuhn–Tucker necessary and sufficient optimality conditions specialized to this problem yields the following equations and inequalities:

$$c + Q\bar{x} - A^T\bar{y} - \bar{u} = 0,$$

$$-A\bar{x} + \bar{v} = -b,$$

$$\bar{x}^T\bar{u} = \bar{y}^T\bar{v} = 0,$$

$$\bar{x} \geq 0,\ \bar{y} \geq 0,\ \bar{u} \geq 0, \bar{v} \geq 0.$$

This gives us the linear complementarity problem LCP(q, M) with

$$M = \begin{bmatrix} Q & -A^T \\ A & 0 \end{bmatrix},\ q = \begin{bmatrix} c \\ -b \end{bmatrix},\ w = \begin{bmatrix} \bar{u} \\ \bar{v} \end{bmatrix},\ \text{and } z = \begin{bmatrix} \bar{x} \\ \bar{y} \end{bmatrix}.$$

Note that $Q = 0$ give rise to a linear program. Thus when Q is PSD, quadratic programming problem is completely equivalent to solving LCP(q, M).

1.3.3 Linear Fractional Programming Problem

The problem of minimizing a linear fractional function subject to linear inequality constraints also leads to a linear complementarity problem via the Karush–Kuhn–Tucker conditions.

Given an $m \times n$ matrix $A \in \mathbb{R}^{m\times n}$, vectors $b \in \mathbb{R}^m$, $c, d \in \mathbb{R}^n$ and $\alpha, \beta \in \mathbb{R}$, the linear fractional programming problem is the following:

$$\text{minimize } f(x) = \frac{c^T x + \alpha}{d^T x + \beta} \tag{1.3}$$

subject to

$$Ax \leq b, -x \leq 0. \tag{1.4}$$

Let $S = \{x|Ax \leq b, x \geq 0\}$. It is assumed that $d^T x + \beta \neq 0$ for all $x \in S$ or without loss of generality, we assume that $d^T x + \beta > 0$ for all $x \in S$. With this assumption, the function $f(x)$ is both pseudoconvex and pseudoconcave. Hence, the

Karush–Kuhn–Tucker optimality conditions are both necessary and sufficient for a point $\bar{x}$ to be a solution to (1.3)–(1.4). Thus, $\bar{x}$ is a solution to (1.3)–(1.4), if and only if there exist $\bar{y}, \bar{u} \in \mathbb{R}^m$, and $\bar{v} \in \mathbb{R}^n$ such that

$$\nabla f(\bar{x}) + A^T \bar{u} - \bar{v} = 0,$$

$$A\bar{x} + \bar{y} = b,$$

$$\bar{x}^T \bar{v} + \bar{y}^T \bar{u} = 0,$$

$$\bar{x} \geq 0, \bar{u} \geq 0,$$

$$\bar{v} \geq 0, \bar{y} \geq 0.$$

Now for the linear fractional programming problem, we can easily calculate $\nabla f(\bar{x})$. This is given by

$$\nabla f(\bar{x}) = (d^T \bar{x} + \beta)^{-2}[(d^T \bar{x} + \beta)c - (c^T \bar{x} + \alpha)d]$$

which reduces to $(d^T \bar{x} + \beta)^{-2}[D\bar{x} + \beta c - \alpha d]$, where D is an $n \times n$ matrix whose ith row jth column element is given by $d_j c_i - d_i c_j$ for $1 \leq i \leq n$, $1 \leq j \leq n$. We see that $\bar{x}$ is a solution to (1.3)–(1.4) if and only if there exist $\bar{y} \in \mathbb{R}^m$, $\bar{u} \in \mathbb{R}^m$, and $\bar{v} \in \mathbb{R}^n$ such that

$$D\bar{x} + \beta c - \alpha d + A^T \bar{u} - \bar{v} = 0,$$

$$A\bar{x} + \bar{y} = b,$$

$$\bar{x}^T \bar{v} + \bar{y}^T \bar{u} = 0,$$

$$\bar{x} \geq 0, \bar{u} \geq 0,$$

$$\bar{v} \geq 0, \bar{y} \geq 0.$$

The above leads to the following linear complementarity problem:

$$\begin{bmatrix} \bar{v} \\ \bar{y} \end{bmatrix} - \begin{bmatrix} D & A^T \\ -A & 0 \end{bmatrix} \begin{bmatrix} \bar{x} \\ \bar{u} \end{bmatrix} = \begin{bmatrix} \beta c - \alpha d \\ b \end{bmatrix}, \begin{bmatrix} \bar{v} \\ \bar{y} \end{bmatrix} \geq 0, \begin{bmatrix} \bar{x} \\ \bar{u} \end{bmatrix} \geq 0,$$

$$\bar{v}^T \bar{x} = 0, \ \bar{y}^T \bar{u} = 0.$$

We note that the diagonal elements of $M = \begin{bmatrix} D & A^T \\ -A & 0 \end{bmatrix}$ are 0 and $M = -M^T$. Such a matrix is PSD and therefore LCP(q, M) corresponding to a linear fractional programming problem is processable by Lemke's algorithm.

1.3.4 Nash Equilibrium and Bimatrix Games

A bimatrix game is a noncooperative nonzero-sum two-person game (Player I and Player II) in which each player has a finite number of actions (called pure strategies). Let player I have m pure strategies and player II, n pure strategies. In a game if player I chooses strategy i and player II chooses strategy j they incur the costs a_{ij} and b_{ij}, respectively, where $A = [a_{ij}] \in \mathbb{R}^{m\times n}$ and $B = [b_{ij}] \in \mathbb{R}^{m\times n}$ are given cost matrices.

A mixed strategy for player I is a probability vector $x \in \mathbb{R}^m$ whose ith component x_i represents the probability of choosing pure strategy i, where $x_i \geq 0$ for $i = 1, \ldots, m$ and $\sum_{i=1}^{m} x_i = 1$. Similarly, a mixed strategy for player II is a probability vector $y \in \mathbb{R}^n$. If player I adopts a mixed strategy x and player II adopts a mixed strategy y, then their *expected costs* are given by $x^T Ay$ and $x^T By$, respectively.

A pair of mixed strategies (x^*, y^*) with $x^* \in \mathbb{R}^m$ and $y^* \in \mathbb{R}^n$ is said to be a *Nash equilibrium pair* if

$$(x^*)^T Ay^* \leq x^T Ay^* \quad \text{for all mixed strategies } x \in \mathbb{R}^m$$

and

$$(x^*)^T By^* \leq (x^*)^T By \quad \text{for all mixed strategies } y \in \mathbb{R}^n.$$

It is easy to show that the addition of a constant to all entries of A or B leaves the set of equilibrium points invariant. Henceforth, we assume that all entries of the matrices A and B are positive. We consider the following LCP:

$$\begin{bmatrix} u \\ v \end{bmatrix} = \begin{bmatrix} -e_m \\ -e_n \end{bmatrix} + \begin{bmatrix} 0 & A \\ B^T & 0 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix}, \quad \begin{bmatrix} u \\ v \end{bmatrix}^T \begin{bmatrix} x \\ y \end{bmatrix} = 0, \quad \begin{bmatrix} u \\ v \end{bmatrix}, \begin{bmatrix} x \\ y \end{bmatrix} \geq 0 \quad (1.5)$$

where e_m and e_n are m vectors and n vectors whose components are all 1s. It is easy to see that if (x^*, y^*) is a Nash equilibrium pair then $(\bar{x}, \bar{y})$ is a solution to (1.5) where

$$\bar{x} = x^*/(x^*)^T By^* \text{ and } \bar{y} = y^*/(x^*)^T Ay^*. \quad (1.6)$$

Conversely, if $(\bar{x}, \bar{y})$ is a solution of (1.5) then $\bar{x} \neq 0$ and $\bar{y} \neq 0$ in (1.6) is ensured from the positivity of the cost matrices A and B. Therefore, (x^*, y^*) is a Nash equilibrium pair where

$$x^* = \bar{x}/e_m^T \bar{x} \text{ and } y^* = \bar{y}/e_n^T \bar{y}.$$

Lemke and Howson [24] gave an efficient and constructive procedure for obtaining an equilibrium pair by solving LCP(q, M), where $M = \begin{bmatrix} 0 & A \\ B^T & 0 \end{bmatrix}$ and $q = \begin{bmatrix} -e_m \\ -e_n \end{bmatrix}$.

Note that a two-person zero-sum matrix game is a special case of a bimatrix game in which $A + B = 0$. In a two-person zero-sum matrix game, player I chooses an integer i $(i = 1, \ldots, m)$ and player II chooses an integer j $(j = 1, \ldots, n)$ simultaneously. Then player I pays player II an amount a_{ij} (which may be positive, negative, or zero). Since player II's gain is player I's loss, the game is said to be zero-sum. $A = (a_{ij})$ is called the payoff matrix. We write $v(A)$ to denote the value of the game corresponding to the payoff matrix A. In the game described above, player I is the minimizer and player II is the maximizer. The value of the game $v(A)$ is *positive (nonnegative)* if there exists a $0 \neq y \geq 0$ such that $Ay > 0$ $(Ay \geq 0)$. Similarly, $v(A)$ is *negative (nonpositive)* if there exists a $0 \neq x \geq 0$ such that $A^T x < 0$ $(A^T x \leq 0)$.

1.3.5 Computational Complexity of LCP

We consider LCP(q, M) where q is an n-dimensional integer column vector and M is a square matrix with integer entries. We consider the following decision-making problem.

$$Does\ LCP(q, M)\ have\ a\ solution?$$

In order to show that the above problem is NP-complete, we consider a known NP-complete problem which is given below.

Problem FKP: The decision problem of checking feasibility of a $0 - 1$ equality constrained knapsack problem. Let $a_1, a_2, \ldots, a_n, b$ be given $(n + 1)$ positive integer values. Does $a_1x_1 + a_2x_2 + \cdots + a_nx_n = b$ have a (0, 1) solution?

The above problem is a known NP-complete problem. To show NP-completeness of the linear complementarity problem, we construct an equivalent LCP(q, M) corresponding to FKP, where $M = (m_{ij})$ is a matrix of order $(n + 2)$ and q is an $(n + 2)$-dimensional vector defined as follows:

$$q_i = \begin{cases} a_i, & \text{for } 1 \leq i \leq n, \\ -b, & \text{for } i = n + 1, \\ b, & \text{for } i = n + 2. \end{cases}$$

$$m_{ij} = \begin{cases} -1, & \text{for } i = j = 1 \text{ to } n + 2 \\ 1, & \text{for } j = 1 \text{ to } n \text{ with } i = n + 1 \\ -1, & \text{for } j = 1 \text{ to } n \text{ with } i = n + 2 \\ 0, & \text{otherwise.} \end{cases}$$

Corresponding to above FKP we get an LCP(q, M) where

$$M = \begin{bmatrix} -1 & 0 \cdots & 0 & 0 & 0 \\ 0 & -1 \cdots & 0 & 0 & 0 \\ \vdots & \vdots \ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 \cdots & -1 & 0 & 0 \\ 1 & 1 \cdots & 1 & -1 & 0 \\ -1 & -1 \cdots & -1 & 0 & -1 \end{bmatrix} \text{ and } q = \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \\ -b \\ b \end{bmatrix}.$$

We present the following lemma and its proof due to Chung [1] for the sake of completeness.

Lemma 1.3.1 ([1]) *Problem FKP has a solution if and only if the corresponding LCP has a solution.*

Proof Let x be a solution of FKP. Define $w_{n+1} = w_{n+2} = z_{n+1} = z_{n+2} = 0$. For $i = 1, \ldots, n$, define

$$w_i = a_i(1 - x_i), \ z_i = a_i x_i.$$

Thus, $w_i \geq 0, z_i \geq 0, w_i z_i = 0, \ i = 1, \ldots, n+2$.

Also it is easy to see that $w_i + z_i = a_i, \ i = 1, \ldots, n$.

$$w_{n+1} - z_1 - \cdots - z_n + z_{n+1} = -b. \tag{1.7}$$

$$w_{n+2} + z_1 + \cdots + z_n + z_{n+2} = b. \tag{1.8}$$

Hence (w, z) is a solution of the LCP(q, M).

On the other hand, let (w, z) be a solution to LCP(q, M).

$$\text{Define } x_i = \frac{z_i}{a_i}, \ i = 1, 2, \ldots, n.$$

Note that $w_i z_i = 0, \ w_i + z_i = a_i, \ i = 1, 2, \ldots, n$. Therefore z_i is either 0 or a_i. This implies $x_i = 0$ or 1. From (1.7) and (1.8), we get $w_{n+1} + w_{n+2} + z_{n+1} + z_{n+2} = 0$. Since $w \geq 0, z \geq 0$, we have $w_{n+1} = w_{n+2} = z_{n+1} = z_{n+2} = 0$. Thus $z_1 + z_2 + \cdots + z_n = b$. But this implies $a_1 x_1 + a_2 x_2 + \cdots + a_n x_n = b$. Hence, x is a solution of the problem FKP. ■

Remark 2 It is shown above that a known NP-complete problem FKP reduces to LCP(q, M). A nondeterministic algorithm can guess a complementarity basic vector and then check its feasibility in polynomial time. Therefore, the problem LCP(q, M) belongs to NP-complete class. Clearly, all the generalizations of LCP(q, M) presented in Sect. 1.8 also belongs to NP-complete class.

1.4 Matrix Classes in LCP

Matrix classes play an important role for studying the theory and algorithms of LCP. Over the years, a variety of classes of matrices are introduced in LCP literature. Most of the matrix classes encountered in the context of LCP are commonly found in several applications. Several of these matrix classes are of interest, because they characterize certain properties of the LCP and they offer certain nice features from the viewpoint of algorithms. It is useful to review some of these matrix classes and their properties which will form the basis for further discussions.

We say that $M \in \mathbb{R}^{n\times n}$ is

- *positive semidefinite* (PSD) if $x^T Mx \geq 0 \; \forall \; x \in \mathbb{R}^n$.
- *positive definite* (PD) if $x^T Mx > 0 \; \forall \; 0 \neq x \in \mathbb{R}^n$.
- $\mathbf{Z}$ if $m_{ij} \leq 0, \; \forall \, i \neq j$.
- $\mathbf{P}$ ($\mathbf{P}_0$) if all its principal minors are positive (nonnegative).
- $\mathbf{K}$ ($\mathbf{K}_0$)-*matrix* if it is in $\mathbf{Z} \cap \mathbf{P}$ ($\mathbf{Z} \cap \mathbf{P}_0$).
- $\mathbf{N}$ ($\mathbf{N}_0$) if all the principal minors of M are negative (nonpositive).
- $\mathbf{N}$-*matrix of the first category* if it has at least one positive entry.
- *almost* $\mathbf{N}$-*matrix* if the determinant is positive and all proper principal minors are negative.
- $\mathbf{N}$-*matrix of the second category* if $M < 0$.
- *column adequate* if $M \in \mathbf{P}_0$ and for each $\alpha \subseteq \{1, \ldots, n\}$, $\det(M_{\alpha\alpha}) = 0$ implies that columns of $M_{\cdot\alpha}$ are linearly dependent.
- *column sufficient* if for all $x \in \mathbb{R}^n$ the following implication holds:

$$x_i(Mx)_i \leq 0 \; \forall \; i \;\text{ implies }\; x_i(Mx)_i = 0 \;\; \forall \; i.$$

- *row sufficient* if M^T is column sufficient.
- *sufficient* if M and M^T are both column sufficient.
- *copositive* ($\mathbf{C}_0$) (*strictly copositive* ($\mathbf{C}$)) if $x^T Mx \geq 0 \; \forall \; x \geq 0$ ($x^T Mx > 0 \; \forall \; 0 \neq x \geq 0$).
- *copositive-plus* ($\mathbf{C}_0^+$) if $M \in \mathbf{C}_0$ and the implication $[x^T Mx = 0, x \geq 0] \Rightarrow (M + M^T)x = 0$ holds.
- *copositive-star* ($\mathbf{C}_0^*$) if $M \in \mathbf{C}_0$ and the implication $[x^T Mx = 0, \; Mx \geq 0, x \geq 0] \Rightarrow M^T x \leq 0$ holds.
- $\mathbf{Q}$-matrix if LCP(q, M) has a solution $\forall \; q \in \mathbb{R}^n$.
- $\mathbf{Q}_0$-matrix if for all $q \in \mathbb{R}^n$, $F(q, M) \neq \emptyset \Rightarrow S(q, M) \neq \emptyset$.
- $\mathbf{S}$-matrix if there exists a vector $0 \neq x \in \mathbb{R}^n_+$ such that $Mx > 0$.
- $\mathbf{R}_0$-*matrix* if LCP$(0, A)$ has only the trivial solution.
- $\mathbf{L}_1$ (semimonotone) if for every $0 \neq y \geq 0, \; y \in \mathbb{R}^n \; \exists$ an i such that $y_i > 0$ and $(My)_i \geq 0$.
- $\mathbf{L}_2$-*matrix* if for each $0 \neq \xi \geq 0, \; \xi \in \mathbb{R}^n$ satisfying $\eta = M\xi \geq 0$ and $\eta^T \xi = 0$ $\exists$ a $0 \neq \hat{\xi} \geq 0$ satisfying $\hat{\eta} = -M^T \hat{\xi}, \;\; \eta \geq \hat{\eta} \geq 0, \; \xi \geq \hat{\xi} \geq 0$.
- $\mathbf{L}$-*matrix* if it is in both $\mathbf{L}_1$ and $\mathbf{L}_2$.

- $\mathbf{E}(d)$: $\mathbf{E}(d)$, $(d \in \mathbb{R}^n)$ if $(\bar{w}, \bar{z})$, $\bar{z} \neq 0$ is a solution for the LCP(d, M) implies that there $\exists$ a $0 \neq x \geq 0$ such that $y = -M^T x \geq 0$, $x \leq \bar{z}$, and $y \leq \bar{w}$.
- $\mathbf{E}^*(d)$: $\mathbf{E}^*(d)$ for a $d \in \mathbb{R}^n$ if $(\bar{w}, \bar{z})$ is a solution to the LCP(d, M) implies that $\bar{w} = d$, $\bar{z} = 0$.
  Note that $\mathbf{E}(d) = \mathbf{E}^*(d)$ for any $d > 0$ or $d < 0$, $\mathbf{E}(0) = \mathbf{L}_2$ of [11] and $\mathbf{L}(d) = \mathbf{E}(d) \cap \mathbf{E}(0)$. Further $\mathbf{L}_1 = \cap_{d>0}\mathbf{E}(d)$. We refer to $\mathbf{L}(d)$ as *Garcia's class* which extends Eaves class $\mathbf{L}$ and $\mathbf{L}(d) \subseteq \mathbf{Q}_0$.

1.5 Lemke's Algorithm

A widely applicable method for solving LCP(q, M) is the method of Lemke, which is a modification of the Lemke–Howson method proposed in [24] for finding an equilibrium point of a bimatrix game. Lemke [22] proposed this algorithm for solving certain classes of linear complementarity problems which is described below.

For solving (1.1) and (1.2), the following algorithm based on pivot steps has been given by Lemke [22]. The initial solution to (1.1) and (1.2) is taken as

$$w = q + d\, z_0$$

$$z = 0$$

where $d \in \mathbb{R}^n$ is any given positive vector which is called *covering vector* and z_0 is an artificial variable which takes a large enough value so that $w > 0$. This is called *primary ray.*

Step 1: Decrease z_0 so that one of the variables w_i, $1 \leq i \leq n$, say w_r is reduced to zero. We now have a basic feasible solution with z_0 in place of w_r and with exactly one pair of complementary variables (w_r, z_r) being nonbasic.

Step 2: At each iteration, the complement of the variable which has been removed in the previous iteration is to be increased. In the second iteration, for instance, z_r will be increased.

Step 3: If the variable selected at step 2 to enter the basis can be arbitrarily increased, then the procedure terminates in a *secondary ray.* If a new basic feasible solution is obtained with $z_0 = 0$, we have solved (1.1) and (1.2). If in the new basic feasible solution $z_0 > 0$, we have obtained a new basic pair of complementary variables (w_s, z_s). We repeat step 2.

Lemke's algorithm consists of the repeated applications of steps 2 and 3. If nondegeneracy is assumed, the procedure terminates either in a secondary ray or in a solution to (1.1) and (1.2). If degenerate almost complementary solutions are generated these can be resolved using the methods discussed by Eaves [11]. We say that an algorithm processes a problem if the algorithm can either compute a solution to it if one exists, or show that no solution exists. For more explanations see [8]. Many

classes of matrices have been identified in the literature on linear complementarity theory for which one can conclude that there is no solution to (1.1) and (1.2), when Lemke's algorithm with the positive vector d terminates in a secondary ray for some q. Lemke's method is applicable for a fairly large class of matrices. For $M \in \mathbf{L}(d)$ where $d > 0$ the success of Lemke's algorithm applied to LCP(q, M) with d as the covering vector is guaranteed if it is feasible [15].

1.6 Some Recent Matrix Classes and Lemke's Algorithm

In what follows we discuss some recently introduced matrix classes and their processability by Lemke's algorithm.

1.6.1 Positive Subdefinite Matrices

Martos [29] introduced the class of symmetric positive subdefinite matrices (a generalization of the class of positive semidefinite (PSD) matrices) in connection with a characterization of a pseudoconvex function. The study of pseudoconvex and quasiconvex quadratic forms leads to this new class of matrices, and it is useful in the study of quadratic programming problem. Cottle and Ferland [5] further obtained converses for some of Martos's results. Since Martos was considering the Hessians of quadratic functions, he was concerned only about symmetric matrices. Crouzeix et al. [2] studied nonsymmetric version of PSBD matrices in the context of generalized monotonicity and the linear complementarity problem. We say that $M \in \mathbb{R}^{n\times n}$ is positive subdefinite (PSBD) if for all $x \in \mathbb{R}^n$

$$x^T Mx < 0 \text{ implies either } M^T x \leq 0 \text{ or } M^T x \geq 0.$$

M is said to be *merely positive subdefinite* (MPSBD) if M is a PSBD matrix but not positive semidefinite (PSD). The concept of PSBD matrices leads to a study of pseudomonotone matrices. Crouzeix et al. [2] have obtained new characterizations for generalized monotone affine maps on $\mathbb{R}^n_+$ using PSBD matrices. Given a matrix $M \in \mathbb{R}^{n\times n}$ and a vector $q \in \mathbb{R}^n$, an affine map $\mathcal{F}(x) = Mx + q$ is said to be *pseudomonotone* on $\mathbb{R}^n_+$ if

$$(y - z)^T (Mz + q) \geq 0,\ y \geq 0, z \geq 0 \Rightarrow (y - z)^T (My + q) \geq 0.$$

$M \in \mathbb{R}^{n\times n}$ is said to be pseudomonotone if $\mathcal{F}(x) = Mx$ is pseudomonotone on the nonnegative orthant. Gowda [16] establishes a connection between affine pseudomonotone mapping and the linear complementarity problem and showed that for an affine pseudomonotone mapping, the feasibility of the LCP implies its solvability.

Crouzeix et al. [2] proved that an affine map $\mathcal{F}(x) = Mx + q$ where $M \in \mathbb{R}^{n\times n}$ and $q \in \mathbb{R}^n$ is pseudomonotone if and only if

$$z \in \mathbb{R}^n,\ \ z^T Mz < 0 \Rightarrow \begin{cases} M^T z \geq 0 \text{ and } z^T q \geq 0 \text{ or} \\ M^T z \leq 0,\ z^T q \leq 0 \text{ and } z^T(Mz^- + q) < 0. \end{cases}$$

Theorem 1.6.1 ([2, Proposition 2.1]) *Let $M = ab^T$ where $a \neq b,\ a, b \in \mathbb{R}^n$. M is PSBD if and only if one of the following conditions holds:*

(i) $\exists\ a\ t > 0$ *such that* $b = ta$;
(ii) for all $t > 0$, $b \neq ta$ *and either* $b \geq 0$ *or* $b \leq 0$.

Further suppose that $M \in MPSBD$. Then $M \in \mathbf{C}_0$ if and only if either ($a \geq 0$ and $b \geq 0$) or ($a \leq 0$ and $b \leq 0$) and $M \in \mathbf{C}_0^*$ if and only if M is copositive and $a_i = 0$ whenever $b_i = 0$.

The following results are obtained by Crouzeix et al. [2].

Theorem 1.6.2 ([2, Theorem 2.1, Proposition 2.5]) *Let $M \in \mathbb{R}^{n\times n}$ is PSBD and $rank(M) \geq 2$. Then M^T is PSBD and at least one of the following conditions holds:*

(i) M *is PSD;*
(ii) $(M + M^T) \leq 0$;
(iii) M *is* $\mathbf{C}_0^*$.

Theorem 1.6.3 ([2, Proposition 2.2]) *Suppose $M \in \mathbb{R}^{n\times n}$ is MPSBD and $rank(M) \geq 2$. Then*

(a) $\nu_-(M + M^T) = 1$,
(b) $(M + M^T)z = 0 \ \Leftrightarrow\ Mz = M^T z = 0$.

Theorem 1.6.4 ([2, Theorem 3.3]) *A matrix $M \in \mathbb{R}^{n\times n}$ is pseudomonotone if and only if M is PSBD and copositive with the additional condition in case $M = ab^T$, that $b_i = 0 \Rightarrow a_i = 0$.*

In fact, the class of pseudomonotone matrices coincides with the class of matrices which are both PSBD and copositive-star.

Theorem 1.6.5 ([16, Corollary 4]) *If M is pseudomonotone, then M is a row sufficient matrix.*

PSBD matrix is introduced as a natural generalization of a PSD matrix. However, many properties of a PSD matrix may not hold for a PSBD matrix. Let $M = \begin{bmatrix} 0 & 2 \\ -1 & 0 \end{bmatrix}$. It is easy to check that $M \in$ PSBD but $(M + M^T)$ and M^{-1} is not a PSBD matrix. The next theorem says that PSBD is a complete class in the sense of [8, 3.9.5].

Theorem 1.6.6 ([37]) *Suppose $M \in \mathbb{R}^{n\times n}$ is a PSBD matrix. Then $M_{\alpha\alpha} \in PSBD$ where $\alpha \subseteq \{1, \ldots, n\}$.*

Proof Let $M \in$ PSBD and $\alpha \subseteq \{1, \ldots, n\}$. Let $x_\alpha \in \mathbb{R}^{|\alpha|}$ and

$$M = \begin{bmatrix} M_{\alpha\alpha} & M_{\alpha\bar{\alpha}} \\ M_{\bar{\alpha}\alpha} & M_{\bar{\alpha}\bar{\alpha}} \end{bmatrix}.$$

Suppose that $x_\alpha^T M_{\alpha\alpha} x_\alpha < 0$. Now define $z \in \mathbb{R}^n$ by taking $z_\alpha = x_\alpha$ and $z_{\bar{\alpha}} = 0$. Then $z^T M z = x_\alpha^T M_{\alpha\alpha} x_\alpha$. Since M is a PSBD matrix, $z^T M z = x_\alpha^T M_{\alpha\alpha} x_\alpha < 0 \Rightarrow$ either $M^T z \geq 0$ which implies that $M_{\alpha\alpha}^T x_\alpha \geq 0$ or $M^T z \leq 0$ (which implies $M_{\alpha\alpha}^T x_\alpha \leq 0$). Therefore $M_{\alpha\alpha} \in$ PSBD. As α is arbitrary, it follows that every principal submatrix of M is a PSBD matrix. ■

Theorem 1.6.7 ([37]) *Assume that $M \in \mathbb{R}^{n\times n}$ is a PSBD matrix. Let $D \in \mathbb{R}^{n\times n}$ be a positive diagonal matrix. Then $M \in PSBD$ if and only if $DMD^T \in PSBD$.*

Proof Let $M \in$ PSBD. For any $x \in \mathbb{R}^n$, let $y = D^T x$. Note that $x^T DMD^T x = y^T My < 0 \Rightarrow M^T y = M^T D^T x \leq 0$ or $M^T y = M^T D^T x \geq 0$. This implies that either $DM^T D^T x \leq 0$ or $DM^T D^T x \geq 0$ since D is a positive diagonal matrix. Thus $DMD^T \in$ PSBD. The converse follows from the fact that D^{-1} is a positive diagonal matrix and $M = D^{-1}(DMD^T)(D^{-1})^T$. ■

Theorem 1.6.8 ([37]) *If $M \in \mathbb{R}^{n\times n}$ is a PSBD matrix and $P \in \mathbb{R}^{n\times n}$ is any permutation matrix, then $PMP^T \in PSBD$, i.e., PSBD matrices are invariant under principal rearrangement.*

Proof Let $M \in$ PSBD and let $P \in \mathbb{R}^{n\times n}$ be any permutation matrix. For any $x \in \mathbb{R}^n$, let $y = P^T x$. Note that $x^T PMP^T x = y^T My < 0 \Rightarrow M^T y = M^T P^T x \leq 0$ or $M^T y = M^T P^T x \geq 0$. This implies that either $PM^T P^T x \leq 0$ or $PM^T P^T x \geq 0$ since P is just a permutation matrix. It follows that PMP^T is a PSBD matrix. The converse of the above theorem follows from the fact that $P^T P = I$ and $M = P^T(PMP^T)(P^T)^T$. ■

Lemma 1.6.2 ([37]) *Suppose $M \in \mathbb{R}^{n\times n}$ is a PSBD matrix with $rank(M) \geq 2$ and $M + M^T \leq 0$. If M is not a skew-symmetric matrix, then $M \leq 0$.*

Theorem 1.6.9 ([37]) *Suppose $M \in \mathbb{R}^{n\times n}$ is a PSBD matrix with $rank(M) \geq 2$. Then M is a $\mathbf{Q}_0$-matrix.*

Proof By Theorem 1.6.2, M^T is a PSBD matrix. Also by the same theorem, either $M \in$ PSD or $(M + M^T) \leq 0$ or $M \in \mathbf{C}_0^*$. If $M \in \mathbf{C}_0^*$ then $M \in \mathbf{Q}_0$ (see [8]). Now if $(M + M^T) \leq 0$, and M is not skew-symmetric then by Lemma 1.6.2 it follows that $M \leq 0$. In this case, $M \in \mathbf{Q}_0$ [8]. However, if M is skew-symmetric then $M \in PSD$. Therefore, $M \in \mathbf{Q}_0$. ■

Corollary 1 ([37]) *Assume that M is a PSBD matrix with rank$(M) \geq 2$. Then LCP(q, M) is processable by Lemke's algorithm. If rank$(M) = 1$, (i.e., $M = ab^T$, $a, b \neq 0$) and $M \in \mathbf{C}_0$ then LCP(q, M) is processable by Lemke's algorithm whenever $b_i = 0 \Rightarrow a_i = 0$.*

Proof Suppose rank$(M) \geq 2$. From Theorem 1.6.2 and the proof of Theorem 1.6.9, it follows that M is either a PSD matrix or $M \leq 0$ or $M \in \mathbf{C}_0^*$. Hence LCP(q, M) is processable by Lemke's algorithm (see [8]). For PSBD $\cap$ $\mathbf{C}_0$ matrices of rank$(M) = 1$, i.e., for $M = ab^T$, $a, b \neq 0$, such that $b_i = 0 \Rightarrow a_i = 0$. Note that $M \in \mathbf{C}_0^*$ by Theorem 1.6.1. Hence LCP(q, M) with such matrices are processable by Lemke's algorithm. ■

Theorem 1.6.10 ([37]) *Suppose M is a PSBD $\cap$ $\mathbf{C}_0$ matrix with rank$(M) \geq 2$. Then $M \in \mathbb{R}^{n \times n}$ is a sufficient matrix.*

Proof Note that by Theorem 1.6.2, M^T is a PSBD $\cap$ $\mathbf{C}_0$ matrix with rank$(M^T) \geq 2$. Now by Theorem 1.6.4, M and M^T are pseudomonotone. Hence, M and M^T are row sufficient by Theorem 1.6.5 Therefore M is sufficient. ■

Note that, in general a PSBD matrix need not be a $\mathbf{P}_0$ matrix. It is easy to check that $M = \begin{bmatrix} 0 & -2 \\ -1 & 0 \end{bmatrix}$ is a PSBD matrix but $M \notin \mathbf{P}_0$.

Theorem 1.6.11 ([37]) *Suppose $A \in \mathbb{R}^{n \times n}$ can be written as $M + N$ where $M \in$ MPSBD $\cap$ $\mathbf{C}_0^+$, rank$(M) \geq 2$ and $N \in \mathbf{C}_0$. If the system $q + Mx - N^T y \geq 0$, $y \geq 0$ is feasible, then Lemke's algorithm for LCP(q, A) with covering vector $d > 0$ terminates with a solution.*

Proof Let the feasibility condition of the theorem holds so that there exist an $x^0 \in \mathbb{R}^n$ and a $y^0 \in \mathbb{R}^n_+$ such that $q + Mx^0 - N^T y^0 \geq 0$. First we need to show that for any $x \in \mathbb{R}^n_+$, if $Ax \geq 0$ and $x^T Ax = 0$, then $x^T q \geq 0$. Note that for given $x \geq 0$, $x^T Ax = 0 \Rightarrow x^T (M + N)x = 0$ and since $M, N \in C_0$, this implies that $x^T Mx = 0$. As M is a MPSBD matrix $x^T Mx = 0 \Leftrightarrow x^T (M + M^T)x = 0 \Leftrightarrow (M + M^T)x = 0 \Leftrightarrow M^T x = 0 \Leftrightarrow Mx = 0$. See Theorem 1.6.3. Also since $Ax \geq 0$, it follows that $Nx \geq 0$ and hence $x^T N^T y^0 \geq 0$. Further since $q + Mx^0 - N^T y^0 \geq 0$ and $x \geq 0$, it follows that $x^T (q + Mx^0 - N^T y^0) \geq 0$. This implies that $x^T q \geq x^T N^T y^0 \geq 0$.

Now from Corollary 4.4.12 and Theorem 4.4.13 of [8, p. 277] it follows that Lemke's algorithm for LCP(q, A) with covering vector $d > 0$ terminates with a solution. ■

The class MPSBD $\cap$ $\mathbf{C}_0^+$ is nonempty. It is easy to check this from the matrix $M = \begin{bmatrix} 2 & 5 & 0 \\ 1 & 4 & 0 \\ 0 & 0 & 0 \end{bmatrix}$. Note that $x^T Mx = 2(x_1 + x_2)(x_1 + 2x_2)$. Using this expression, it is easy to verify that $x^T Mx < 0 \Rightarrow$ either $M^T x \leq 0$ or $M^T x \geq 0$. Also it is easy to see that $M \in \mathbf{C}_0^+$.

1.6.2 $\bar{N}$ (Almost $\bar{N}$-Matrix)

The class of $\bar{\mathbf{N}}$-matrices was introduced by Mohan and Sridhar in [31]. The class of almost **N**-matrices is studied in [32, 45]. We discuss here a new matrix class almost $\bar{\mathbf{N}}$ studied in [43], which is a subclass of the almost $\mathbf{N}_0$-matrices.

Definition 1 A matrix $M \in \mathbb{R}^{n\times n}$ is said to be an $\bar{\mathbf{N}}$(*almost* $\bar{\mathbf{N}}$)-matrix if there exists a sequence $\{M^{(k)}\}$, where $M^{(k)} = [m_{ij}^{(k)}]$ are **N**(almost **N**)-matrix such that $m_{ij}^{(k)} \to m_{ij}$ for all $i, j \in \{1, 2, \ldots, n\}$.

Example 1 Let $M = \begin{bmatrix} -1 & 2 & 2 \\ 0 & 0 & 2 \\ 1 & 1 & -1 \end{bmatrix}$. Note that M is an almost $\mathbf{N}_0$-matrix. It is easy to verify that $M \in$ almost $\bar{\mathbf{N}}$ since we can get M as a limit point of the sequence of almost **N**-matrices

$M^{(k)} = \begin{bmatrix} -1 & 2 & 2 \\ \frac{1}{k} & -\frac{1}{k} & 2 \\ 1 & 1 & -1 \end{bmatrix}$.

It is well known that for $\mathbf{P}_0$ (almost $\mathbf{P}_0$)-matrices, by perturbing the diagonal entries alone one can get a sequence of **P** (almost **P**)-matrices that converges to an element of $\mathbf{P}_0$ (almost $\mathbf{P}_0$). However, this is not true for $\mathbf{N}_0$ (almost $\mathbf{N}_0$)-matrices. One of the reasons is that an **N** (almost **N**)-matrix needs to have all its entries nonzero. However, in the above example, we can see that even though the matrix $M \in$ almost $\mathbf{N}_0$, it cannot be obtained as a limit point of almost **N**-matrices by perturbing the diagonal. However, we show in the above example that $M \in$ almost $\bar{\mathbf{N}}$.

The following example shows that an almost $\mathbf{N}_0$-matrix need not be an almost $\bar{\mathbf{N}}$-matrix.

Example 2 Let $M = \begin{bmatrix} 0 & -1 & -1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 \\ 1 & 1 & 0 & -1 \end{bmatrix}$. Here M is an almost $\mathbf{N}_0$-matrix. However, it is easy to see that M is not an almost $\bar{\mathbf{N}}$-matrix since we cannot get M as a limit point of a sequence of almost **N**-matrices.

Suppose $M \in$ almost $\mathbf{N}_0$. Then is it true that (i) $M \in \mathbf{Q}$ implies $M \in \mathbf{R}_0$? The following example demonstrates that $M \in$ almost $\mathbf{N}_0 \cap \mathbf{Q}$ but $M \notin \mathbf{R}_0$.

Example 3 Consider the matrix $M = \begin{bmatrix} -1 & 1 & 1 & 1 \\ 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & -1 \\ 1 & 0 & -1 & 0 \end{bmatrix}$. It is easy to verify that $M \in$ almost $\mathbf{N}_0$. Now taking a PPT with respect to $\alpha = \{1, 3\}$ we get $\wp_\alpha(M) = \begin{bmatrix} 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & 1 \\ 1 & -1 & 1 & 0 \\ -1 & 1 & 0 & 1 \end{bmatrix}$. Now $M \in \mathbf{Q}$ since $\wp_\alpha(M)$ (a PPT of M) $\in \mathbf{Q}$ (*see* [42, p. 193]). However $(0, 1, 0, 0)$ solves LCP$(0, M)$. Hence $M \notin \mathbf{R}_0$.

The next example [45, p. 120] shows that an almost $\mathbf{N}_0$-matrix, even with value positive, need not be a $\mathbf{Q}$-matrix or an $\mathbf{R}_0$-matrix.

Example 4 Let $M = \begin{bmatrix} -2 & -2 & -2 & 2 \\ -2 & -1 & -3 & 3 \\ -2 & -3 & -1 & 3 \\ 2 & 3 & 3 & 0 \end{bmatrix}$ $q = \begin{bmatrix} -1001 \\ -500 \\ -500 \\ 500 \end{bmatrix}$. It is easy to verify that $M \in$ almost $\mathbf{N}_0$ but $M \notin \mathbf{Q}$ even though $v(M)$ is positive. Furthermore, $M \notin \mathbf{R}_0$.

However, if $M \in$ almost $\bar{\mathbf{N}} \cap \mathbf{R}_0$ and $v(M) > 0$, then we show that $M \in \mathbf{Q}$.

In the statement of some theorems that follow, we assume that $n \geq 4$, to make use of the sign pattern stated in the following lemma.

Lemma 1.6.3 ([43]) *Suppose $M \in \mathbb{R}^{n \times n}$ is an almost $\bar{\mathbf{N}}$-matrix of order $n \geq 4$. Then there exists a nonempty subset α of $\{1, 2, \ldots, n\}$ such that M can be written in the partitioned form as (if necessary, after a principal rearrangement of its rows and columns)*

$$M = \begin{bmatrix} M_{\alpha\alpha} & M_{\alpha\bar{\alpha}} \\ M_{\bar{\alpha}\alpha} & M_{\bar{\alpha}\bar{\alpha}} \end{bmatrix}$$

where $M_{\alpha\alpha} \leq 0$, $M_{\bar{\alpha}\bar{\alpha}} \leq 0$, $M_{\bar{\alpha}\alpha} \geq 0$, and $M_{\alpha\bar{\alpha}} \geq 0$.

Proof This follows from Remark 3.1 in [32, p. 623] and from the definition of almost $\bar{\mathbf{N}}$-matrices. ■

In the proof of the sign pattern in Lemma 1.6.3, we assume $n \geq 4$ since lemma requires that all the principal minors of order 3 or less are negative.

Theorem 1.6.12 ([43]) *Suppose $M \in \mathbf{E}_0 \cap$ almost $\bar{\mathbf{N}}$ ($n \geq 4$). Then there exists a principal rearrangement*

$$B = \begin{bmatrix} B_{\alpha\alpha} & B_{\alpha\bar{\alpha}} \\ B_{\bar{\alpha}\alpha} & B_{\bar{\alpha}\bar{\alpha}} \end{bmatrix}$$

of M where $B_{\alpha\alpha}$, $B_{\bar{\alpha}\bar{\alpha}}$ are nonpositive strict upper triangular matrices and $B_{\bar{\alpha}\alpha}$, $B_{\alpha\bar{\alpha}}$ are nonnegative matrices.

Proof Note that M is an almost $\bar{\mathbf{N}}$-matrix of order $n \geq 4$. By Lemma 1.6.3 there exists a nonempty subset α of $\{1, 2, \ldots, n\}$ satisfying

$$M = \begin{bmatrix} M_{\alpha\alpha} & M_{\alpha\bar{\alpha}} \\ M_{\bar{\alpha}\alpha} & M_{\bar{\alpha}\bar{\alpha}} \end{bmatrix}$$

where $M_{\alpha\alpha} \leq 0$, $M_{\bar{\alpha}\bar{\alpha}} \leq 0$, $M_{\bar{\alpha}\alpha} \geq 0$ and $M_{\alpha\bar{\alpha}} \geq 0$.

$M \in \mathbf{E}_0$ implies $M_{\alpha\alpha} \in \mathbf{E}_0$. It is easy to see that there exist permutation matrices $\mathcal{L} \in \mathbb{R}^{|\alpha| \times |\alpha|}$ and $\mathcal{M} \in \mathbb{R}^{|\bar{\alpha}| \times |\bar{\alpha}|}$ such that $\mathcal{L} M_{\alpha\alpha} \mathcal{L}^T$ and $\mathcal{M} M_{\bar{\alpha}\bar{\alpha}} \mathcal{M}^T$ are strict upper triangular matrices. Let

$$P = \begin{bmatrix} \mathcal{L} & 0 \\ 0 & \mathcal{M} \end{bmatrix}$$

be a permutation matrix. Then

$$B = PMP^T = \begin{bmatrix} \mathcal{L}M_{\alpha\alpha}\mathcal{L}^T & \mathcal{L}M_{\alpha\bar{\alpha}}\mathcal{M}^T \\ \mathcal{M}M_{\bar{\alpha}\alpha}\mathcal{L}^T & \mathcal{M}M_{\bar{\alpha}\bar{\alpha}}\mathcal{M}^T \end{bmatrix}$$

where $\mathcal{L}M_{\alpha\alpha}\mathcal{L}^T$ and $\mathcal{M}M_{\bar{\alpha}\bar{\alpha}}\mathcal{M}^T$ are nonpositive strict upper triangular matrices and $\mathcal{L}M_{\alpha\bar{\alpha}}\mathcal{M}^T$, $\mathcal{M}M_{\bar{\alpha}\alpha}\mathcal{L}^T$ are nonnegative matrices. Hence the result. ■

An example of almost $\bar{\mathbf{N}} \cap \mathbf{E}_0$-matrix is given below.

Example 5 Let $M = \begin{bmatrix} 0 & -1 & 0 & 2 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & -1 \\ 1 & 0 & 0 & 0 \end{bmatrix}$. Here M is an $\mathbf{E}_0 \cap$ almost $\mathbf{N}_0$-matrix. It is easy to see that $M \in$ almost $\bar{\mathbf{N}}$ since we can get M as a limit point of the sequence $M^{(k)} = \begin{bmatrix} -\frac{1}{k} & -1 & \frac{2}{k} & 2 \\ -\frac{1}{k} & -\frac{1}{k} & 1 & \frac{2}{k} \\ \frac{4}{k} & 1 & -\frac{1}{k} & -1 \\ 1 & \frac{2}{k} & -\frac{1}{k} & -\frac{1}{k} \end{bmatrix}$ of almost $\mathbf{N}$-matrices which converges to M as $k \to \infty$.

We need the following results in sequel.

Theorem 1.6.13 ([38, 48]) *If $M \in \mathbf{R}_0$ and LCP(q, M) has an odd number of solutions for a nondegenerate q, then $M \in \mathbf{Q}$.*

Theorem 1.6.14 ([40, p. 1271]) *Suppose $M \in \mathbf{Q}$ ($\mathbf{Q}_0$). Assume that $M_{i\cdot} \geq 0$ for some $i \in \{1, 2, \ldots, n\}$. Then $M_{\alpha\alpha} \in \mathbf{Q}$ ($\mathbf{Q}_0$), where $\alpha = \{1, 2, \ldots, n\}\backslash\{i\}$.*

Theorem 1.6.15 ([48, p. 45]) *A sufficient condition for LCP(q, M) to have even number of solutions for all q for which each solution is nondegenerate is that there exists a vector $z > 0$ such that $z^T M < 0$.*

Theorem 1.6.16 ([43]) *Suppose $M \in \mathbb{R}^{n\times n}$ is an almost $\bar{\mathbf{N}} \cap \mathbf{Q}_0 \cap \mathbf{E}_0$-matrix with $n \geq 4$. Then, there exists a principal rearrangement B of M such that all the leading principal submatrices of B are $\mathbf{Q}_0$-matrices.*

Proof Note that M is an almost $\bar{\mathbf{N}} \cap \mathbf{Q}_0 \cap \mathbf{E}_0$-matrix with $n \geq 4$. Then by Theorem 1.6.12 there exists a principal rearrangement

$$B = \begin{bmatrix} B_{\alpha\alpha} & B_{\alpha\bar{\alpha}} \\ B_{\bar{\alpha}\alpha} & B_{\bar{\alpha}\bar{\alpha}} \end{bmatrix}$$

of A such that $B_{\alpha\alpha}$, $B_{\bar{\alpha}\bar{\alpha}}$ are nonpositive strict upper triangular matrices and $B_{\bar{\alpha}\alpha}$, $B_{\alpha\bar{\alpha}}$ are nonnegative matrices. It is easy to conclude from the structure of B that $B_{n\cdot} \geq 0$. Note that $B \in \mathbf{Q}_0$, since B is a principal rearrangement of A. Therefore, by Theorem 1.6.14, $B_{\beta\beta} \in \mathbf{Q}_0$ where $\beta = \{1, 2, \ldots, n\}\backslash\{n\}$. Repeating the same argument, it follows that all leading principal submatrices of B are $\mathbf{Q}_0$. ■

Theorem 1.6.17 ([43]) *Suppose* $M \in$ *almost* $\bar{\mathbf{N}} \cap \mathbb{R}^{n\times n}, n \geq 4$ *with* $v(M) > 0$. *Then* $M \in \mathbf{Q}$ *if* $M \in \mathbf{R}_0$.

Proof Let $M \in$ almost $\bar{\mathbf{N}} \cap \mathbf{R}_0$. Then by Lemma 1.6.3, there exists $\emptyset \neq \alpha \subseteq \{1, 2, \ldots, n\}$, $M = \begin{bmatrix} M_{\alpha\alpha} & M_{\alpha\bar{\alpha}} \\ M_{\bar{\alpha}\alpha} & M_{\bar{\alpha}\bar{\alpha}} \end{bmatrix}$ where $M_{\alpha\alpha} \leq 0$, $M_{\bar{\alpha}\bar{\alpha}} \leq 0$, $M_{\bar{\alpha}\alpha} \geq 0$ and $M_{\alpha\bar{\alpha}} \geq 0$.

Now consider $M_{\alpha\alpha}$. Suppose $M_{\alpha\alpha}$ contains a nonnegative column vector. Then clearly LCP$(0, M)$ has a nontrivial solution, which contradicts our hypothesis that $M \in \mathbf{R}_0$. Hence, every column of $M_{\alpha\alpha}$ should have at least one negative entry. Hence there exists an $x \in \mathbb{R}^{|\alpha|}$, $x > 0$, such that $x^T M_{\alpha\alpha} < 0$. It now follows from Theorem 1.6.15 that for any $q_\alpha > 0$, where q_α is nondegenerate with respect to $M_{\alpha\alpha}$, LCP$(q_\alpha, M_{\alpha\alpha})$ has r solutions ($r \geq 2$ and even). Similarly, LCP$(q_{\bar{\alpha}}, M_{\bar{\alpha}\bar{\alpha}})$ has s solutions ($s \geq 2$ and even) for any $q_{\bar{\alpha}} > 0$, where $q_{\bar{\alpha}}$ is nondegenerate with respect to $M_{\bar{\alpha}\bar{\alpha}}$. Now suppose (w^i_α, z^i_α) is a solution for LCP$(q_\alpha, M_{\alpha\alpha})$. Note that $w = \begin{bmatrix} w^i_\alpha \\ q_{\bar{\alpha}} \end{bmatrix}$ and $z = \begin{bmatrix} z^i_\alpha \\ 0 \end{bmatrix}$ solves LCP(q, M). Similarly, associated with every solution $(w^i_{\bar{\alpha}}, z^i_{\bar{\alpha}})$ we can construct a solution of LCP(q, M). Thus, LCP(q, M) has $(r + s - 1)$ solutions accounting for only once the solution $w = q, z = 0$. Thus, there are an odd number $(r + s - 1 \geq 3)$ of solutions to LCP(q, M) with all solutions nondegenerate. We shall show that $(r + s - 1) \leq 3$ and hence there are only 3 solutions to LCP(q, M). Since q is nondegenerate with respect to M, this is a finite set [38, p. 85]. Suppose $(\bar{w}, \bar{z})$ is a nondegenerate solution to LCP(q, M). Then $(\bar{w}, \bar{z}) \in S(q, M)$. Now since M is a limit point of almost $\mathbf{N}$-matrices $\{M^{(k)}\}$, we note that the complementary basis corresponding $(\bar{w}, \bar{z})$ will also yield a solution to LCP$(q, M^{(k)})$ for all k sufficiently large. By Theorem 3.2 [32, p. 625], which asserts that there are exactly 3 solutions for LCP$(q, M^{(k)})$ for any nondegenerate $q(> 0)$ with respect to $M^{(k)}$, we obtain $(r + s - 1) \leq |S(q, M)| \leq |S(q, M^{(k)})| = 3$. But $(r + s - 1) \geq 3$. Hence LCP(q, M) has exactly 3 solutions for any nondegenerate $q(> 0)$ with respect to M. Since $M \in \mathbf{R}_0$ and LCP(q, M) has an odd number of solutions, it follows from Theorem 1.6.13 that $M \in \mathbf{Q}$. ■

1.6.3 Fully Cospositive Matrices

In this section, we discuss about a class of matrices that are defined based on principal pivot transforms and show that the matrices in this class have nonnegative principal minors. A matrix M is said to be *fully copositive* ($\mathbf{C}^f_0$) if $\wp_\alpha(M)$ is a $\mathbf{C}_0$-matrix for all $\alpha \subseteq \{1, \ldots, n\}$. It is known that $\mathbf{C}^f_0 \cap \mathbf{Q}_0$ matrices are sufficient. The elements of $\mathbf{C}^f_0 \cap \mathbf{Q}_0$ are completely $\mathbf{Q}_0$-matrices [41] and share many properties of positive semidefinite (PSD) matrices. Symmetric $\mathbf{C}^f_0 \cap \mathbf{Q}_0$ matrices are PSD.

Theorem 1.6.18 ([41, Theorem 4.5]) *Suppose* $M \in \mathbf{C}_0^f \cap \mathbf{Q}_0$. *Then* $M \in \mathbf{P}_0$.

Theorem 1.6.19 ([41, Theorem 3.3]) *Let* $M \in \mathbf{C}_0^f$. *The following statements are equivalent:*

(a) M *is a* $\mathbf{Q}_0$*-matrix.*
(b) for every PPT M' *of* M, $m'_{ii} = 0 \Rightarrow m'_{ij} + m'_{ji} = 0, \ \forall\, i, j \in \{1, 2, \ldots, n\}$.
(c) M *is a completely* $\mathbf{Q}_0$*-matrix.*

Theorem 1.6.20 ([41, Theorem 4.9]) *If* $M \in \mathbb{R}^{2\times 2} \cap \mathbf{C}_0^f \cap \mathbf{Q}_0$, *then* M *is a PSD matrix.*

Theorem 1.6.21 ([6, Theorem 2′, p.73]) $M \in \mathbb{R}^{n\times n}$ *is sufficient if and only if every matrix obtained from it by means of a PPT operation is sufficient of order 2.*

As a consequence we have the following theorem.

Theorem 1.6.22 ([36]) *Let* $M \in \mathbf{C}_0^f \cap \mathbf{Q}_0$. *Then* M *is sufficient.*

Proof Note that all 2×2 submatrices of M or its PPTs are $\mathbf{C}_0^f \cap \mathbf{Q}_0$ matrices since M and $\wp_\alpha(M)$ are completely $\mathbf{Q}_0$-matrices. Now by Theorem 1.6.20, all 2×2 submatrices of M or $\wp_\alpha(M)$ for all α are positive semidefinite, and hence sufficient. Therefore, M or every matrix obtained by means of a PPT operation is sufficient of order 2. Now by Theorem 1.6.21, M is sufficient. ■

1.7 Hidden Z-Matrices

The class of hidden **Z**-matrices generalizes the class of **Z**-matrices. Mangasarian introduced this generalization for studying the class of linear complementarity problems solvable as linear programs [10, 25–27]. Let us recall the definition of hidden **Z**-matrix.

Definition 2 A matrix $M \in \mathbb{R}^{n\times n}$ is said to be a hidden **Z**-matrix if there exist Z-matrices $X,\ Y \in \mathbb{R}^{n\times n}$ and $r, s \in \mathbb{R}^n_+$ satisfying the following two conditions:
(i) $MX = Y$,
(ii) $r^T X + s^T Y > 0$.

The class of hidden **Z**-matrices is denoted by hidden **Z**. Pang [47] established a necessary and sufficient condition for a hidden **Z**-matrix to be a **P**-matrix. Many of the results which hold for the **Z** class admit an extension to the hidden **Z** class [8, 47]. The idea of solving a LCP as linear programs follows from well known fact that if LCP has a solution then the solution is one of the extreme points of the feasible set $S(q, M)$. Therefore, if an appropriate linear form is known whose minimum over $S(q, M)$ would necessarily occur at a complementary solution then LCP could be solved as an LP [7]. Mangasarian observed the following result for solving LCPs as linear programs.

Proposition 1 *If the linear complementarity problem LCP*(q, M) *has a solution then there exist a* $p \in \mathbb{R}^n$ *such that the linear program LP (*$\min p^T z$ *subject to* $w = Mz + q \geq 0, z \geq 0$*) has a (unique) solution* $\bar{z}$ *that also solves LCP*(q, M).

Mangasarian [25] also obtain the expressions for such p for the class hidden $\mathbf{Z}$ matrices as stated in the following theorem.

Theorem 1.7.23 ([25]) *Let* $M \in$ *hidden* $\mathbf{Z}$ *and* $F(q, M) \neq \emptyset$. *Then the LCP* (q, M) *has a solution which can be obtained by solving the linear program* $LP(p, q, M)$:

$$\begin{aligned} \min \quad & p^T x \\ \text{subject to} \quad & \\ & q + Mx \geq 0 \\ & x \geq 0, \end{aligned}$$

where $p = r + M^T s$, r *and* s *are as in the Definition 2.*

The following result identifies some more subclasses of hidden $\mathbf{Z}$-matrices, where the vector p can be easily specified in the following theorem:

Lemma 1.7.4 ([10]) *Let* $M \in \mathbb{R}^{n \times n}$ *be a hidden* $\mathbf{Z}$*-matrix. Let* $\wp_\alpha(M)$ *be a PPT of* M *with respect to* $\alpha \subseteq \{1, \ldots, n\}$. *Then* $\wp_\alpha(M)$ *is a hidden* $\mathbf{Z}$*-matrix.*

Proof Let $M \in$ hidden $\mathbf{Z}$, X and Y are any two $\mathbf{Z}$-matrices satisfying the conditions in Definition 2 with $r,\ s \geq 0$ and $\alpha \subseteq \{1, \ldots, n\}$. Suppose M, X, and Y are partitioned as follows:

$$M = \begin{bmatrix} M_{\alpha\alpha} & M_{\alpha\bar{\alpha}} \\ M_{\bar{\alpha}\alpha} & M_{\bar{\alpha}\bar{\alpha}} \end{bmatrix},\ X = \begin{bmatrix} X_{\alpha\alpha} & X_{\alpha\bar{\alpha}} \\ X_{\bar{\alpha}\alpha} & X_{\bar{\alpha}\bar{\alpha}} \end{bmatrix},\ Y = \begin{bmatrix} Y_{\alpha\alpha} & Y_{\alpha\bar{\alpha}} \\ Y_{\bar{\alpha}\alpha} & Y_{\bar{\alpha}\bar{\alpha}} \end{bmatrix}.$$

Then by Lemma 1.3.1 [47], we have

$$\begin{bmatrix} M_{\alpha\alpha}^{-1} & -M_{\alpha\alpha}^{-1} M_{\alpha\bar{\alpha}} \\ M_{\bar{\alpha}\alpha} M_{\alpha\alpha}^{-1} & M_{\bar{\alpha}\bar{\alpha}} - M_{\bar{\alpha}\alpha} M_{\alpha\alpha}^{-1} M_{\alpha\bar{\alpha}} \end{bmatrix} \begin{bmatrix} Y_{\alpha\alpha} & Y_{\alpha\bar{\alpha}} \\ X_{\bar{\alpha}\alpha} & X_{\bar{\alpha}\bar{\alpha}} \end{bmatrix} = \begin{bmatrix} X_{\alpha\alpha} & X_{\alpha\bar{\alpha}} \\ Y_{\bar{\alpha}\alpha} & Y_{\bar{\alpha}\bar{\alpha}} \end{bmatrix}.$$

Let $\tilde{X} = \begin{bmatrix} Y_{\alpha\alpha} & Y_{\alpha\bar{\alpha}} \\ X_{\bar{\alpha}\alpha} & X_{\bar{\alpha}\bar{\alpha}} \end{bmatrix}$ and $\tilde{Y} = \begin{bmatrix} X_{\alpha\alpha} & X_{\alpha\bar{\alpha}} \\ Y_{\bar{\alpha}\alpha} & Y_{\bar{\alpha}\bar{\alpha}} \end{bmatrix}$. Note that $\tilde{X},\ \tilde{Y} \in \mathbf{Z}$. Define $\tilde{r} = (s_\alpha, r_{\bar{\alpha}})$ and $\tilde{s} = (r_\alpha, s_{\bar{\alpha}})$. Note that $\tilde{r}$ and $\tilde{s}$ are nonnegative and $\tilde{r}^T \tilde{X} + \tilde{s}^T \tilde{Y} = r^T X + s^T Y > 0$. Therefore, $\wp_\alpha(M) \in$ hidden $\mathbf{Z}$. ■

Mangasarian [26] gave a table consisting of a summary of the cases for which LCP (q, M) can be solvable as LP (p, q, M), where p is specified along with the conditions on M.

A partial table from [26] is given below.

In the next theorem, we identify some more subclasses of hidden $\mathbf{Z}$-matrices where the vector p can be easily specified in the following theorem:

Table 1.1 Vector p in the linear program

Matrix M	Condition on M	Vector p in the LP
$M = Z_2 Z_1^{-1}$, $r^T Z_1 + s^T Z_2 > 0, r, s \geq 0$	$Z_1, Z_2 \in \mathbf{Z}$	$p = r + M^T s$
M	$M \in \mathbf{Z}$	$p > 0$
M	$M^{-1} \in \mathbf{Z}$	$p = M^T s,\ s > 0$
M	$M > 0,\ n \geq 3$	$p = e$ where e is a vector of all 1's
M	$M_{jj} \geq \sum_{i \neq j} \lvert M_{ij} \rvert, ; j = 1, 2, \ldots, n$	$p = M^T e$

Theorem 1.7.24 ([10]) *Consider the LCP* (q, M). *Let* $\wp_\alpha(M)$ *be the PPT of* M *with respect to* $\alpha \subseteq \{1, \ldots, n\}$, *which belongs to the subclasses of hidden* $\mathbf{Z}$*-matrices listed in the Table 1.1 [26]. Then the LCP* $(q', \wp_\alpha(M))$ *obtained from taking the PPT of LCP* (q, M) *with respect to* $\alpha \subseteq \{1, \ldots, n\}$ *can be solved by solving the linear program* $LP(p', q', \wp_\alpha(M))$:

$$\begin{aligned} \min \quad & p'^T x \\ \text{subject to} \quad & \\ & q' + \wp_\alpha(M)x \geq 0 \\ & x \geq 0, \end{aligned}$$

where p' *is specified in the Table 1.1 [26].*

Proof Suppose $M \in$ hidden $\mathbf{Z}$ matrix which does not belong to the classes listed in Table 1.1 [26] and there exists a PPT $\wp_\alpha(M)$ where $\alpha \subseteq \{1, \ldots, n\}$ such that $\wp_\alpha(M)$ belongs to one of the classes listed in Table 1.1 [26]. Note that PPT of LCP (q, M) with respect to $\alpha \subseteq \{1, \ldots, n\}$ is given by LCP $(q', \wp_\alpha(M))$. Further, note that $|S(q, M)| = |S(q', \wp_\alpha(M))|$ and we can obtain a solution of LCP (q, M) by solving LCP $(q', \wp_\alpha(M))$. Since $\wp_\alpha(M)$ belongs to the class listed in the Table 1.1 [26], we can take p' as 1 as p specified in the Table 1.1 [26]. By Lemma 1.7.4 and Theorem 1.7.23, it follows that by solving $LP(p', q', \wp_\alpha(M))$ we can obtain a solution of LCP $(q', \wp_\alpha(M))$ and hence we can obtain a solution of LCP (q, M). ■

The following example demonstrates that the above scheme extends the class of LCPs which can be solved as a linear program.

Example 6 Consider the following matrix

$$M = \begin{bmatrix} -1 & -1 & -13 \\ 3 & 1 & 7 \\ -5 & -1 & -14 \end{bmatrix}.$$

Note that its inverse is $\begin{bmatrix} 0.2692 & 0.0385 & -0.2308 \\ -0.2692 & 1.9615 & 1.2308 \\ -0.0769 & -0.1538 & -0.0769 \end{bmatrix}$, which is not a **Z**-matrix. But,

$$\wp_\alpha(M) = \begin{bmatrix} 2 & -1 & -6 \\ -3 & 1 & -7 \\ -2 & -1 & -7 \end{bmatrix}$$

with respect to $\alpha = \{2\}$ is a **Z**-matrix.

Remark 3 Mangasarian [26] provides the following example of LCP(q, M) where $M = \begin{bmatrix} 0 & 3 & 4 \\ 1 & -1 & 0 \\ 0 & -1 & -3 \end{bmatrix}$, $q = \begin{bmatrix} -2 \\ 0 \\ 1 \end{bmatrix}$ for which the solution can be obtained by solving LP(p, q, M) with $p = M^T e = e$ but application of Lemke's algorithm on LCP(q, M) leads to ray termination.

1.8 Various Generalizations of LCP

In this section, we discuss various generalizations of the linear complementarity problem appeared in the literature. A number of generalizations of the linear complementarity problem have been proposed by several researchers in the context of real-life problems arising from management, engineering, or game theoretical applications. Researchers over the decade have developed theory and algorithms for each of the generalizations exclusively. These generalizations deals with various types of mixed complementarity conditions for which standard literature is not available.

1.8.1 Vertical Linear Complementarity Problem

While defining LCP(q, M), it is assumed that the given matrix is a square matrix. However, in many real-life applications, we may not get a square matrix and each complementarity pair may not exist as it appears in the definition of the problem LCP(q, M). In order to overcome the difficulties associated with a square matrix, the concept of a vertical block matrix (a rectangular matrix) was introduced by Cottle and Dantzig [4] in connection with the generalization of the linear complementarity problem.

Consider a rectangular matrix $\mathcal{A}$ of order $m \times k$ with $m \geq k$. Suppose $\mathcal{A}$ is partitioned row-wise into k blocks in the form

$$\mathcal{A} = \begin{bmatrix} \mathcal{A}^1 \\ \mathcal{A}^2 \\ \vdots \\ \mathcal{A}^k \end{bmatrix}$$

where each $\mathcal{A}^j = ((a_{rs}^j)) \in \mathbb{R}^{m_j \times k}$ with $\sum_{j=1}^{k} m_j = m$. Then $\mathcal{A}$ is called a *vertical block matrix of type* $(m_1, \ldots, m_k)$. If $m_j = 1, \ \forall\ j = 1, \ldots, k$, then $\mathcal{A}$ is a square matrix. The rth block of $\mathcal{A}$ is denoted by $\mathcal{A}^r$ and is a matrix of order $m_r \times k$. We then use the notation $J_1 = \{1, 2, \ldots, m_1\}$ to denote the set of row indices of the first block in $\mathcal{A}$ and $J_r = \left\{ \sum_{j=1}^{r-1} m_j + 1, \sum_{j=1}^{r-1} m_j + 2, \ldots, \sum_{j=1}^{r} m_j \right\}$ to denote the set of row indices of the rth block in $\mathcal{A}$ for $r = 2, 3, \ldots, k$. A vertical block matrix is a natural generalization of a square matrix. For example, the vertical block matrix structure given above arises naturally in the literature of stochastic games where the states are represented by the columns and actions in each state are represented by rows in a particular block. See [34, 35].

We shall now present a generalization of the linear complementarity problem by Cottle and Dantzig [4] involving a vertical block matrix known as vertical linear complementarity problem and it is stated as follows:

Given a vertical block matrix $\mathcal{A}$ of type $(m_1, \ldots, m_k)$ and a vector $q \in \mathbb{R}^m$, the vertical linear complementarity problem (VLCP$(q, \mathcal{A})$) is to find $w \in \mathbb{R}^m$ and $z \in \mathbb{R}^k$ such that

$$w - \mathcal{A}z = q, \ w \geq 0, \ z \geq 0 \tag{1.9}$$

$$z_j \prod_{i=1}^{m_j} w_i^j = 0, \ \text{for } j = 1, 2, \ldots, k. \tag{1.10}$$

Cottle–Dantzig's generalization was designated later by the name *vertical linear complementarity problem* [8] and this problem is denoted as VLCP$(q, \mathcal{A})$. For details on vertical linear complementarity problem see [30, 33] and the references therein. Ebiefung and Kostreva [12] presented a generalized version of Leontief input–output linear model as a vertical linear complementarity problem and mentioned that this model can be used for the problem of choosing a new technology, solving problems related to energy commodity demands, international trade, multinational army personnel assignment, and pollution control. Another general form of the VLCP$(q, \mathcal{A})$ arises in different areas of control theory through discretization of Hamilton–Jacobi–Bellman equations [52, 53]. Oh [44] formulated a mixed lubrication problem as a generalized nonlinear complementarity problem. Another nice application of the VLCP is the formulation of the global stability of a two-species piecewise linear Volterra ecosystem [14]. Gowda and Sznajder [19] present an extension of the bimatrix game model and the problem of computing a pair of equilibrium strategies for

this extended model leads to a VLCP formulation. This generalized bimatrix game model can be used in many applications in economics. A number of natural applications of vertical linear complementarity problem arise in stochastic games of special structure in the payoff and transition probability matrix. See [35] and the references cited therein. This sort of applications and the potential future applications have motivated to study VLCP theory and algorithms for the VLCP.

Mohan et al. [33] obtained an equivalent formulation of VLCP$(q, \mathcal{A})$ as LCP $(q, \mathcal{M})$ in order to extend various matrix theoretic results and applicability of Cottle Dantzig' algorithm (a generalization of Lemke's algorithm). The problem VLCP$(q, \mathcal{A})$ can be formulated as LCP$(q, \mathcal{M})$ as follows.

Consider a vertical block matrix $\mathcal{A}$ of type $(m_1, \ldots, m_k)$, where m_j is the size of the jth block. We construct an *equivalent square matrix* $\mathcal{M}$ of order $m \times m$ of $\mathcal{A}$ by copying $\mathcal{A}_{.j}$, m_j times for $j = 1, 2, \ldots, k$ (for example, $\mathcal{A}_{.1}$ is copied m_1 times, $\mathcal{A}_{.2}$ is copied m_2 times etc.). Thus $\mathcal{M}_{.p} = \mathcal{A}_{.s}\ \forall\ p \in J_s$. Note that $\mathcal{M}$ is singular if $m > k$. Mohan et al. [33] observe the following result. The proof of the following lemma presents a construction procedure for a solution (u, v) to LCP$(q, \mathcal{M})$ from a solution (w, z) of VLCP$(q, \mathcal{A})$.

Lemma 1.8.5 *Given the VLCP$(q, \mathcal{A})$, let $\mathcal{M}$ be the equivalent square matrix of $\mathcal{A}$. VLCP$(q, \mathcal{A})$ has a solution if and only if LCP$(q, \mathcal{M})$ has a solution.*

Proof We obtain a solution (u, v) to LCP$(q, \mathcal{M})$ from a solution (w, z) of VLCP $(q, \mathcal{A})$ as follows:

We choose $u = w$. Note that $z_j > 0$ implies $\exists\ p(j) \in J_j$ such that $w_{p(j)} = 0$. Define

$$v_r = \begin{cases} 0, & \text{if } r \neq p(j) \text{ for any } 1 \leq j \leq k \\ z_j, & \text{if } \exists \text{ a } j,\ 1 \leq j \leq k, \text{ such that } r = p(j) \end{cases}$$

Clearly, v_r is well defined. Now it is easy to see that (u, v) solves LCP$(q, \mathcal{M})$.

Conversely, suppose (u, v) is a solution to the LCP$(q, \mathcal{M})$. Define the vector $z \in \mathbb{R}^k$ by taking

$$z_j = \sum_{i \in J_j} v_i.$$

Note that if $z_j > 0$, $\exists\ i \in J_j$ such that $v_i > 0$ and hence $u_i = 0$. Hence with $w = u$, (w, z) solves VLCP$(q, \mathcal{A})$. ■

1.8.2 Scarf's Complementarity Problem

Scarf [50] introduced a generalization of the linear complementarity problem to diversify the field of applications. Scarf [50] introduced the following interesting generalization of the linear complementarity problem involving a vertical block matrix A of type $(m_1, m_2, \ldots, m_k)$ described in earlier section. Let $M^j(x)$ where $0 \leq x \in \mathbb{R}^k$

be k homogeneous linear functions, each of which is the maximum of a finite number of linear functions and $q = [q^1, q^2, \ldots, q^k]^T \in \mathbb{R}^k$ be a vector. Scarf posed the following problem. Under what conditions can we say that the equations

$$\begin{aligned} M^1(x) - r_1 &= q^1 \\ M^2(x) - r_2 &= q^2 \\ &\vdots \\ M^k(x) - r_k &= q^k \end{aligned}$$

have a solution in nonnegative variables x and r with $x_j r_j = 0$ for all j?

Note that the important difference between Scarf's problem and LCP (see [50]) is that each linear function is replaced by the maximum of several linear functions. Scarf [50] pointed out that if $M^j(x)$ were the minimum rather than the maximum of linear functions, the problem could be solved by a trivial reformulation of LCP.

A slightly generalized version of Scarf's complementarity problem stated by Lemke [23] is as follows.

Given an $m \times k$, $m \geq k$ vertical block matrix A of type $(m_1, m_2, \ldots, m_k)$ and $\bar{q} \in \mathbb{R}^m$ where $m = \sum_{j=1}^{k} m_j$, find $x \in \mathbb{R}^k$ such that

$$r_j(x) = \max_{i \in J_j} (A^j x + \bar{q}^j)_i \geq 0, \; j = 1, \ldots, k, \; x \geq 0 \tag{1.11}$$

$$\sum_{j=1}^{k} x_j r_j(x) = 0. \tag{1.12}$$

We refer to this generalization as Scarf's complementarity problem and denote this problem by SCP($\bar{q}$, A). Lemke [23] formulated the Scarf's complementarity problem as a linear complementarity problem LCP(q, M) but he remained silent about the processability of this problem by his algorithm. Lemke [23] showed that this formulation arises for calculating a vector in the core of an n person game [51].

1.8.3 Other Generalizations

We now briefly mention some more generalizations which are proposed in the literature to accommodate more real-life problems.

- *The Horizontal Linear Complementarity Problem*: Given two matrices A, $B \in \mathbb{R}^{n \times n}$ and a vector $q \in \mathbb{R}$, the horizontal linear complementarity problem (HLCP(q, A, B)) is to find vectors $x \in \mathbb{R}^n$ and $y \in \mathbb{R}^n$ such that

$$Ax - By = q, \quad x \geq 0, \; y \geq 0 \tag{1.13}$$

$$x^T y = 0. \tag{1.14}$$

The HLCP was apparently introduced by Samelson, Thrall, and Wesler [49], motivated by a problem in structural engineering. Clearly, this problem reduces to the standard problem LCP(q, M) when $A = I,\ B = M$.

- *The extended horizontal linear complementarity problem*: Consider a rectangular matrix C of order $n \times m$ $(m > n)$. Suppose C is partitioned into $(k+1)$ blocks of the form

$$\left[C^0\ C^1\ C^2 \cdots C^k \right]$$

where $C^j \in \mathbb{R}^{n\times n},\ j = 0, 1, 2, \ldots, k$ and $m = (k+1)n$. Let c be a block vector which is defined as q for $k = 1$ and as $[q, d^1, d^2, \ldots, d^{k-1}]$ for $k \geq 2$, where $q \in \mathbb{R}^n$ and $0 < d^j \in \mathbb{R}^n$ for $j = 1, 2, \ldots, k-1$. The extended horizontal LCP(c, C) is to find vectors $x^j \in \mathbb{R}^n,\ j = 0, 1, 2, \ldots, k$ such that

$$C^0 x^0 = q + \sum_{j=1}^{k} C^j x^j,$$

$$x^0 \wedge x^1 = 0,\ \ (d^j - x^j) \wedge x^{j+1} = 0,\ \ j = 1, 2, \ldots, k-1,$$

where for $k = 1$, only the first complementarity condition is considered. The above form of the extended HLCP has been considered by Sznajder and Gowda [54]. Kaneko [20] considers the extended HLCP for the case $C^0 = I$ and cites applications in mathematical programming and structural mechanics. See [21, 46, 55] for applications in inventory theory, statistics, and modeling piecewise linear electrical networks. The study of HLCP is important due to the fact that any piecewise linear system can be formulated as a HLCP.

- *Extended Generalized Order Linear Complementarity Problem*: Given a block matrix $B \in \mathbb{R}^{n(k+1)\times n}$ and a block vector $b \in \mathbb{R}^{n(k+1)\times 1}$ where $B = [B^0, B^1, \ldots, B^k],\ B^j \in \mathbb{R}^{n\times n},\ j = 0, 1, \ldots, k$ and $b = [b^0, b^1, \ldots, b^k],\ b^j \in \mathbb{R}^n,\ j = 0, 1, \ldots, k$, the extended generalized order linear complementarity problem (EGOLCP (b, B)) is to find $z \in \mathbb{R}^n$ such that

$$(B^0 z + b^0) \wedge (B^1 z + b^1) \wedge (B^2 z + B^2) \wedge \cdots \wedge (B^k z + b^k) = 0. \tag{1.15}$$

This was introduced by Gowda and Sznajder [18]. The problem reduces to generalized order linear complementarity problem (GOLCP) by taking $B^0 = I$.

- *Generalized LCP of Ye*: This was introduced by Ye [56]. Given matrices $A, B \in \mathbb{R}^{m\times n}$, $C \in \mathbb{R}^{m\times k}$ and a vector $q \in \mathbb{R}^m$, find $x, y \in \mathbb{R}^n$ and $z \in \mathbb{R}^k$ such that

$$Ax + By + Cz = q,\ \ x, y, z \geq 0,$$

$$x^T y = 0.$$

As mentioned in [56], this generalized LCP arises in economic equilibrium problems, noncooperative games, traffic assignment problems, and, of course, in optimization problems. It is related to the variational inequality problem, the stationary point problem, bilinear programming problem, and nonlinear equations.

- *Mangasarian and Pang's Extended Linear Complementarity Problem*: This generalization of LCP was introduced by Mangasarian and Pang [28]. Given two matrices $M, N \in \mathbb{R}^{m\times n}$ and a polyhedral set $\mathcal{K}$ in $\mathbb{R}^m$, Mangasarian and Pang's extended linear complementarity problem (denoted by XLCP$(M, N, \mathcal{K})$) is to find two vectors $x, y \in \mathbb{R}^n$ such that

$$Mx - Ny \in \mathcal{K}, \quad x \geq 0, y \geq 0,$$

$$x^T y = 0.$$

The generalized LCP of Ye and XLCP are equivalent [17].

- *The Mixed Linear Complementarity Problem*: Given matrices $A \in \mathbb{R}^{n\times n}$, $B \in \mathbb{R}^{n\times m}$, $C \in \mathbb{R}^{m\times n}$, $D \in \mathbb{R}^{m\times m}$ and vectors $a \in \mathbb{R}^n$, $b \in \mathbb{R}^m$, find $x \in \mathbb{R}^n$ and $y \in \mathbb{R}^m$ such that

$$Ax + By + a = 0, \; Cx + Dy + b \geq 0, \; y \geq 0,$$

$$y^T(Cx + Dy + b) = 0.$$

In this formulation x is the free variable. If A is a nonsingular square matrix then MLCP is equivalent to LCP [8].

- *Extended Linear Complementarity Problem*: Given two matrices $C \in \mathbb{R}^{p\times n}$, $D \in \mathbb{R}^{q\times n}$, two vectors $c \in \mathbb{R}^p$, $d \in \mathbb{R}^q$ and m subsets $\theta_1, \ldots, \theta_m$ of $\{1, 2, \ldots, p\}$, the extended linear complementarity problem (ELCP(C, D, c, d, Θ)) is to find $x \in \mathbb{R}^n$ such that

$$Cx \geq c \tag{1.16}$$

$$Dx = d \tag{1.17}$$

$$\prod_{i\in\theta_j} (Cx - c)_i = 0, \; \forall \; j = 1, 2, \ldots, m, \tag{1.18}$$

or show that no such vector exists.
Here, $\Theta = \{\theta_1, \ldots, \theta_m\}$ is the collection of subsets θ_j of $\{1, 2, \ldots, p\}$ and $|\Theta| = m$. If $x \in \mathbb{R}^n$ satisfies (1.16) and (1.17) then the problem ELCP(C, D, c, d, Θ) is said to have a *feasible solution.* The *complementarity condition* (1.18) implies that for each set θ_j, $j = 1, 2, \ldots, m$ corresponds to a group of inequalities in $Cx \geq c$ and for each θ_j at least one inequality should hold as equality. If the feasible solution $x \in \mathbb{R}^n$ satisfies the complementarity condition (1.18) then we say that it is a solution of ELCP(C, D, c, d, Θ). This generalization is proposed

by Schutter and De Moor [9] and it is shown that the generalization like VLCP, HLCP, XLCP etc can be obtained as a special case of ELCP. The formulation of ELCP arises in the study of discrete event systems, examples of which are flexible manufacturing systems, subway traffic networks, parallel processing systems, and telecommunication networks. Many important problems in the max algebra such as solving a set of multivariate polynomial equalities and inequalities, matrix decompositions, state-space transformations, minimal state-space realization of max-linear discrete event systems and some problems in structured stochastic game can be reformulated as an ELCP. Schutter and De Moor [9] have proposed an algorithm for solving ELCP(C, D, c, d, Θ).

Acknowledgements The authors would like to thank the anonymous referees for their constructive suggestions, which considerably improve the overall presentation of the chapter. The first author wants to thank the Science and Engineering Research Board, DST, Government of India for financial support for this research.

References

1. Chung, S.J.: NP-completeness of the linear complementarity problem. J. Optim. Theory Appl. **60**, 393–399 (1989)
2. Crouzeix, J.-P., Hassouni, A., Lahlou, A., Schaible, S.: Positive subdefinite matrices, generalized monotonicity and linear complementarity problems. SIAM J. Matrix Anal. Appl. **22**, 66–85 (2000)
3. Cottle, R.W.: The principal pivoting method revisited. Math. Program. **48**, 369–385 (1990)
4. Cottle, R.W., Dantzig, G.B.: A generalization of the linear complementarity problem. J. Comb. Theory **8**, 79–90 (1970)
5. Cottle, R.W., Ferland, J.A.: Matrix-theoretic criteria for the quasiconvexity and pseudoconvexity of quadratic functions. Linear Algebr. Its Appl. **5**, 123–136 (1972)
6. Cottle, R.W., Guu, S.-M.: Two characterizations of sufficient matrices. Linear Algebr. Its Appl. **170**, 56–74 (1992)
7. Cottle, R.W., Pang, J.S.: On solving linear complementarity problems as linear programs. Math. Program. Study **7**, 88–107 (1978)
8. Cottle, R.W., Pang, J.S., Stone, R.E.: The Linear Complementarity Problem. Academic Press, Boston (1992)
9. De Schutter, B., De Moor, B.: The extended linear complementarity problem. Math. Program. **71**, 289–325 (1995)
10. Dubey, D., Neogy, S.K.: On hidden **Z**-matrices and the linear complementarity problem. Linear Algebr. Its Appl. **496**, 81–100 (2016)
11. Eaves, B.C.: The linear complementarity problem. Manag. Sci. **17**, 612–634 (1971)
12. Ebiefung, A.A., Kostreva, M.: The generalized Leontief input-output model and its application to the choice of new technology. Ann. Oper. Res. **44**, 161–172 (1993)
13. Ferris, M.C., Pang, J.S.: Engineering and economic applications of complementarity problems. SIAM Rev. **39**, 669–713 (1997)
14. Habetler, G.J., Haddad, C.A.: Global stability of a two-species piecewise linear Volterra ecosystem. Appl. Math. Lett. **5**(6), 25–28 (1992)
15. Garcia, C.B.: Some classes of matrices in linear complementarity theory. Math. Program. **5**, 299–310 (1973)
16. Gowda, M.S.: Affine pseudomonotone mappings and the linear complementarity problem. SIAM J. Matrix Anal. Appl. **11**, 373–380 (1990)

17. Gowda, M.S.: On the extended linear complementarity problem. Math. Program. **72**, 33–50 (1996)
18. Gowda, M.S., Sznajder, R.: The generalized order linear complementarity problem. SIAM J. Matrix Anal. Appl. **15**, 779–795 (1994)
19. Gowda, M.S., Sznajder, R.: A generalization of the Nash equilibrium theorem on bimatrix games. Int. J. Game Theory **25**, 1–12 (1996)
20. Kaneko, I.: A linear complementarity problem with an n by $2n$ "P"-matrix. Math. Program. Study **7**, 120–141 (1978)
21. Kaneko, I., Pang, J.S.: Some n by dn linear complementarity problems. Linear Algebr. Its Appl. **34**, 297–319 (1980)
22. Lemke, C.E.: Bimatrix equilibrium points and mathematical programming. Manag. Sci. **11**, 681–689 (1965)
23. Lemke, C.E.: Recent results on complementarity problems. In: Rosen, J.B., Mangasarian, O.L., Ritter, K. (eds.) Nonlinear Programming, pp. 349–384. Academic Press, New York (1970)
24. Lemke, C.E., Howson Jr., J.T.: Equilibrium points of bimatrix games. SIAM J. Appl. Math. **12**, 413–423 (1964)
25. Mangasarian, O.L.: Linear complementarity problems sovable by a single linear program. Math. Program. **10**, 263–270 (1976)
26. Mangasarian, O.L.: Characterization of linear complementarity problems as linear programs. Math. Program. Study **7**, 74–87 (1978)
27. Mangasarian, O.L.: Simplified characterization of linear complementarity problems as linear programs. Math. O.R. **4**, 268–273 (1979)
28. Mangasarian, O.L., Pang, J.S.: The extended linear complementarity problem. SIAM J. Matrix Anal. Appl. **16**, 359–368 (1995)
29. Martos, B.: Subdefinite matrices and quadratic forms. SIAM J. Appl. Math. **17**, 1215–1223 (1969)
30. Mohan, S.R., Neogy, S.K.: Generalized linear complementarity in a problem of n person games. OR Spektrum **18**, 231–239 (1996)
31. Mohan, S.R., Parthasarathy, T., Sridhar, R.: $\bar{N}$ matrices and the class Q. In: Dutta, B., et al. (eds.) Lecture Notes in Economics and Mathematical Systems, vol. 389, pp. 24–36. Springer, Berlin (1992)
32. Mohan, S.R., Parthasarathy, T., Sridhar, R.: The linear complementarity problem with exact order matrices. Math. Oper. Res. **19**, 618–644 (1994)
33. Mohan, S.R., Neogy, S.K., Sridhar, R.: The generalized linear complementarity problem revisited. Math. Program. **74**, 197–218 (1996)
34. Mohan, S.R., Neogy, S.K., Parthasarathy, T., Sinha, S.: Vertical linear complementarity and discounted zero-sum stochastic games with ARAT structure. Math. Program. Ser. A **86**, 637–648 (1999)
35. Mohan, S.R., Neogy, S.K., Parthasarathy, T.: Pivoting algorithms for some classes of stochastic games: a survey. Int. Game Theory Rev. **3**, 253–281 (2001)
36. Mohan, S.R., Neogy, S.K., Das, A.K.: On the class of fully copositive and fully semimonotone matrices. Linear Algebr. Its Appl. **323**, 87–97 (2001)
37. Mohan, S.R., Neogy, S.K., Das, A.K.: More on positive subdefinite matrices and the linear complementarity problem. Linear Algebr. Its Appl. **338**, 275–285 (2001)
38. Murty, K.G.: On the number of solutions to the linear complementarity problem and spanning properties of complementarity cones. Linear Algebr. Its Appl. **5**, 65–108 (1972)
39. Murty, K.G.: Linear Complementarity. Linear and Nonlinear Programming. Heldermann, Berlin (1988)
40. Murthy, G.S.R., Parthasarathy, T.: Some properties of fully semimonotone matrices. SIAM J. Matrix Anal. Appl. **16**, 1268–1286 (1995)
41. Murthy, G.S.R., Parthasarathy, T.: Fully copositive matrices. Math. Program. **82**, 401–411 (1998)
42. Murthy, G.S.R., Parthasarathy, T., Ravindran, G.: On copositive semi-monotone Q-matrices. Math. Program. **68**, 187–203 (1995)

43. Neogy, S.K., Das, A.K.: On almost type classes of matrices with Q-property. Linear Multilinear Algebr. **53**, 243–257 (2005)
44. Oh, K.P.: The formulation of the mixed lubrication problem as a generalized nonlinear complementarity problem. Trans. ASME, J. Tribol. **108**, 598–604 (1986)
45. Olech, C., Parthasarathy, T., Ravindran, G.: Almost N-matrices in linear complementarity. Linear Algebr. Its Appl. **145**, 107–125 (1991)
46. Pang, J.S.: On a class of least element complementarity problems. Math. Program. **16**, 111–126 (1976)
47. Pang, J.S.: Hidden Z-matrices with positive principal minors. Linear Algebra Appl. **23**, 201–215 (1979)
48. Saigal, R.: A characterization of the constant parity property of the number of solutions to the linear complementarity problem. SIAM J. Appl. Math. **23**, 40–45 (1972)
49. Samelson, H., Thrall, R.M., Wesler, O.: A partition theorem for Euclidean n-space. Proc. Am. Math. Soc. **9**, 805–807 (1958)
50. Scarf, H.E.: An algorithm for a class of non-convex programming problems. Cowles Commission Discussion Paper No. **211**, Yale University (1966)
51. Scarf, H.E.: The core of an N person game. Econometrics **35**, 50–69 (1967)
52. Sun, M.: Singular control problems in bounded intervals. Stochastics **21**, 303–344 (1987)
53. Sun, M.: Monotonicity of Mangasarian's iterative algorithm for generalized linear complementarity problems. J. Math. Anal. Appl. **144**, 474–485 (1989)
54. Sznajder, R., Gowda, M.S.: Generalizations of P_0- and P- properties; extended vertical and horizontal linear complementarity problems. Linear Algebr. Its Appl. **223**(224), 695–715 (1995)
55. Vandenberghe, L., De Moor, B., Vandewalle, J.: The generalized linear complementarity problem applied to the complete analysis of resistive piecewise-linear circuits. IEEE Trans. Circuits Syst. **11**, 1382–1391 (1989)
56. Ye, Y.: A fully polynomial time approximation algorithm for computing a stationary point of the General linear Complementarity Problem. Math. Oper. Res. **18**, 334–345 (1993)

Chapter 2
Maximizing Spectral Radius and Number of Spanning Trees in Bipartite Graphs

Ravindra B. Bapat

2.1 Introduction

We consider simple graphs which have no loops or parallel edges. Thus, a graph $G = (V, E)$ consists of a finite set of vertices, $V(G)$, and a set of edges, $E(G)$, each of whose elements is a pair of distinct vertices. We will assume familiarity with basic graph-theoretic notions, see, for example, Bondy and Murty [5].

There are several matrices that one normally associates with a graph. We introduce some such matrices which are important. Let G be a graph with $V(G) = \{1, \ldots, n\}$. The *adjacency matrix* A of G is an $n \times n$ matrix with its rows and columns indexed by $V(G)$ and with the (i, j)-entry equal to 1 if vertices i, j are adjacent and 0 otherwise. Thus, A is a symmetric matrix with its ith row (or column) sum equal to $d(i)$, which by definition is the degree of the vertex $i, i = 1, 2, \ldots, n$. Let D denote the $n \times n$ diagonal matrix, whose ith diagonal entry is $d(i), i = 1, 2, \ldots, n$. The *Laplacian matrix* of G, denoted by L, is the matrix $L = D - A$.

By the eigenvalues of a graph, we mean the eigenvalues of its adjacency matrix. Spectral graph theory is the study of the relationship between the eigenvalues of a graph and its structural properties. The spectral radius of a graph is the largest eigenvalue, in modulus, of the graph. It is a topic of much investigation. It evolved during the study of molecular graphs by chemists. We refer to [12] for the subject of spectral graph theory.

A connected graph without a cycle is called a tree. Trees constitute an important subclass of graphs both from theoretical and practical considerations. A spanning tree in a graph is a spanning subgraph which is a tree. Spanning trees arise in several applications. If we are interested in establishing a network of locations with minimal links, then it corresponds to a spanning tree. We may also be interested in the spanning

R. B. Bapat (✉)
Indian Statistical Institute, 7, S. J. S Sansanwal Marg, New Delhi 110016, India
e-mail: rbb@isid.ac.in

S. K. Neogy et al. (eds.), *Mathematical Programming and Game Theory*,
Indian Statistical Institute Series, https://doi.org/10.1007/978-981-13-3059-9_2

tree with the least weight, where each edge in the graph is associated a weight, and the weight of a spanning tree is the sum of the weights of its edges.

If G is connected, then L is singular with rank $n-1$. Furthermore, the well-known Matrix-Tree Theorem asserts that any cofactor of L equals the number of spanning trees $\tau(G)$ in G. For basic results concerning matrices associated with a graph, we refer to [2].

A graph G is bipartite if its vertex set can be partitioned as $V(G) = X \cup Y$ such that no two vertices in X, or in Y, are adjacent. We often denote the bipartition as (X, Y). A graph is bipartite if and only if it has no cycle of odd length.

The adjacency matrix of a bipartite graph G has a particularly simple form viewed as a partitioned matrix

$$A(G) = \begin{bmatrix} 0 & B \\ B' & 0 \end{bmatrix}.$$

This form is especially useful in dealing with matrices associated with a bipartite graph.

In this chapter, we consider two optimization problems over bipartite graphs under certain constraints. One of the problems is to maximize the spectral radius, while the other is to maximize the number of spanning trees.

We now describe the contents of this chapter. In Sect. 2.2, we introduce the class of Ferrers graphs which are bipartite graphs such that the edges of the graph are in direct correspondence with the boxes in a Ferrers diagram. This class is of interest in both the maximization problems that we consider.

The problem of maximizing the spectral radius of a bipartite graph is considered in Sect. 2.3. We give a brief survey of the problem and provide references to the literature containing results and open problems.

In Sect. 2.4, we state an elegant formula for the number of spanning trees in a Ferrers graph due to Ehrenborg and van Willigenburg [13]. We give references to the proofs of the formula available in the literature. The formula leads to a conjectured upper bound for the number of spanning trees in a bipartite graph and is considered in Sect. 2.5. A reformulation of the conjecture in terms of majorization due to Slone is described in Sect. 2.6.

Sections 2.7 and 2.8 contain new results. The concept of resistance distance [17] between two vertices in a graph captures the notion of the degree of communication in a better way than the classical distance. The resistance distance can be defined in several equivalent ways, see, for example [3]. It is known, and intuitively obvious, that the resistance distance between any two vertices does not decrease when an edge, which is not a cut edge, is deleted from the graph. In Sect. 2.7, we first give an introduction to resistance distance. We then examine the situation when the removal of an edge in a graph does not affect the resistance distance between the end vertices of another edge. Several equivalent conditions are given for this to hold. This result, which appears to be of interest by itself, is then used in Sect. 2.8 to give another proof of the formula for the number of spanning trees in a Ferrers graph. Ehrenborg and van Willigenburg [13] also use electrical networks and resistances in their proof of the formula but our approach is different.

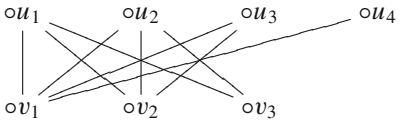
2.2 Ferrers Graphs

A Ferrers graph is defined as a bipartite graph on the bipartition (U, V), where $U = \{u_1, \ldots, u_m\}$, $V = \{v_1, \ldots, v_n\}$ such that

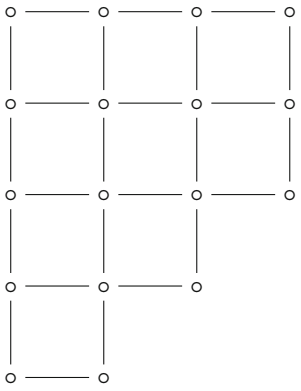
- if (u_i, v_j) is an edge, then so is (u_p, v_q), where $1 \leq p \leq i$ and $1 \leq q \leq j$, and
- (u_1, v_n) and (u_m, v_1) are edges.

For a Ferrers graph G, we have the associated partition $\lambda = (\lambda_1, \ldots, \lambda_m)$, where λ_i is the degree of vertex u_i, $i = 1, \ldots, m$, Similarly, we have the dual partition $\lambda' = (\lambda'_1, \ldots, \lambda'_n)$ where λ'_j is the degree of vertex v_j, $j = 1, \ldots, n$. Note that $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_m$ and $\lambda'_1 \geq \lambda'_2 \geq \cdots \geq \lambda'_n$. The associated Ferrers diagram is the diagram of boxes where we have a box in position (i, j) if and only if (u_i, v_j) is an edge in the Ferrers graph.

Example 1 The Ferrers graph with the degree sequences (3, 3, 2, 1) and (4, 3, 2) is shown below:



The associated Ferrers diagram is



The definition of Ferrers graph is due to Ehrenborg and van Willigenburg [13]. Chestnut and Fishkind [10] defined the class of bipartite graphs called *difference graphs.* A bipartite graph with parts X and Y is a difference graph if there exist a function $\phi : X \cup Y \to R$ and a threshold $\alpha \in R$ such that for all $x \in X$ and $y \in Y$, x is adjacent to y if and only if $\phi(x) + \phi(y) \geq \alpha$. It turns out that the class of Ferrers

graphs coincides with the class of difference graphs, as shown by Hammer et al. [16]. A more direct proof of this equivalence is given by Cheng Wai Koo [18]. The same class is termed *chain graphs* in [4].

2.3 Maximizing the Spectral Radius of a Bipartite Graph

We introduce some notation. Let $G = (V \cup W, E)$ be a bipartite graph, where $V = \{v_1, \ldots, v_m\}$, $W = \{w_1, \ldots, w_n\}$ are the two partite sets. We view the undirected edges E of G as a subset of $V \times W$. Let

$$D(G) = d_1(G) \geq d_2(G) \geq \cdots \geq d_m(G)$$

be the rearranged set of the degrees of $v_1, \ldots, v_m$. Note that $e(G) = \sum_{i=1}^{m} d_i(G)$ is the number of edges in G. Recall that the eigenvalues of G are simply the eigenvalues of the adjacency matrix of G. Since the adjacency matrix is entry-wise non-negative, it follows from the Perron–Frobenius Theorem that the spectral radius of the adjacency matrix is an eigenvalue of the matrix. Denote by $\lambda_{max}(G)$ the maximum eigenvalue of G. It is known [4] that

$$\lambda_{max}(G) \leq \sqrt{e(G)} \tag{2.1}$$

and equality occurs if and only if G is a complete bipartite graph, with possibly some isolated vertices.

We now consider refinements of (2.1) for non-complete bipartite graphs. For positive integers p, q, let $K_{p,q}$ be the complete bipartite graph $G = (V \cup W, E)$ where $|V| = p$, $|W| = q$. Let $\mathcal{K}(p, q, e)$ be the family of subgraphs of $K_{p,q}$ with e edges, with no isolated vertices, and which are not complete bipartite graphs. The following problem was considered in [4].

Problem 1 Let $2 \leq p \leq q$, $1 < e < pq$ be integers. Characterize the graphs which solve the maximization problem

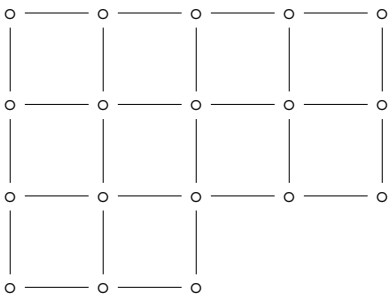
$$\max_{G \in \mathcal{K}(p,q,e)} \lambda_{max}(G). \tag{2.2}$$

Motivated by a conjecture of Brualdi and Hoffman [7] for non-bipartite graphs, which was proved by Rowlinson [21], the following conjecture was proposed in [4].

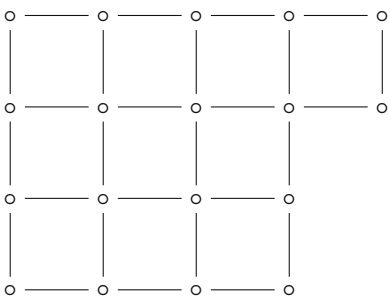
Conjecture 1 *Under the assumptions of Problem 1, an extremal graph that solves the maximal problem (2.2) is obtained from a complete bipartite graph by adding one vertex and a corresponding number of edges.*

As an example, consider the class $\mathcal{K}(3, 4, 10)$. There are two graphs in this class which satisfy the description in Conjecture 1. The graph G_1 is obtained from the

complete bipartite graph $K_{2,4}$ by adding an extra vertex of degree 2, and the graph G_2, is obtained from $K_{3,3}$ by adding an extra vertex of degree 1. The graph G_1 is associated with the Ferrers diagram



while G_2 is associated with the Ferrers diagram



It can be checked that $\lambda_{max}(G_2) = 3.0592 > \lambda_{max}(G_1) = 3.0204$. Thus according to Conjecture 1, G_2 maximizes $\lambda_{max}(G)$ over $G \in \mathcal{K}(3, 4, 10)$.

Conjecture 1 is still open, although some special cases have been settled, see [4, 14, 20, 23]. We now mention a result from [4] towards the solution of Problem 1 which is of interest by itself and is related to Ferrers graphs.

Let $D = \{d_1, d_2, \ldots, d_m\}$ be a set of positive integers where $d_1 \geq d_2 \geq \cdots \geq d_m$ and let $\mathcal{B}_D$ be the class of bipartite graphs $G = (X \cup Y, E)$ with no isolated vertices, with $|X| = m$, and with degrees of vertices in X being $d_1, \ldots, d_m$. Then, it is shown in [4] that $\max_{G \in \mathcal{B}_D} \lambda_{max}(G)$ is achieved, up to isomorphism, by the Ferrers graph, with the Ferrers diagram having $d_1, d_2, \ldots, d_m$ boxes in rows $1, 2, \ldots, m$, respectively.

It follows that an extremal graph solving Problem 1 is a Ferrers graph.

2.4 The Number of Spanning Trees in a Ferrers Graph

Definition 1 Let $G = (V, E)$ be a bipartite graph with bipartition $V = X \cup Y$. The Ferrers invariant of G is the quantity

$$F(G) = \frac{1}{|X||Y|} \prod_{v \in V} deg(v).$$

Recall that we denote the number of spanning trees in a graph G as $\tau(G)$. Ehrenborg and van Willigenburg [13] proved the following interesting formula.

Theorem 1 *If G is a Ferrers graph, then $\tau(G) = F(G)$.*

Let G be the Ferrers graph with bipartition (U, V), where $|U| = m$, $|V| = n$. We assume $U = \{u_1, \ldots, u_m\}$, $V = \{v_1, \ldots, v_n\}$. Let $d_1 \geq \cdots \geq d_m$ and $d'_1 \geq \cdots \geq d'_n$ be the degrees of $u_1, \ldots, u_m$ and $v_1, \ldots, v_n$, respectively. We may assume G to be connected, since otherwise, $\tau(G) = 0$. If G is connected, then $d_1 = |V|$ and $d'_1 = |U|$. Thus according to Theorem 1, $\tau(G) = d_2 \cdots d_m d'_2 \cdots d'_n$.

As an example, the Ferrers graph in Example 1 has degree sequences (3, 3, 2, 1) and (4, 3, 2). Thus, according to Theorem 1, it has $3 \cdot 2 \cdot 1 \cdot 3 \cdot 2 = 36$ spanning trees.

The complete graph $K_{m,n}$ has $m^{n-1}n^{m-1}$ spanning trees, and this can also be seen as a consequence of Theorem 1.

Theorem 1 can be proved in many ways. The proof given by Ehrenborg and van Willigenburg [13] is based on electrical networks. A purely bijective proof is given by Burns [8]. We give yet another proof based on resistance distance, which is different than the one in [13], see Sect. 2.8.

It is tempting to attempt a proof of Theorem 1 using the Matrix-Tree Theorem. As an example, the Laplacian matrix of the Ferrers graph in Example 1 is given by

$$L = \begin{bmatrix} 3 & 0 & 0 & 0 & -1 & -1 & -1 \\ 0 & 3 & 0 & 0 & -1 & -1 & -1 \\ 0 & 0 & 2 & 0 & -1 & -1 & 0 \\ 0 & 0 & 0 & 1 & -1 & 0 & 0 \\ -1 & -1 & -1 & -1 & 4 & 0 & 0 \\ -1 & -1 & -1 & 0 & 0 & 3 & 0 \\ -1 & -1 & 0 & 0 & 0 & 0 & 2 \end{bmatrix}.$$

Let $L(1|1)$ be the submatrix obtained from L be deleting the first row and column. According to the Matrix-Tree Theorem, the number of spanning trees in the graph is equal to the determinant of $L(1|1)$. Thus, Theorem 1 will be proved if we can evaluate the determinant of $L(1|1)$. But this does not seem easy in general.

A weighted analogue of Theorem 1 has also been given in [13] which we describe now. Consider the Ferrers graph G on the vertex partition $U = \{u_0, \ldots, u_n\}$ and $V = \{v_0, \ldots, v_m\}$. For a spanning tree T of G, define the weight $\sigma(T)$ to be

$$\sigma(T) = \prod_{p=0}^{n} x_p^{deg_T(u_p)} \prod_{q=0}^{m} y_q^{deg_T(v_q)},$$

where $x_0, \ldots, x_n; y_0, \ldots, y_m$ are indeterminates.

For a Ferrers graph G define $\Sigma(G)$ to be the sum $\Sigma(G) = \sum_T \sigma(T)$, where T ranges over all spanning trees T of G.

Theorem 2 ([13]) *Let G be the Ferrers graph corresponding to the partition λ and the dual partition λ'. Then*

$$\Sigma(G) = x_0 \cdots x_n \cdot y_0 \cdots y_m \prod_{p=1}^{n} (y_0 + \cdots + y_{\lambda_p - 1}) \prod_{q=1}^{m} (x_0 + \cdots + x_{\lambda'_q - 1}).$$

Theorem 1 follows from Theorem 2 by setting $x_0 = \cdots = x_n = y_0 = \cdots = y_m = 1$.

2.5 Maximizing the Number of Spanning Trees in a Bipartite Graph

For general bipartite graphs, the following conjecture was proposed by Ehrenborg [18, 22].

Conjecture 2 (Ferrers bound conjecture) *Let $G = (V, E)$ be a bipartite graph with bipartition $V = X \cup Y$. Then*

$$\tau(G) \leq \frac{1}{|X||Y|} \prod_{v \in V} deg(v),$$

that is, $\tau(G) \leq F(G)$.

Conjecture 2 is open in general. In this section, we describe some partial results towards its solution, mainly from [15, 18]. The following result has been proved in [15].

Theorem 3 *Let G be a connected bipartite graph for which Conjecture 2 holds. Let u be a new vertex not in $V(G)$, and let v be a vertex in $V(G)$. Let G' be the graph obtained by adding the edge $\{u, v\}$ to G. Then Conjecture 2 holds for G' as well.*

Note that Conjecture 2 clearly holds for the graph consisting of a single edge. Any tree can be constructed from such a graph by repeatedly adding a pendant vertex. Thus as an immediate consequence of Theorem 3, we get the following.

Corollary 1 *Conjecture 2 holds when the graph is a tree.*

Using explicit calculations with homogeneous polynomials, the following result is also established in [15].

Theorem 4 *Let G be a bipartite graph with bipartition $X \cup Y$. Then Conjecture 2 holds when $|X| \leq 5$.*

The following result is established in [18].

Proposition 1 *Let G and G' be bipartite graphs for which Conjecture 2 holds. Let X and Y be the parts of G, and let X' and Y' be the parts of G'. Choose vertices $x \in X$ and $x' \in X'$. Define the graph H with $V(H) = V(G) \cup V(G')$ and $E(H) = E(G) \cup E(G') \cup \{xx'\}$. Then the conjecture holds for H also.*

It may be remarked that Corollary 1 can be proved using Proposition 1 and induction as well. The following bound has been obtained in [6].

Theorem 5 *Let G be a bipartite graph on $n \geq 2$ vertices. Then*

$$\tau(G) \leq \frac{\prod_v d_v}{|E(G)|}, \tag{2.3}$$

with equality if and only if G is complete bipartite.

Since there can be at most $|X||Y|$ edges in a bipartite graph with parts X and Y, if Conjecture 2 were true, then Theorem 5 would follow. Thus, the assertion of Conjecture 2 improves upon Theorem 5 by a factor of $E(G)|/(|X||Y|)$. This motivates the following definition introduced in [18].

Definition 2 Let G be a bipartite graph with parts X and Y. The bipartite density of G, denoted $\rho(G)$, is the ratio $E(G)/(|X||Y|)$. Equivalently, G contains $\rho(G)$ times as many edges as the complete bipartite graph $K_{|X|,|Y|}$.

Let G be a graph with n vertices. Let A be the adjacency matrix of G and let D be the diagonal matrix of vertex degrees of G. Note that $L = D - A$ is the Laplacian of G. The matrix $K = D^{-\frac{1}{2}} L D^{-\frac{1}{2}}$ is termed as the normalized Laplacian of G. If G is connected, then K is positive semi-definite with rank $n - 1$. Let $\mu_1 \geq \mu_2 \cdots \geq \mu_{n-1} > \mu_n = 0$ denote the eigenvalues of K. It is known, see [11], that $\mu_{n-1} \leq 2$, with equality if and only if G is bipartite. Conjecture 2 can be shown to be equivalent to the following, see [18].

Conjecture 3 *Let G be a bipartite graph on $n \geq 3$ vertices with parts X and Y. Then*

$$\prod_{i=1}^{n-2} \mu_i \leq \rho(G).$$

Yet another result from [18] is the following.

Lemma 1 *Let G be a bipartite graph on $n \geq 3$ vertices with parts X and Y. Suppose, for some $1 \leq k \leq \lfloor \frac{n-1}{2} \rfloor$ we have*

$$\prod_{i=1}^{k} \mu_i(2 - \mu_i) \leq \rho(G).$$

Then Conjecture 2 holds for G.

We conclude this section by stating the following result [18]. It asserts that Conjecture 2 holds for a sufficiently edge-dense graph with a cut vertex of degree 2.

Theorem 6 *Let G be a bipartite graph. Suppose that $\rho(G) \geq 0.544$ and that G contains a cut vertex x of degree 2. Then Conjecture 2 holds for G.*

2.6 A Reformulation in Terms of Majorization

This section is based on [22]. Call a bipartite graph G *Ferrers-good* if $\tau(G) \leq F(G)$. Thus, Conjecture 2 may be expressed more briefly as the claim that all bipartite graphs are Ferrers-good.

In 2009, Jack Schmidt (as reported in [22]) computationally verified by an exhaustive search that all bipartite graphs on at most 13 vertices are Ferrers-good. For a bipartite graph, we refer to the vertices in the two parts as red vertices and blue vertices. In 2013, Praveen Venkataramana proved an inequality weaker than Conjecture 2 valid for all bipartite graphs:

Proposition 2 (Venkataramana) *Let G be a bipartite graph with red vertices having degrees $d_1, \ldots, d_p$ and blue vertices having degrees $e_1, \ldots, e_q$. Then*

$$\tau(G) \leq \prod_{i=1}^{p}(d_i + \frac{1}{2}) \prod_{j=1}^{q}(e_j + \frac{1}{2})\sqrt{e_1}.$$

Conjecture 2 can be expressed in terms of majorization, for which the standard reference is [19]. For a vector $a = (a_1, \ldots, a_n)$, the vector $(a[1], \ldots, a[n])$ denotes the rearrangement of the entries of a in non-increasing order. Recall that a vector $a = (a_1, \ldots, a_n)$ is majorized by another vector $b = (b_1, \ldots, b_n)$, written $a \prec b$, provided that the inequality

$$\sum_{i=1}^{k} a_{[i]} \leq \sum_{i=1}^{k} b_{[i]}$$

holds for $1 \leq k \leq n$ and holds with equality for $k = n$.

Given a finite sequence a, let $\ell(a)$ denote its number of parts and $|a|$ denote its sum. For example, if $a = (4, 3, 1)$, then $\ell(a) = 3$ and $|a| = 8$.

Definition 3 (*Conjugate sequence*) Let a be a partition of an integer. The conjugate partition of a is the partition a^*

$$a_i^* = \#\{j : 1 \le j \le \ell(a) \text{ and } a_j \ge i\}.$$

For example, $(5, 5, 4, 2, 2, 1)^* = (6, 5, 3, 3, 2)$.

Definition 4 (*Concatenation of sequences*) Let $a = (a_1, \ldots, a_p)$ and $b = (b_1, \ldots, b_q)$ be sequences. Then their concatenation is the sequence

$$a \oplus b = (a_1, \ldots, a_p, b_1, \ldots, b_q).$$

With this notation, we can now state the following conjecture.

Conjecture 4 *Let d be a partition with $\ell(d) = n$, and let λ be a non-increasing sequence of positive real numbers with $\ell(\lambda) = n - 1$. Suppose $d = a \oplus b$ for some a, b with $\ell(a) = p$ and $\ell(b) = q$. If $a \prec b^*$ and $d \prec \lambda \prec d^*$, then*

$$\frac{1}{n}\prod_{i=1}^{n-1}\lambda_i \le \frac{1}{pq}\prod_{i=1}^{n} d_i.$$

Conjecture 4 implies Conjecture 2 in view of the following two theorems.

Theorem 7 (Gale–Ryser) *Let a and b be partitions of an integer. There is a bipartite graph whose blue degree sequence is a and whose red degree sequence is b if and only if $a \prec b^*$.*

Theorem 8 (Grone–Merris conjecture, proved in [1]) *The Laplacian spectrum of a graph is majorized by the conjugate of its degree sequence.*

Now let us show that Conjecture 4 implies Conjecture 2. Assume Conjecture 4 is true. Let G be a bipartite graph on n vertices, with p blue vertices and q red vertices. Let d be its degree sequence, with blue degree sequence a and red degree sequence b, and let λ be its Laplacian spectrum. By Theorem 7, $a \prec b^*$. Since the Laplacian is a Hermitian matrix, $d \prec \lambda$, and by Theorem 8, $\lambda \prec d^*$. Hence, the assumptions of Conjecture 2 apply. We conclude that

$$\frac{1}{n}\prod_{i=1}^{n-1}\lambda_i \le \frac{1}{pq}\prod_{i=1}^{n} d_i. \tag{2.4}$$

By the Matrix-Tree Theorem, the left-hand side of (2.4) is $\tau(G)$. Hence, Conjecture 2 holds as well.

2.7 Resistance Distance in G and $G \setminus \{f\}$

We recall some definitions that will be useful. Given a matrix A of order $m \times n$, a matrix G of order $n \times m$ is called a generalized inverse (or a g-inverse) of A if it satisfies $AGA = A$. Furthermore, G is called Moore–Penrose inverse of A if it satisfies $AGA = A$, $GAG = G$, $(AG)' = AG$ and $(GA)' = GA$. It is well known that the Moore–Penrose inverse exists and is unique. We denote the Moore–Penrose inverse of A by A^+. We refer to [9] for background material on generalized inverses.

Let G be a connected graph with vertex set $V = \{1, \ldots, n\}$ and let $i, j \in V$. Let H be a g-inverse of the Laplacian matrix L of G. The resistance distance $r(i, j)$ between i and j is defined as

$$r_G(i, j) = h_{ii} + h_{jj} - h_{ij} - h_{ji}. \tag{2.5}$$

It can be shown that the resistance distance does not depend on the choice of the g-inverse. In particular, choosing the Moore–Penrose inverse, we see that

$$r_G(i, j) = \ell^+_{ii} + \ell^+_{jj} - 2\ell^+_{ij}.$$

Let G be a connected graph with $V(G) = \{1, \ldots, n\}$. We assume that each edge of G is given an orientation. If $e = \{i, j\}$ is an edge of G oriented from i to j, then the incidence vector x_e of e is and $n \times 1$ vector with $1(-1)$ at ith (jth) place and zeros elsewhere. The Laplacian L of G has rank $n - 1$ and any vector orthogonal to $\mathbf{1}$ is in the column space of L. In particular, x_e is in the column space of L.

For a matrix A, we denote by $A(i|j)$ the matrix obtained by deleting row i and column j from A. We denote $A(i|i)$ simply as $A(i)$. Similar notation applies to vectors. Thus for a vector x, we denote by $x(i)$ the vector obtained by deleting the ith coordinate of x. Let L be the Laplacian matrix of a connected graph G with vertex set $\{1, \ldots, n\}$. Fix $i, j \in \{1, \ldots, n\}$, $i \neq j$, and let H be the matrix constructed as follows. Set $H(i) = L(i)^{-1}$ and let the ith row and column of H be zero. Then H is a g-inverse of L ([2], p.133). It follows from (2.5) that $r(i, j) = h_{jj}$. For basic properties of resistance distance, we refer to [2, 3].

In the next result, we give several equivalent conditions under which deletion of an edge does not affect the resistance distance between the end vertices of another edge. This result, which appears to be of interest by itself, will be used in Sect. 2.8 to give another proof of Theorem 1. We denote an arbitrary g-inverse of the matrix L by L^-.

Theorem 9 *Let G be a graph with $V(G) = \{1, \ldots, n\}$, $n \geq 4$. Let $e = \{i, j\}$, $f = \{k, \ell\}$ be edges of G with no common vertex such that $G \setminus \{e\}$ and $G \setminus \{f\}$ are connected subgraphs. Let L, L_e and L_f be the Laplacians of G, $G \setminus \{e\}$ and $G \setminus \{f\}$, respectively. Let x_e, x_f be the incidence vector of e, f, respectively. Then the following statements are equivalent:*

(i) $r_G(i, j) = r_{G\setminus\{f\}}(i, j)$.
(ii) $r_G(k, \ell) = r_{G\setminus\{e\}}(k, \ell)$.
(iii) $\tau(G \setminus \{e\})\tau(G \setminus \{f\}) = \tau(G)\tau(G \setminus \{e, f\})$.
(iv) *The ith and the jth coordinates of L^+x_f are equal.*
(v) *The ith and the jth coordinates of L^-x_f are equal for any L^-.*
(vi) *The ith and the jth coordinates of $L_f^+x_f$ are equal.*
(vii) *The ith and the jth coordinates of $L_f^-x_f$ are equal for any L_f^-.*
(viii) *The kth and the ℓth coordinates of L^+x_e are equal.*
(ix) *The kth and the ℓth coordinates of L^-x_e are equal for any L^-.*
(x) *The kth and the ℓth coordinates of $L_e^+x_e$ are equal.*
(xi) *The kth and the ℓth coordinates of $L_e^-x_e$ are equal for any L_e^-.*

Proof Let $u = L_f^+x_f$, $w = L^+x_f$. Since x_f is in the column space of L_f, we have $x_f = L_fz$ for some z. It follows that $L_fu = L_fL_f^+x_f = L_fL_f^+L_fz = L_fz = x_f$. Similarly $Lw = x_f$. Since $L = L_f + x_fx_f'$ then $Lw = L_fw + x_fx_f'w$ and hence

$$L_f(u - w) = x_fx_f'w. \tag{2.6}$$

Also,

$$(x_f'w)L_fu = x_f(x_f'w). \tag{2.7}$$

Subtracting (2.7) from (2.6) gives $L_f(u - w - (x_f'w))u = 0$, which implies $u - w - (x_f'w)u = \alpha\mathbf{1}$ for some α. It follows that $(1 - x_f'w)u = w + \alpha\mathbf{1}$. If $1 - x_f'w = 0$, then all coordinates of w are equal, which would imply $Lw = 0$, contradicting $x_f = Lw$. Thus $1 - x_f'w \neq 0$ and hence $u = \frac{w+\alpha\mathbf{1}}{1-x_f'w}$. Thus, any two coordinates of u are equal if and only if the corresponding coordinates of w are equal. This implies the equivalence of (iv) and (vi). A similar argument shows that $(iv) - (vii)$ are equivalent and that $(viii) - (xi)$ are equivalent.

Note that $r_G(i, j) = \frac{\det L(i,j)}{\det L(i)} = \frac{\tau(G\setminus\{e\})}{\tau(G)}$, $r_{G\setminus\{f\}}(i, j) = \frac{\det L_f(i,j)}{\det L_f(i)} = \frac{\tau(G\setminus\{e,f\})}{\tau(G\setminus\{f\})}$. and $r_{G\setminus\{e\}}(k, \ell) = \frac{\det L_e(k,\ell)}{\det L_e(k)} = \frac{\tau(G\setminus\{e,f\})}{\tau(G\setminus\{e\})}$. Thus, (i), (ii) and (iii) are equivalent.

We turn to the proof of $(iv) \Rightarrow (i)$. Let $w = L^+x_f$ and suppose $w_i = w_j$. Since the vector $\mathbf{1}$ is in the null space of L^+, we may assume, without loss of generality, that $w_i = w_j = 0$. As seen before, $Lw = x_f$.

Since $L(i) = L_f(i) + x_f(i)x_f(i)'$, by the Sherman–Morrison formula,

$$\begin{aligned} L(i)^{-1} &= (L_f(i) + x_f(i)x_f(i)')^{-1} \\ &= L_f(i)^{-1} - \frac{L_f(i)^{-1}x_f(i)x_f(i)'L_f(i)^{-1}}{1 - x_f(i)'L_f(i)^{-1}x_f(i)}. \end{aligned} \tag{2.8}$$

Since $x_f = Lw$, $w_i = 0$ and $(x_f(i))_j = 0$, we have

$$\begin{aligned}(x_f(i))_j &= (L(i)w(i))_j \\ &= ((L_f(i) + x_f(i)x_f(i)')w(i))_j \\ &= (L_f(i)w(i)_j + x_f(i)'w(i)(L_f(i)x_f(i))_j.\end{aligned}$$

Hence $(L_f(i)^{-1}x_f(i))_j = 0$. It follows from (2.8) that the (j, j)th element of $L(i)^{-1}$ and $L_f(i)^{-1}$ are identical. In view of the observation preceding the Theorem, the (j, j)-element of $L(i)^{-1}$ (respectively, $L_f(i)^{-1}$) is the resistance distance between i and j in G (respectively, $G \setminus \{f\}$). Therefore the resistance distance between i and j is the same in G and $G \setminus \{f\}$ if the ith and the jth coordinates of L^+x are equal.

Before proceeding we remark that if (v) holds for a particular g-inverse, then it can be shown that it holds for any g-inverse. Similar remark applies to (vii), (ix) and (x).

Now suppose (i) holds. Then $(L(i))^{-1}_{jj} = (L_f(i))^{-1}_{jj}$, and using (2.8) we conclude that $(L_f(i)^{-1}x_f(i)x_f(i)'L_f(i))_{jj} = 0$, which implies

$$(L_f(i)^{-1}x_f(i))_j = 0. \tag{2.9}$$

If we augment $L_f(i)^{-1}$ by introducing the ith row and ith column, both equal to zero vectors, then we obtain a g-inverse L_f^- of L_f. Since the ith coordinate of x_f is zero, we conclude from (2.9) that $(L_f^- x_f)_j = 0$. Since the ith row of L_f^- is zero, $(L_f^- x_f)_i = 0$. It follows that the ith and the jth coordinates of $L_f^- x_f = 0$ and thus (vii) holds (for a particular g-inverse and hence for any g-inverse). Similarly, it can be shown that $(ii) \Rightarrow (xi)$. This completes the proof. ■

2.8 The Number of Spanning Trees in Ferrers Graphs

We now prove a preliminary result.

Lemma 2 *Consider the Ferrers graph G with bipartition (U, V), where $U = \{u_1, \ldots, u_m\}$, $V = \{v_1, \ldots, v_n\}$. Let λ_i be the degree of u_i, $i = 1, \ldots, m$ and let λ'_j be the degree of v_j, $j = 1, \ldots, n$. Let $p \in \{1, \ldots, m-1\}$ be such that $\lambda_i = n$, $i = 1, \ldots, p$ and $\lambda_{p+1} = k < n$. Let f be the edge $\{u_p, v_n\}$. Then*

$$r_G(u_{p+1}, v_k) = r_{G\setminus\{f\}}(u_{p+1}, v_k). \tag{2.10}$$

Proof The bipartite adjacency matrix of G is given by

$$M = \begin{array}{c} \\ 1 \\ 2 \\ \vdots \\ p \\ p+1 \\ \vdots \\ m \end{array} \begin{array}{c} \begin{array}{ccccc} 1 & 2 & \cdots & \cdots & n \end{array} \\ \begin{pmatrix} 1 & 1 & \cdots & \cdots & 1 \\ 1 & 1 & \cdots & \cdots & 1 \\ 1 & 1 & \cdots & \cdots & 1 \\ 1 & 1 & \cdots & \cdots & 1 \\ 1 & 1 & \cdots & 0 & 0 \\ 1 & 1 & \cdots & 0 & 0 \\ 1 & 1 & \cdots & \cdots & 0 \end{pmatrix} \end{array},$$

and the Laplacian matrix L of G is given by

$$L = diag(\lambda_1, \ldots, \lambda_m, \lambda'_1, \ldots, \lambda'_n) - \begin{bmatrix} 0 & M \\ M' & 0 \end{bmatrix}.$$

Let

$$w = \frac{1}{p}[\underbrace{-\frac{1}{n}, \cdots, -\frac{1}{n}}_{p-1}, \frac{p-1}{n}, 0, \cdots, 0, -1]'.$$

It can be verified that Lw is the $(m+n) \times 1$ vector with 1 at position p, -1 at position $m+n$ and zeros elsewhere. Thus, $Lw = x_f$, the incidence vector of the edge $f = \{u_p, v_n\}$.

It follows from basic properties of the Moore–Penrose inverse [9] that

$$L^+L = \left(I - \frac{1}{m+n}\mathbf{1}\mathbf{1}'\right).$$

Hence

$$L^+x_f = L^+Lw = \left(I - \frac{1}{m+n}\mathbf{1}\mathbf{1}'\right)w = w - \alpha\mathbf{1}\mathbf{1}', \tag{2.11}$$

where $\alpha = \mathbf{1}'w/(m+n)$. Let e be the edge $\{u_{p+1}, v_k\}$. Since the coordinates $p+1$ and $m+k$ of w are zero, it follows from (2.11) and the implication $(iv) \Rightarrow (i)$ of Theorem 9 that (2.10) holds. This completes the proof. ■

Let G be a connected graph with $V(G) = \{1, \ldots, n\}$, and let $i, j \in V(G)$. Let L be the Laplacian of G. We denote by $L(i, j)$ the submatrix of L obtained by deleting rows i, j and columns i, j. Recall that $\tau(G)$ denotes the number of spanning trees of G. It is well known that

$$r_G(i, j) = \frac{\det L(i|j)}{\tau(G)}. \tag{2.12}$$

Furthermore, $\det L(i, j)$ is the number of spanning forests of G with two components, one containing i and the other containing j. Now suppose that i and j are adjacent and let $f = \{i, j\}$ be the corresponding edge. Let $\tau'(G)$ and $\tau''(G)$ denote the number

of spanning trees of G, containing f, and not containing f, respectively. Then in view of the preceding remarks, $\tau'(G) = \det L_1(i, j)$, where L_1 is the Laplacian of $G \setminus \{e\}$.

Theorem 10 ([13]) *Let G be the Ferrers graph with the bipartition (U, V), where $U = \{u_1, \ldots, u_m\}$, $V = \{v_1, \ldots, v_n\}$ and let $\lambda = (\lambda_1, \ldots, \lambda_m)$, $\lambda' = (\lambda'_1, \ldots, \lambda'_n)$ be the associated partitions. Then the number of spanning trees in G is*

$$\frac{1}{mn} \prod_{i=1}^{m} \lambda_i \prod_{i=1}^{n} \lambda'_i.$$

Proof We assume λ_m, λ'_n to be positive, for otherwise, the graph is disconnected and the result is trivial. We prove the result by induction on the number of edges. Let $e = \{p + 1, m + k\}$, $f = \{p, m + n\}$ be edges of G.

By the induction assumption, we have

$$\tau(G \setminus \{e\}) = \frac{1}{mn} \prod_{i=1}^{m} \lambda_i \prod_{i=1}^{n} \lambda'_i \frac{(\lambda_{p+1} - 1)(\lambda'_k - 1)}{\lambda_{p+1}\lambda'_k}, \tag{2.13}$$

$$\tau(G \setminus \{f\}) = \frac{1}{mn} \prod_{i=1}^{m} \lambda_i \prod_{i=1}^{n} \lambda'_i \frac{(\lambda_p - 1)(\lambda'_n - 1)}{\lambda_p\lambda'_n}, \tag{2.14}$$

and

$$\tau(G \setminus \{e, f\}) = \frac{1}{mn} \prod_{i=1}^{m} \lambda_i \prod_{i=1}^{n} \lambda'_i \frac{(\lambda_{p+1} - 1)(\lambda_p)(\lambda'_k - 1)(\lambda'_n - 1)}{\lambda_{p+1}\lambda_p\lambda'_k\lambda'_n}. \tag{2.15}$$

It follows from (2.13), (2.14), (2.15) and Theorem 9 that

$$\tau(G) = \frac{\tau(G \setminus \{e\})(\tau(G \setminus \{f\})}{\tau(G \setminus \{e, f\})} = \frac{1}{mn} \prod_{i=1}^{m} \lambda_i \prod_{i=1}^{n} \lambda'_i,$$

and the proof is complete. ■

Acknowledgements I sincerely thank Ranveer Singh for a careful reading of the manuscript. Support from the JC Bose Fellowship, Department of Science and Technology, Government of India, is gratefully acknowledged.

References

1. Bai, H.: The Grone-Merris conjecture. Trans. Am. Math. Soc. **363**, 4463–4474 (2011)
2. Bapat, R.B.: Graphs and Matrices, 2nd edn. Hindustan Book Agency, New Delhi and Springer (2014)
3. Bapat, R.B.: Resistance distance in graphs. Math. Stud. **68**, 87–98 (1999)
4. Bhattacharya, A., Friedland, S., Peled, U.N.: On the first eigenvalue of bipartite graphs. Electron. J. Comb. **15**, # R144 (2008)
5. Bondy, J.A., Murty, U.S.R.: Graph Theory, Graduate Texts in Mathematics, vol. 244. Springer, New York (2008)
6. Bozkurt, S.B.: Upper bounds for the number of spanning trees of graphs. J. Inequal. Appl. **269** (2012)
7. Brualdi, R.A., Hoffman, A.J.: On the spectral radius of (0, 1)-matrices. Linear Algebra Appl. **65**, 133146 (1985)
8. Burns, J.: Bijective Proofs for Enumerative Properties of Ferrers Graphs. arXiv: math/0312282v1 [math CO] (2003)
9. Campbell, S.L., Meyer, C.D.: Generalized Inverses of Linear Transformations. Pitman, London (1979)
10. Chestnut, S.R., Fishkind, D.E.: Counting spanning trees in threshold graphs. arXiv: 1208.4125v2 (2013)
11. Chung, F.R.K.: Spectral Graph Theory. CBMS Regional Conference Series in Mathematics. American Mathematical Society, Providence (1997)
12. Cvetković, D.M., Doob, M., Sachs, H.: Spectra of Graphs. Theory and Applications, 3rd edn. Johann Ambrosius Barth, Heidelberg (1995)
13. Ehrenborg, R., Willigenburg, S.V.: Enumerative properties of Ferrers graphs. Discrete Comput. Geom. **32**, 481–492 (2004)
14. Friedland, S.: Bounds on the spectral radius of graphs with e edges. Linear Algebra Appl. **101**, 8186 (1988)
15. Garrett, F., Klee, S.: Upper bounds for the number of spanning trees in a bipartite graph. Preprint. http://fac-staff.seattleu.edu/klees/web/bipartite.pdf (2014)
16. Hammer, P.L., Peled, U.N., Sun, X.: Difference graphs. Discrete Appl. Math. **28**, 35–44 (1990)
17. Klein, D.J., Randić, M.: Resistance distance. J. Math. Chem. **12**, 81–95 (1993)
18. Koo, C.W.: A bound on the number of spanning trees in bipartite graphs. Senior thesis. https://www.math.hmc.edu/~ckoo/thesis/ (2016)
19. Marshall, A.W., Olkin, I., Arnold, B.C.: Inequalities: Theory of Majorization and Its Applications. Springer, New York (2011)
20. Petrović, M., Simić, S.K.: A note on connected bipartite graphs of fixed order and size with maximal index. Linear Algebra Appl. **483**, 21–29 (2015)
21. Rowlinson, P.: On the maximal index of graphs with a prescribed number of edges. Linear Algebra Appl. **110**, 43–53 (1988)
22. Slone, M.: A conjectured bound on the spanning tree number of bipartite graphs. arXiv:1608.01929v2 [math.CO] (2016)
23. Stanley, R.P.: A bound on the spectral radius of graphs with e edges. Linear Algebra Appl. **87**, 267–269 (1987)

Chapter 3
Optimization Problems on Acyclic Orientations of Graphs, Shellability of Simplicial Complexes, and Acyclic Partitions

Masahiro Hachimori

3.1 An Optimization Problem on Acyclic Orientation of Graphs in the Theory of Polytopes

For an undirected graph $G = (V(G), E(G))$ and its orientation O, we denote by G^O the resulted directed graph. In this chapter, we consider optimization problems such that the values of the objective functions are determined by the out-degrees of G^O, where we vary the orientations O of G under some given restrictions. A typical example is the following problem.

$$(P1): \quad \min \sum_{v \in G} 2^{\text{out-deg}(v;\, G^O)} \quad \text{s. t. } O \text{ is acyclic,}$$

where the minimum is taken by varying the orientations O of G under the restriction that O is acyclic, i.e., there are no directed cycles on G^O. Here, out-deg$(v;\, G^O)$ is the out-degree of v in G^O. This optimization problem appears in the theory of polytopes. In [6], Blind and Mani showed the following theorem.

Theorem 1 (Blind and Mani [6]) *The combinatorial structure of a simple polytope P is determined by its graph $G(P)$.*

Here, the graph $G(P)$ of a polytope P is a graph consisting of the vertices and edges of P. In other words, two simple polytopes have isomorphic face lattices if and only if their graphs are isomorphic.

Later, Kalai [14] gave a simple short proof for Theorem 1. In his proof, the key is the notion of "good orientation." An orientation O of $G(P)$ is a *good orientation*

M. Hachimori (✉)
Faculty of Engineering, Information and Systems, University of Tsukuba,
Tsukuba, Ibaraki 305-8573, Japan
e-mail: hachi@sk.tsukuba.ac.jp

S. K. Neogy et al. (eds.), *Mathematical Programming and Game Theory*,
Indian Statistical Institute Series, https://doi.org/10.1007/978-981-13-3059-9_3

if the restriction of $G(P)^O$ to every face of P (including P itself) has exactly one source. (Remark: In this chapter, we orient all the edges in a reverse way to the original paper. Originally, it is defined that an orientation is good if all the restriction of $G(P)^O$ to every face of P has exactly one sink. Here, a *source* is a node in a directed graph such that all the edges incident to the node are oriented from the node, and a *sink* is a node such that all the edges incident to it are oriented into the node.) Using this definition, it is shown that a set of vertices A of $G(P)$ forms a face of P if and only if the induced subgraph $G(P)[A]$ is k-regular and A is an ending set with respect to some good orientation O (i.e., all the edges connecting a vertex a of A and a vertex a' outside of A are oriented from a' to a.). By this fact, the remaining thing to be shown is to determine which orientations are good without knowing which set A of vertices forms a face of P. The following theorem is the answer to this.

Theorem 2 (Kalai [14]) *For a simple polytope P, an orientation O of $G(P)$ is a good orientation if and only if it is a minimizer of the problem (P1) with $G = G(P)$.*

Since Theorem 2 assures that whether an orientation is good or not is determined only by $G(P)$ (no information of the faces of P is needed), this gives the proof of Theorem 1. A comprehensive introduction of this story can be found in Ziegler [21, Lect. 3.4].

In this chapter, we introduce optimization problems similar to (P1) in the following sections in relation to the combinatorial structures of simplicial complexes and cubical complexes.

3.2 Shellability of Simplicial Complexes and Orientations of Facet-Ridge Incidence Graphs

A (finite) *simplicial complex* Γ is a nonempty set of simplices in some Euclidean space $\mathbb{R}^N$ such that (i) every face of $\sigma \in \Gamma$ is a member of Γ, and (ii) $\sigma \cap \tau$ is a face of both σ and τ for any $\sigma, \tau \in \Gamma$. (Remark: we treat the empty set as a (-1)-dimensional simplex, and in this definition, the empty set is always a member of a simplicial complex. Also we remark that we assume all the simplicial complexes are finite in this chapter.) The members of a simplicial complex Γ are *faces* of Γ. We adopt the conventional terminology to mention 0-dimensional faces as *vertices*, 1-dimensional faces as *edges*, and the maximal faces with respect to inclusion as *facets*. The dimension of a simplicial complex Γ is the maximum dimension of its faces. A simplicial complex is *pure* if all the facets are of the same dimension.

The combinatorial structures of simplicial complexes have been important subjects of study from several contexts, as a high-dimensional generalization of graphs, in the theory of polytopes (e.g., Ziegler [21, Lect. 8]), as a tool of topological methods in combinatorics (Björner [2]), or a way to address applications like computing network reliability (Colbourn [7]). One reason simplicial complexes appear in many contexts in combinatorics is because they are equivalent to a set family closed under

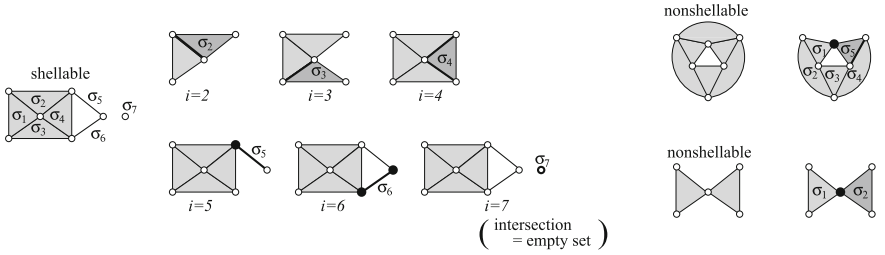


Fig. 3.1 Shellable and nonshellable simplicial complexes

taking subsets (i.e., "*abstract simplicial complex*") which can be found quite commonly in many combinatorial structures. Among several combinatorial properties of simplicial complexes, shellability is one of the most famous and important properties, and it appears in many places.

Definition 1 A simplicial complex Γ is *shellable* if the facets $\sigma_1, \sigma_2, \ldots, \sigma_t$ of Γ can be ordered such that $(\bigcup_{j=1}^{i-1} \overline{\sigma}_j) \cap \overline{\sigma_i}$ is a $(\dim \sigma_i - 1)$-dimensional pure subcomplex for each $2 \leq i \leq t$, where $\overline{\sigma}$ denotes the simplicial complex consisting of all the faces of σ. An ordering of facets satisfying this condition is called a *shelling*.

See Fig. 3.1 for examples of shellable and nonshellable simplicial complexes. During the previous century, shellability of simplicial complexes are only defined for pure simplicial complexes (e.g., [2, 21]). To define shellability for nonpure simplicial complexes is suggested by Björner and Wachs [4, 5] and now this generalized version has become the standard definition. Our definition above follows this version.

To distinguish shellable simplicial complexes and nonshellable ones is a difficult problem. All zero-dimensional simplicial complexes are shellable, and one-dimensional simplicial complexes are shellable if and only if its 1-dimensional edges are connected (i.e., a connected graph with some isolated vertices). However, for two- and higher dimensional simplicial complexes, no efficient way is known in general to recognize whether a given simplicial complex is shellable or not. The recognition problem is in the class NP, but it is neither known whether it is in P or not, nor whether it is NP-complete or not (Kaibel and Pfetsch [16, Sec. 34]). There is an efficient way to recognize shellability for the two-dimensional case if restricted to the class of pseudomanifolds (Danaraj and Klee [8]), but it is not known whether there exist efficient algorithms to recognize shellability for three-dimensional pseudomanifolds, even for the triangulations of spheres.

In this section, we give a characterization of shellability by an optimization problem on orientations of graphs. First, we restrict ourselves to the pure case for simplicity. Later, we give a generalized formulation including nonpure complexes. Our result in this section first appeared in Hachimori and Moriyama [13], and also appeared in Hachimori [11] with a generalized treatment. We here follow the proof given in Hachimori [11] and present in a somewhat more easily comprehensible way.

3.2.1 The Case of Pure Simplicial Complexes

Though our result of this section is valid for general simplicial complexes including both pure and nonpure simplicial complexes, we first present the result restricted to pure simplicial complexes in this subsection, since the pure case is essential in this result. The generalization to include the nonpure case, which will be presented in the next subsection, is just a technical revision and easy to follow after understanding the pure case.

For a pure d-dimensional simplicial complex Γ, we say that a face τ is a *ridge* of Γ if it is *covered* by a facet, i.e., if $\tau \subseteq \sigma$ with $\dim \tau = \dim \sigma - 1$ for some facet σ. Let $\mathcal{F}(\Gamma)$ be the set of facets, and $\mathcal{R}(\Gamma)$ the set of ridges of Γ. Since Γ is pure, $\mathcal{F}(\Gamma)$ is exactly the set of d-dimensional faces and $\mathcal{R}(\Gamma)$ is the set of $(d-1)$-dimensional faces of Γ. (Remark that we need to change the definition of ridges for nonpure complexes in the next subsection.) We let the graph $G(\Gamma)$ be the *facet-ridge incidence graph*, i.e., the bipartite graph with the partite sets $\mathcal{F}(\Gamma)$ and $\mathcal{R}(\Gamma)$, and two nodes $\sigma \in \mathcal{F}(\Gamma)$ and $\tau \in \mathcal{R}(\Gamma)$ are adjacent if and only if $\sigma \supseteq \tau$ in Γ.

We consider orientations of the graph $G(\Gamma)$. We denote the oriented arc from α to β in $G(\Gamma)$ by $\alpha \to \beta$, and denote the directed path from α to β by $\alpha \rightsquigarrow \beta$. We say an orientation O is *admissible* if in-deg$(\tau) \geq 1$ for every $\tau \in \mathcal{R}(\Gamma)$. We have the following characterization of shellability of pure simplicial complexes.

Theorem 3 *For a pure d-dimensional simplicial complex Γ, let us consider the following minimization problem:*

$$(P2): \quad \min \sum_{\sigma \in \mathcal{F}(\Gamma)} 2^{\text{out-deg}(\sigma;\, G^O(\Gamma))} \quad \text{s. t. } O \text{ is acyclic and admissible.}$$

Then the optimum value V^* of (P2) satisfies $V^* \geq f(\Gamma)$, where $f(\Gamma)$ is the number of all the faces of Γ. Further, the equality holds if and only if Γ is shellable.

The proof of Theorem 3 follows the following lemmas.

First, we define the set $S^O(\sigma)$ as follows.

$$S^O(\sigma) = \{\eta \in \Gamma : \sigma \to \tau \text{ in } G^O(\Gamma) \text{ for every ridge } \tau \text{ with } \eta \subseteq \tau \subseteq \sigma\} \cup \{\sigma\}. \tag{3.1}$$

Note that the complement of $S^O(\sigma)$ in $\overline{\sigma}$, denoted as $S^{cO}(\sigma)$, is given as follows.

$$\begin{aligned} S^{cO}(\sigma) &= \overline{\sigma} - S^O(\sigma) \\ &= \{\eta \in \Gamma : \sigma \leftarrow \tau \text{ in } G^O(\Gamma) \text{ for some ridge } \tau \text{ with } \eta \subseteq \tau \subseteq \sigma\} \\ &= \bigcup \{\overline{\tau} : \tau \in \mathcal{R}(\Gamma),\ \sigma \leftarrow \tau \text{ in } G^O(\Gamma)\}. \end{aligned} \tag{3.2}$$

Lemma 1 *Let Γ be a pure simplicial complex, and let $\eta \in \Gamma$ and $\sigma \in \mathcal{F}(\Gamma)$. Then, for any orientation O, $\eta \in S^O(\sigma)$ if and only if σ is a source node in $G^O_{\supseteq\eta}(\Gamma)$, where $G^O_{\supseteq\eta}(\Gamma)$ is the subgraph induced by the nodes corresponding to the facets and the ridges of Γ containing η.*

Proof The proof is obvious from the definition of $S^O(\sigma)$. □

The inequality of the theorem follows the following lemma.

Lemma 2 *Let Γ be a pure simplicial complex and O an orientation of $G(\Gamma)$ that is acyclic and admissible. Then, we have $\sum_{\sigma\in\mathcal{F}(\Gamma)} 2^{\text{out-deg}(\sigma;\, G^O(\Gamma))} \geq f(\Gamma)$.*

Proof We have the graph $G^O_{\supseteq\eta}(\Gamma)$ acyclic since $G^O(\Gamma)$ is acyclic, and this implies that $G^O_{\supseteq\eta}(\Gamma)$ has at least one source node. This source node should be a facet, not a ridge, by the condition that O is admissible. By Lemma 1, this implies that the family $\{S^O(\sigma) : \sigma \in \mathcal{F}(\Gamma)\}$ covers Γ (i.e., for any $\eta \in \Gamma$ there exists a $\sigma \in \mathcal{F}(\Gamma)$ such that $\eta \in S^O(\sigma)$). On the other hand, we have $|S^O(\sigma)| = 2^{\text{out-deg}(\sigma;\, G^O(\Gamma))}$. (This follows from the fact that $S^O(\sigma)$ forms a boolean lattice with respect to inclusion relation. In fact, the smallest face in $S^O(\sigma)$ is given by $\sigma \cap \{\tau \in \mathcal{R}(\Gamma) : \sigma \to \tau \text{ in } G^O(\Gamma)\} =: \Psi^O(\sigma)$ and $S^O(\sigma)$ equals the interval $[\Psi^O(\sigma), \sigma]$ in the face poset of Γ. This interval is a boolean lattice since every proper interval in the face poset of a simplicial complex is boolean.) Hence the inequality is verified. □

By the proof of Lemma 2, we have the following natural consequence for the equality case.

Lemma 3 *Let Γ be a pure simplicial complex and O an orientation of $G(\Gamma)$ that is acyclic and admissible. The equality $\sum_{\sigma\in\mathcal{F}(\Gamma)} 2^{\text{out-deg}(\sigma;\, G^O(\Gamma))} = f(\Gamma)$ holds if and only if $\{S^O(\sigma) : \sigma \in \mathcal{F}(\Gamma)\}$ forms a partition of Γ.*

Proof By the proof of Lemma 2, $\{S^O(\sigma) : \sigma \in \mathcal{F}(\Gamma)\}$ covers Γ. Since $\sum_{\sigma\in\mathcal{F}(\Gamma)} 2^{\text{out-deg}(\sigma;\, G^O(\Gamma))}$ counts the number of the faces of Γ with multiplicity in this covering, the equality means that each face of Γ is contained in exactly one $S^O(\sigma)$. □

Here we define a graph $\widetilde{G}^O(\Gamma)$ whose nodes are the facets of Γ and arcs $\sigma \to \sigma'$ are defined if there is a face $\eta \subseteq \sigma'$ with $\eta \in S^O(\sigma)$. We have the following lemma.

Lemma 4 *When $\{S^O(\sigma) : \sigma \in \mathcal{F}(\Gamma)\}$ is a partition of a pure simplicial complex Γ, $\widetilde{G}^O(\Gamma)$ is acyclic if and only if $G^O(\Gamma)$ is acyclic.*

Proof Let us assume $\widetilde{G}^O(\Gamma)$ has a directed cycle. Assume $\sigma \to \sigma'$ is an arc in $\widetilde{G}^O(\Gamma)$. From the definition of $\widetilde{G}^O(\Gamma)$, there exists a face η with $\eta \subseteq \sigma'$ and $\eta \in S^O(\sigma)$. Here, $\eta \subseteq \sigma'$ implies that both σ and σ' are nodes of $G^O_{\supseteq\eta}(\Gamma)$. From Lemma 1 and the assumption that $\{S^O(\sigma) : \sigma \in \mathcal{F}(\Gamma)\}$ is a partition, $\eta \in S^O(\sigma)$ implies that σ is a unique source in $G^O_{\supseteq\eta}(\Gamma)$. This assures the existence of a directed path from σ to σ' in $G^O_{\supseteq\eta}(\Gamma)$, and thus in $G^O(\Gamma)$. Hence, the existence of a directed cycle in $\widetilde{G}^O(\Gamma)$ implies the existence of a directed cycle in $G^O(\Gamma)$.

On the other hand, let us assume $G^O(\Gamma)$ has a directed cycle. The cycle is of the form $\sigma_1 \to \tau_1 \to \sigma_2 \to \tau_2 \to \cdots \to \sigma_s \to \tau_s \to \sigma_{s+1} = \sigma_1$, where $\sigma_i \in \mathcal{F}(\Gamma)$ for all $1 \leq i \leq s$ and $\tau_j \in \mathcal{R}(\Gamma)$ for all $1 \leq j \leq s$. Then we have $\tau_i \subseteq \sigma_i$ and $\tau_i \in S^O(\sigma_{i+1})$ for all $1 \leq i \leq s$, and this implies there is a directed cycle in $\widetilde{G}^O(\Gamma)$. □

The following last lemma shows that having an acyclic admissible orientation O of $G(\Gamma)$ such that the family $\{S^O(\sigma) : \sigma \in \mathcal{F}(\Gamma)\}$ is a partition of Γ and $\widetilde{G}^O(\Gamma)$ is acyclic is equivalent to the shellability of Γ.

Lemma 5 *For a pure simplicial complex Γ, there exists an acyclic admissible orientation O of $G(\Gamma)$ such that $\{S^O(\sigma) : \sigma \in \mathcal{F}(\Gamma)\}$ is a partition of Γ with $\widetilde{G}^O(\Gamma)$ acyclic if and only if Γ is shellable.*

Proof To show the "only if" part, let us assume $\{S^O(\sigma) : \sigma \in \mathcal{F}(\Gamma)\}$ is a partition of Γ and $\widetilde{G}^O(\Gamma)$ is acyclic. Let $\sigma_1, \sigma_2, \ldots, \sigma_t$ be a linear extension (or a "topological sort") of $\widetilde{G}^O(\Gamma)$, i.e., a total ordering such that the existence of a directed arc $\sigma_i \to \sigma_j$ in $\widetilde{G}^O(\Gamma)$ implies $i < j$. From the fact that $\{S^O(\sigma) : \sigma \in \mathcal{F}(\Gamma)\}$ is a partition together with the Eqs. (3.1) and (3.2), we have that $(\bigcup_{j=1}^{i-1} \overline{\sigma_j}) \cap \overline{\sigma_i} = S^{cO}(\sigma_i) = \bigcup\{\overline{\tau} : \tau \in \mathcal{R}(\Gamma),\ \sigma_i \leftarrow \tau \text{ in } G^O(\Gamma)\}$ for every $1 \leq i \leq t-1$. Hence, $\sigma_1, \sigma_2, \ldots, \sigma_t$ is a shelling and Γ is shellable since $(\bigcup_{j=1}^{i-1} \overline{\sigma_j}) \cap \overline{\sigma_i}$ is $(\dim \sigma_i - 1)$-dimensional and pure. (Note that, the set $\{\tau \in \mathcal{R}(\Gamma) : \sigma_i \leftarrow \tau \text{ in } G^O(\Gamma)\}$ is not empty for $i > 1$ since $G(\Gamma) = G(\Gamma)_{\geq \emptyset}$ has only one source node and it should be σ_1.) For the "if" part, let Γ be a pure shellable simplicial complex and $\sigma_1, \sigma_2, \ldots, \sigma_t$ be its shelling. It is well known that this shelling induces a partition of Γ by $\bigcup_{i=1}^{t} [Res(\sigma_i), \sigma_i]$ with $Res(\sigma_i)$ the minimum face of σ_i not contained in the facets $\sigma_1, \sigma_2, \ldots, \sigma_{i-1}$, see for example [4, Sec. 2] or [21, Lect. 8]. (Here, $[a, b] = \{z \in \Gamma : a \subseteq z \subseteq b\}$. Remark that $\Psi^O(\sigma)$ mentioned in the proof of Lemma 2 coincides with this $Res(\sigma)$.) This $Res(\sigma_i)$ is called the "restriction" of σ_i, and given by $Res(\sigma_i) = \bigcap\{\tau \in \mathcal{R}(\Gamma) \cap \overline{\sigma_i} : \text{there is no } j < i \text{ with } \tau \subseteq \sigma_j\}$. We construct an orientation O such that, for each ridge τ incident to σ_i, $\tau \to \sigma_i$ if $\tau \subseteq \sigma_j$ for some $j < i$, and $\tau \leftarrow \sigma_i$ otherwise. Under this orientation, we have $Res(\sigma_i) = \bigcap\{\tau \in \mathcal{R}(\Gamma) : \tau \leftarrow \sigma_i\}$, and thus $[Res(\sigma_i), \sigma_i] = \{\eta \in \Gamma : \tau \to \sigma_i \text{ for all } \tau \in \mathcal{R}(\Gamma) \text{ with } \eta \subseteq \tau \subseteq \sigma_i\} = S^O(\sigma_i)$. Hence $\{S^O(\sigma_i) : 1 \leq i \leq t\}$ forms a partition of Γ. Here, O is obviously acyclic, and thus we have $\widetilde{G}^O(\Gamma)$ acyclic by Lemma 4, hence Lemma 5 is verified. □

Proof (*Proof of Theorem 3*) The inequality $V^* \geq f(\Gamma)$ follows from Lemma 2. Further, Lemma 3 shows that the equality holds if and only if $\{S^O(\sigma) : \sigma \in \mathcal{F}(\Gamma)\}$ is a partition of Γ, and for this partition we have $\widetilde{G}(\Gamma)$ acyclic by Lemma 4. Finally, Lemma 5 shows this is equivalent to the shellability of Γ. □

3.2.2 The Case of Nonpure Simplicial Complexes

In the case of pure simplicial complexes, we defined the faces covered by a facet as ridges and considered the adjacency between facets and ridges. For the case of general

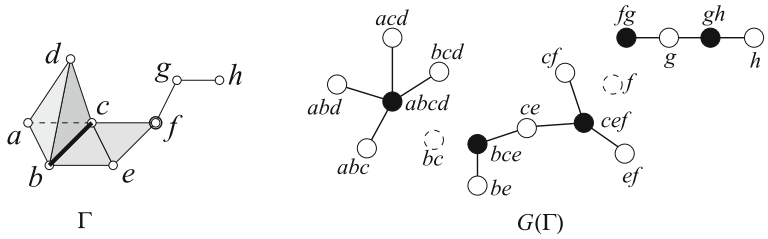


Fig. 3.2 The simplicial complex Γ has facets $abcd$, bce, cef, fg, and gh. The faces bc and f are pseudoridges. In the figure of $G(\Gamma)$, the black nodes are facets and white nodes are ridges. The pseudoridges are indicated by the node with dashed circle but they are not contained in $G(\Gamma)$

simplicial complexes including nonpure complexes, we need to discriminate these faces covered by a facet into ridges and pseudoridges. Let Γ be a simplicial complex not necessarily pure. Let τ be a face covered by some facet. We say τ is a *ridge* if all its superfaces (i.e., faces strictly containing τ) are facets, and a *pseudoridge* otherwise. We denote the set of facets, ridges, and pseudoridges of Γ, by $\mathcal{F}(\Gamma)$, $\mathcal{R}(\Gamma)$, and $\mathcal{R}'(\Gamma)$, respectively.

We define the facet-ridge incidence graph $G(\Gamma)$ as the bipartite graph with partite sets $\mathcal{F}(\Gamma)$ and $\mathcal{R}(\Gamma)$, where the two nodes $\sigma \in \mathcal{F}(\Gamma)$ and $\tau \in \mathcal{R}(\Gamma)$ are joined by an edge if $\sigma \supseteq \tau$. Note that we do not include pseudoridges in $G(\Gamma)$. (See Fig. 3.2 for example. Here, note that the adjacency between a facet σ and a ridge τ occurs in $G(\Gamma)$ only when $\dim \sigma = \dim \tau + 1$.)

Under this setting, we have the same statement as the pure case.

Theorem 4 *For a d-dimensional (not necessarily pure) simplicial complex Γ, let us consider the following minimization problem:*

$$(P3): \quad \min \sum_{\sigma \in \mathcal{F}(\Gamma)} 2^{\text{out-deg}(\sigma;\, G^O(\Gamma))}$$
$$\text{s. t.} \quad O \text{ is acyclic and admissible.}$$

Then, the optimum value V^* of (P3) satisfies $V^* \geq f(\Gamma)$, where $f(\Gamma)$ is the number of all the faces of Γ. Further, the equality holds if and only if Γ is shellable.

Note that Theorem 4 contains Theorem 3 as a special case.

For the proof of Theorem 4, we introduce a graph $G'(\Gamma)$ and $G'^O(\Gamma)$ as follows. The graph $G'(\Gamma)$ is the graph obtained from $G(\Gamma)$ by adding pseudoridges as nodes and edges between pseudoridges and facets such that an edge is introduced between $\tau \in \mathcal{R}'(\Gamma)$ and $\sigma \in \mathcal{F}(\Gamma)$ if $\tau \subseteq \sigma$. (Here, σ and τ with $\dim \sigma > \dim \tau + 1$ can be joined by an edge.) For an orientation O of $G(\Gamma)$, we extend the orientation to that of $G'(\Gamma)$ to obtain $G'^O(\Gamma)$. In this extended orientation, for $\tau \in \mathcal{R}'(\Gamma)$ and $\sigma \in \mathcal{F}(\Gamma)$ with $\tau \subseteq \sigma$, we orient $\tau \to \sigma$ if $\dim \sigma = \dim \tau + 1$ and $\tau \leftarrow \sigma$ if $\dim \sigma > \dim \tau + 1$. (See Fig. 3.3 for example.) Further, for a face $\eta \in \Gamma$, we let

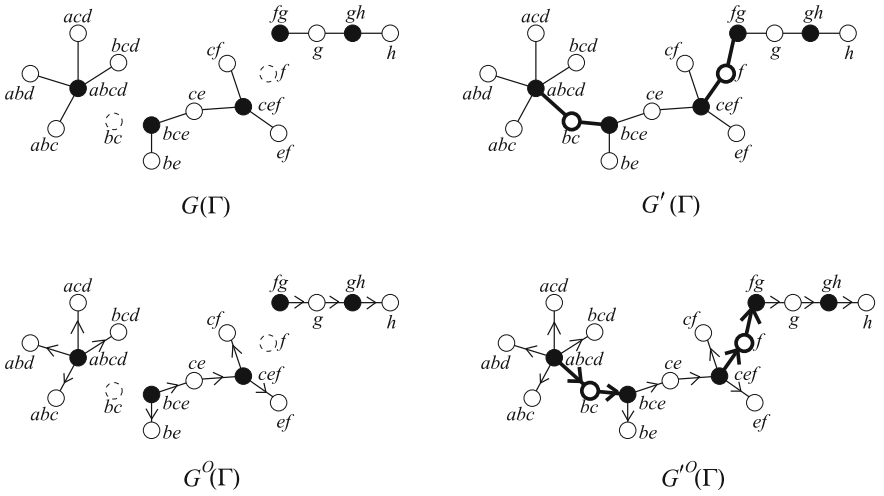


Fig. 3.3 The graphs $G(\Gamma)$ and $G'(\Gamma)$, and their orientations

$G'^{O}_{\supseteq\eta}(\Gamma)$ be the subgraph of $G'^{O}(\Gamma)$ induced by the facets, ridges, and pseudoridges containing η.

The proof of Theorem 4 is given completely in parallel to that of Theorem 3 by replacing $G^{O}_{\supseteq\eta}(\Gamma)$ by $G'^{O}_{\supseteq\eta}(\Gamma)$. In the definitions of $S^{O}(\sigma)$ and $S^{cO}(\sigma)$, we also replace $G^{O}(\Gamma)$ by $G'^{O}(\Gamma)$ as follows. (Formally, $S^{O}(\sigma)$ is the same as the original definition (1). The replacement is essential for the description of $S^{cO}(\sigma)$.)

$$\begin{aligned} S^{O}(\sigma) &= \{\eta \in \Gamma : \sigma \to \tau \text{ in } G'^{O}(\Gamma) \text{ for every (pseudo)ridge } \tau \text{ with } \eta \subseteq \tau \subseteq \sigma\} \cup \{\sigma\} \\ &= \{\eta \in \Gamma : \sigma \to \tau \text{ in } G^{O}(\Gamma) \text{ for every ridge } \tau \text{ with } \eta \subseteq \tau \subseteq \sigma\} \cup \{\sigma\}, \qquad (3.3) \\ S^{cO}(\sigma) &= \overline{\sigma} - S^{O}(\sigma) \\ &= \{\eta \in \Gamma : \sigma \leftarrow \tau \text{ in } G'^{O}(\Gamma) \text{ for some (pseudo)ridge } \tau \text{ with } \eta \subseteq \tau \subseteq \sigma\} \\ &= \bigcup\{\overline{\tau} : \tau \in \mathcal{R}(\Gamma) \cup \mathcal{R}'(\Gamma),\ \sigma \leftarrow \tau \text{ in } G'^{O}(\Gamma)\}. \qquad (3.4) \end{aligned}$$

When $\{S^{O}(\sigma) : \sigma \in \mathcal{F}(\Gamma)\}$ is a partition, we define $\widetilde{G}(\Gamma)$ as same as the pure case. That is, we define a graph $\widetilde{G}^{O}(\Gamma)$ whose nodes are facets of Γ and arcs $\sigma \to \sigma'$ are defined if there is a face $\eta \subseteq \sigma'$ with $\eta \in S^{O}(\sigma)$.

By this replacement, the whole argument in Theorem 3 works for the nonpure case. Theorem 4 is verified by examining the following lemmas.

Lemma 6 *Let Γ be a simplicial complex and let $\eta \in \Gamma$ and $\sigma \in \mathcal{F}(\Gamma)$. Then, for any orientation O, $\eta \in S^{O}(\Gamma)$ if and only if σ is a source node in $G'^{O}_{\supseteq\eta}(\Gamma)$.*

Lemma 7 *Let Γ be a simplicial complex and O an orientation of $G(\Gamma)$ that is acyclic and admissible. Then we have $\sum_{\sigma\in\mathcal{F}(\Gamma)} 2^{\text{out-deg}(\sigma;\, G^{O}(\Gamma))} \geq f(\Gamma)$.*

Lemma 8 *Let Γ be a simplicial complex and O an orientation of $G(\Gamma)$ that is acyclic and admissible. The equality $\sum_{\sigma\in\mathcal{F}(\Gamma)} 2^{\text{out-deg}(\sigma;\, G^O(\Gamma))} = f(\Gamma)$ holds if and only if $\{S^O(\sigma) : \sigma \in \mathcal{F}(\Gamma)\}$ forms a partition of Γ.*

Lemma 9 *When $\{S^O(\sigma) : \sigma \in \mathcal{F}(\Gamma)\}$ is a partition of a simplicial complex Γ, the following are equivalent.*
- *$\widetilde{G}^O(\Gamma)$ is acyclic,*
- *$G'^O(\sigma)$ is acyclic,*
- *$G^O(\sigma)$ is acyclic.*

Lemma 10 *For a simplicial complex Γ, there exists an acyclic and admissible orientation O of $G(\Gamma)$ such that $\{S^O(\sigma) : \sigma \in \mathcal{F}\}$ is a partition of Γ with $\widetilde{G}(\Gamma)$ acyclic if and only if Γ is shellable.*

The proofs of Lemmas 6 to 10 are completely the same as the pure case. The proof of Theorem 4 is also the same as the pure case.

Proof (Proof of Theorem 4) The inequality $V^* \geq f(\Gamma)$ follows from Lemma 7. Further, Lemma 8 shows that the equality holds if and only if $\{S^O(\sigma) : \sigma \in \mathcal{F}(\Gamma)\}$ is a partition of Γ, and for this partition we have $\widetilde{G}(\Gamma)$ acyclic by Lemma 9. Finally, Lemma 10 shows this is equivalent to the shellability of Γ. □

The trick of generalizing the pure case of Theorem 3 to the nonpure case of Theorem 4 can be understood from the well-known "Rearrangement lemma" of Björner and Wachs [4, Lemma 2.6]. According to the Rearrangement lemma, any shelling of a shellable simplicial complex Γ can be rearranged such that the facets in the shelling are ordered in a descending order with respect to dimension, without changing the restriction maps. In our theorems, setting restriction maps corresponds to giving orientations to the facet-ridge incidence graph, and shellings with fixed restriction maps are derived as linear extensions of $G'^O(\Gamma)$ restricted to facets. As remarked in [4, p. 1305], (after the rearrangement) any shelling of a nonpure simplicial complex of dimension d has the structure such that first d-dimensional facets are shelled, and after that $(d-1)$-dimensional facets follow extending a shelling of the $(d-1)$-skeleton of the d-dimensional part, and then $(d-2)$-dimensional facets follow in the same way. This process continues until all the facets are shelled. The orientation of $G'^O(\Gamma)$ extending $G^O(\Gamma)$ forces this structure.

The result of Theorem 4 is first shown in [13], and also later appears in [11] with a generalized framework for cell complexes.

Remark 1 In the optimization problem (P2) or (P3), in the optimal orientation, every ridge has in-degree equal to 1. To see this, assume in an acyclic and admissible orientation O, there is a ridge node τ that has in-degree $k \geq 2$ with $\sigma_1 \to \tau, \sigma_2 \to \tau, \ldots \sigma_k \to \tau$. Then, we can observe there is at least one σ_i such that reversing the orientation to $\sigma_i \leftarrow \tau$ remains the orientation acyclic (and obviously also admissible) as follows. If reversing $\sigma_1 \to \tau$ to $\sigma_1 \leftarrow \tau$ in O makes a cycle, then there should exist a directed path from σ_1 to some σ_{i_1}. If reversing $\sigma_{i_1} \to \tau$ to $\sigma_{i_1} \leftarrow \tau$ in O makes

a cycle, then there should exist a directed path from σ_{i_1} to some σ_{i_2}. By continuing this way, at some $l \leq k$, we will find a σ_{i_l} such that reversing $\sigma_{i_l} \to \tau$ to $\sigma_{i_l} \leftarrow \tau$ in O remain the orientation acyclic, since otherwise we have a cycle $\sigma_{i_j} \rightsquigarrow \sigma_{i_{j+1}} \rightsquigarrow \sigma_{i_{j+2}} \rightsquigarrow \cdots \rightsquigarrow \sigma_{i_l} = \sigma_{i_j}$ because k is finite. Since reversing one $\sigma_{i_l} \to \tau$ to $\sigma_{i_l} \leftarrow \tau$ makes the value of the objective function smaller, we conclude that an orientation O cannot be an optimal solution if there is a ridge node with in-degree ≥ 2.

Remark 2 The optimization problem (P1) in the setting of Theorem 2 (setting $G = G(P)$ for a simple polytope P) is in fact a special case of the problem (P2) in Theorem 3. For a simple polytope P, let P^* be the polar dual of P. P^* is a simplicial polytope, and thus its boundary ∂P^* is a simplicial complex. Then, the facet-ridge incidence graph $G(\partial P^*)$ is isomorphic to a subdivision of the graph $G(P)$ introducing one node (corresponding to a ridge) on each edge. Note that each ridge node in $G(\partial P^*)$ has degree 2. Here, as is explained in the previous remark, the optimal orientation of the problem (P2) has in-deg$(\tau) = 1$ for each ridge node τ. Since each ridge in $G(\partial P^*)$ has exactly two adjacent facets, the orientation optimal for (P2) can be naturally translated to an orientation for (P1), and the resulted orientation is an optimal orientation for (P1). This relation shows that the optimal orientations of (P1) give shellings of P^* as their linear extensions. Such a relation between good orientations of simple polytopes and shellings of their duals has been known already, see [20] for example.

The optimization problem (P1) can be used for characterizing shellability of pseudomanifolds. A (closed) *pseudomanifold* is a pure simplicial complex such that each ridge is contained by exactly two facets. As is noted in Sect. 3.1, the recognition of shellability of pseudomanifolds is easy for the 2-dimensional case [8], but no efficient algorithms are known for 3-dimensional and higher cases. Since each ridge node has exactly two facet nodes in the facet-ridge incidence graph of a pseudomanifold, (P2) can be reduced to (P1) for the case of pseudomanifolds by the same reason as for ∂P^*. The facet-ridge incidence graph of a d-dimensional pseudomanifold is a $(d+1)$-regular graph. This suggests that the problem (P1) is likely a difficult optimization problem even if we restrict the graph G to be a k-regular graph with $k \geq 4$.

3.3 Cubical Complexes and Acyclic Partitions

A simplicial complex, discussed in Sect. 3.2 is a cell complex in which each cell is a simplex. Likewise, a *cubical complex* is a cell complex in which each cell is (combinatorially equivalent to) a (hyper)cube. In this section, we develop a theory for cubical complexes similar to that for simplicial complexes. (More precisely, what we are considering here is a regular CW complex in which each cell is combinatorially equivalent to a (hyper)cube. Usually, it is required that cubical complexes satisfy the intersection property, i.e., the nonempty intersection of two cells is always a cell in the complex, but we do not need this condition.) This result appeared in Hachimori [11]. We here follow the discussion in [11].

Recall the story of our theory for simplicial complexes in the previous section. In the optimization problem of (P2) or (P3), the objective function $\sum_{\sigma\in\mathcal{F}(\Gamma)} 2^{\text{out-deg}(\sigma;G^O(\Gamma))}$ is equal to $\sum_{\sigma\in\mathcal{F}(\Gamma)} |S^O(\sigma)|$, where $S^O(\sigma)$ is the set of faces of a facet σ generated by the ridges τ with orientation $\sigma \to \tau$. On the other hand, the constraint of the optimization problem that the orientations must be acyclic and admissible (i.e., each ridge has in-degree at least 1) assures that the family $\{S^O(\sigma)\}$ always forms a covering of Γ. Hence, the condition that the minimum value of the optimization problem equals the number of the faces of Γ turns out to be equivalent to that $\{S^O(\sigma) : \sigma \in \mathcal{F}(\Gamma)\}$ is a partition with an acyclic structure, i.e., such that the graph $\widetilde{G}^O(\Gamma)$ is acyclic. We say such a partition an "*acyclic partition*." In this story for simplicial complexes, the existence of acyclic partitions happens to be equivalent to be shellable, and this concludes the proof of Theorem 4.

For cubical complexes, the same story can be developed except the last part. We define $G(\Gamma)$ and $G'(\Gamma)$ analogously to Sect. 3.2 with the same definition of facets, ridges, and pseudoridges. For a given orientation O of $G(\Gamma)$, we extend the orientation to $G'(\Gamma)$ by the same rule. We say an orientation O is admissible if in-deg$(\tau) \geq 1$ for every $\tau \in \mathcal{R}(\Gamma)$. In a cubical complex Γ, each facet σ contains $\dim\sigma$ antipodal pairs of (pseudo)ridges of dimension $\dim\sigma - 1$. (For example, a three-dimensional cube has three antipodal pairs of two-dimensional (pseudo)ridges.) According to the orientation O of $G(\Gamma)$ and thus of $G'(\Gamma)$, we define $(t_0^O(\Gamma), t_1^O(\Gamma), t_2^O(\Gamma))$ the *type* of the facet σ, where

$$
\begin{aligned}
t_0^O(\sigma) &= \#\text{ of antipodal pairs of (pseudo)ridges } \{\tau, \tau'\} \text{ with } \sigma \to \tau \text{ and } \sigma \to \tau',\\
t_2^O(\sigma) &= \#\text{ of antipodal pairs of (pseudo)ridges } \{\tau, \tau'\} \text{ with } \sigma \leftarrow \tau \text{ and } \sigma \leftarrow \tau',\\
t_1^O(\sigma) &= \dim\sigma - t_0^O(\sigma) - t_2^O(\sigma).
\end{aligned}
$$

For cubical complexes, we develop the theory on $\check{\Gamma} = \Gamma - \emptyset$ instead of Γ. For $\sigma \in \mathcal{F}(\Gamma)$, we define $\check{S}^O(\sigma) = S^O(\sigma) - \emptyset$ and $\check{S}^{cO}(\sigma) = \overline{\sigma} - \check{S}^O(\sigma)$. As same as in the case of simplicial complexes, define a graph $\widetilde{G}^O(\Gamma)$ whose nodes are facets of Γ and arc $\sigma \to \sigma'$ is defined if there is a face $\eta \subseteq \sigma'$ with $\eta \in \check{S}^O(\sigma)$. If there exists an orientation O for a cubical complex Γ such that $\{\check{S}^O(\sigma) : \sigma \in \mathcal{F}(\Gamma)\}$ is a partition of $\check{\Gamma}$ with $\widetilde{G}^O(\Gamma)$ acyclic, we say $\check{\Gamma}$ has an *acyclic partition*. Now we have the following theorem.

Theorem 5 *For a cubical complex Γ, let us consider the following minimization problem:*

$$
\begin{aligned}
(P4): \quad \min \quad & \sum_{\sigma\in\mathcal{F}(\Gamma)} 2^{t_1^O(\sigma)} 3^{t_0^O(\sigma)}\\
\text{s. t.} \quad & O \text{ is acyclic and admissible.}
\end{aligned}
$$

Then the optimum value V^* of (P4) satisfies $V^* \geq f(\check{\Gamma})$, where $\check{\Gamma} = \Gamma - \emptyset$ and $f(\check{\Gamma})$ is the number of all the faces of $\check{\Gamma}$. Further, the equality holds if and only if $\check{\Gamma}$ has an acyclic partition.

The proof of this theorem is completely the same as Theorem 4. Here, in the objective function of (P4), $2^{t_1^O(\sigma)}3^{t_0^O(\sigma)}$ equals the number of faces contained in $\check{S}^O(\sigma)$. (One reason we removed the empty set and replaced Γ by $\check{\Gamma}$ is to represent the number of faces by this formula.) In Theorem 4 of the case of simplicial complexes, the existence of acyclic partitions is equivalent to shellability as Lemma 10. Unfortunately, however, we lack this equivalence for cubical complexes.

Remark 3 $S^O(\sigma)$ and $\check{S}^O(\sigma)$ differ only when $S^O(\sigma) = \overline{\sigma}$, in this case $\check{S}(\sigma) = S(\sigma) - \emptyset$. The difference between an acyclic partition of Γ and an acyclic partition of $\check{\Gamma}$ is the treatment of the empty set. For an acyclic partition of Γ, we require that the empty set should be contained in exactly one $S^O(\sigma)$. This requires that the oriented graph $G^O(\Gamma)$ has exactly one source node. On the other hand, for an acyclic partition of $\check{\Gamma}$, we remove the empty set from $\check{\Gamma}$ and from each $\check{S}^O(\sigma)$. Hence $G^O(\sigma)$ can have more than one source nodes. If O induces an acyclic partition of $\check{\Gamma}$ such that $G^O(\Gamma)$ has only one source node, then the orientation O also induces an acyclic partition of Γ.

For cubical complexes, more generally for a general class of cell complexes called "regular CW complexes" (including polytopal complexes), shellability is defined in the following recursive form.

Definition 2 (*Björner and Wachs* [5, Sec. 13]) In a regular CW complex Γ, an ordering $\sigma_1, \sigma_2, \ldots, \sigma_t$ of the facets of Γ is called a shelling if either $\dim \Gamma = 0$ or if $\dim \Gamma \geq 1$ and satisfies the following:

(i) $\partial\sigma_1$ has a shelling,
(ii) $\partial\sigma_i \cap (\bigcup_{j=1}^{i-1} \partial\sigma_i)$ is pure and $(\dim \sigma_i - 1)$-dimensional, for $2 \leq i \leq t$,
(iii) $\partial\sigma_i$ has a shelling such that facets of $\partial\sigma_i$ in $\partial\sigma_i \cap (\bigcup_{j=1}^{i-1} \partial\sigma_i)$ come first in the shelling, for $2 \leq i \leq t$,

where $\partial\sigma$ is the boundary complex of σ, i.e., the subcomplex of $\overline{\sigma}$ consisting of all proper faces of σ (i.e., all the faces of σ except σ itself). Γ is shellable if it has a shelling.

This kind of generalized version of shellability has been studied classically for pure complexes, see Björner and Wachs [3]. For a comprehensive exposition of shellability for pure polytopal complexes, see Ziegler [21, Lecture 8]. For regular CW complexes, see Björner [1].

The equivalence of acyclic partition and shellability like Lemma 10 is valid only in the class of simplicial complexes. Unfortunately, this equivalence does not hold for general cell complexes. For example, the simple example in Fig. 3.4 has an acyclic partition with the orientation shown in the figure, but it is not shellable. Hence, the optimization in Theorem 5 does not characterize shellability.

For cubical complexes, however, we can retrieve some topological information as follows. Let O be an orientation on a cubical complex Γ. We say a facet σ is *critical* if $t_1^O(\sigma) = 0$, and count the number of critical facets as follows:

$$p_i^O(\Gamma) = \#\{\sigma \in \mathcal{F}(\Gamma) : \sigma \text{ is critical and } t_2^O(\sigma) = i\}.$$

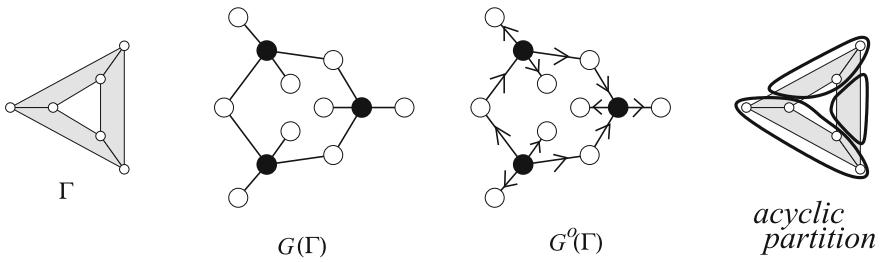


Fig. 3.4 A nonshellable cubical complex that has an acyclic partition

We say that a facet is a critical facet of *index* i if σ is critical and $t_2^O(\sigma) = i$. Thus, $p_i^O(\Gamma)$ is the number of critical facets of index i. We have the following theorem.

Theorem 6 *Let Γ be a cubical complex, and O an orientation such that $\{\check{S}^O(\sigma) : \sigma \in \mathcal{F}(\Gamma)\}$ is an acyclic partition. Then we have the following inequalities:*

$$\beta_k(\Gamma) - \beta_{k-1}(\Gamma) + \cdots m + (-1)^{k-1}\beta_0(\Gamma) \leq p_k^O(\Gamma) - p_{k-1}^O(\Gamma) + \cdots m + (-1)^{k-1}p_0^O(\Gamma), \quad (0 \leq k \leq \dim \Gamma)$$

$$\chi(\Gamma) = p_0^O(\Gamma) - p_1^O(\Gamma) + \cdots m + (-1)^{\dim \Gamma - 1} p_{\dim \Gamma}^O(\Gamma),$$
$$\beta_i \leq p_i^O, \quad (0 \leq i \leq \dim \Gamma)$$

where $\beta_i(\Gamma)$ is the ith Betti number of Γ and $\chi(\Gamma)$ is the Euler characteristic of Γ.

Proof Let $\sigma_1, \sigma_2, \ldots, \sigma_t$ be a linear extension of $\widetilde{G}^O(\Gamma)$, and $\Gamma_i = \bigcup_{j=1}^{i} \overline{\sigma}_j$. As same as the discussion in the proof of Theorem 3, $\Gamma_{i-1} \cap \overline{\sigma_i} = \check{S}^{cO}(\sigma_i)$. For each i, Γ_i is a cubical complex and we observe the following.

- If $t_1^O(\sigma_i) \geq 1$, then $\check{S}^{cO}$ is homeomorphic to a ball (of dimension $\dim \sigma_i$), and thus Γ_i is homotopy equivalent to Γ_{i-1}. (This can be verified from the fact that $S^{cO}(\sigma_i)$ is shellable, see for example [21, Exercise 8.1 (i)].)
- If $t_1^O(\sigma_i) = 0$, then Γ_i is a union of Γ_{i-1} and $\overline{\sigma}_i$, where $\overline{\sigma}_i$ is homeomorphic to the direct product of intervals $I^{t_0^O(\sigma_i)} \times I^{t_2^O(\sigma_i)}$ with the intersection $\Gamma_{i-1} \cap \overline{\sigma}_i$ corresponds to $I^{t_0(\sigma_i)} \times \{0, 1\}^{t_2(\sigma_i)}$. Thus, Γ_i is homotopy equivalent to the union of Γ_{i-1} and a $t_2^O(\sigma_i)$-dimensional cell (i.e., adding a $t_2^O(\sigma_i)$-handle to Γ_i).

By these observations, we conclude that $\Gamma = \Gamma_t$ is homotopy equivalent to a CW complex with p_i cells for each i. (See Fig. 3.5 for a simple illustrative example of this procedure.) The inequalities follow from this by following the standard argument in Morse theory, see for example [10, 17]. □

As we see in the proof of Theorem 6, acyclic partitions can be seen as a kind of discrete analogue of Morse functions on smooth manifolds. The critical facets of index i correspond to the critical points of index i of Morse functions. There is a famous discrete analogue of Morse theory by Forman [10], but our cubical analogue

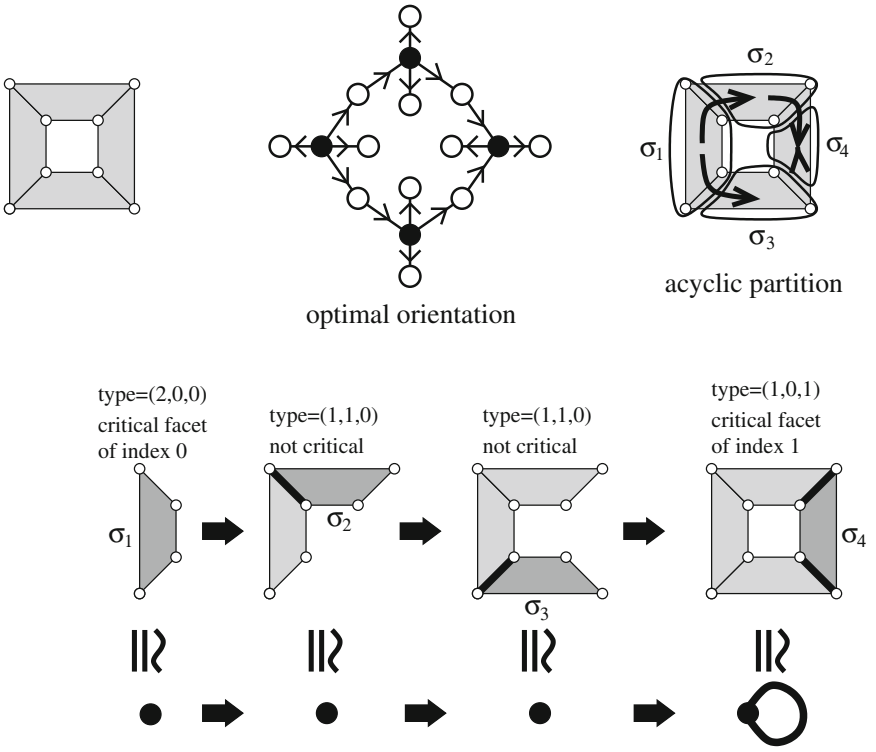


Fig. 3.5 An acyclic partition of a cubical complex homotopy equivalent to a cell complex with one 0-cell and one 1-cell

seems different from this. The similarity to Morse function can be observed further as follows. This is an analogue of the "Sphere Theorem".

Theorem 7 *Let Γ be a cubical decomposition of a closed manifold (i.e., a cubical complex homeomorphic to a closed manifold). If Γ has an acyclic partition such that $p_0 = p_{\dim \Gamma} = 1$ and $p_i = 0$ for $0 < i < \dim \Gamma$, then Γ is a PL-sphere.*

Proof This is just a consequence of that Γ is shellable if $p_0 = 1$ and $p_i = 0$ for $0 < i < \dim \Gamma$, which is easy to verify. It is well known that a regular CW decomposition of a closed manifold is a PL-sphere if it is shellable, see Björner [1] for example. □

3.4 Optimization of Orientation of Graphs Without Acyclicity Constraint

As is remarked in the end of Sect. 3.2, the problem (P1) seems a difficult optimization problem in general. The difficulty of the problem (P1) lies in the constraint that the

orientations must be acyclic. Without this constraint, the problem is easy to solve. To see this, let us consider the following optimization problem.

$$(P5): \quad \min \sum_{v \in G} 2^{\text{out-deg}(v;\, G^O)} \quad (=: \varphi(O)) \qquad \text{s. t. } O \text{ is any orientation.}$$

Lemma 11 *An orientation O is optimal for the problem (P5) if and only if there is no directed path in G^O from u to v for any $u, v \in V(G)$ with* $\text{out-deg}(v) \leq \text{out-deg}(u) - 2$.

Proof The "only if" part is easy. If there is a directed path p in G^O from u to v, let O_p be the orientation reversing the orientations of edges on the path p in O. Then we have

$$\begin{aligned} \text{out-deg}(u; G^{O_p}) &= \text{out-deg}(u; G^O) - 1, \\ \text{out-deg}(v; G^{O_p}) &= \text{out-deg}(v; G^O) + 1, \\ \text{out-deg}(w; G^{O_p}) &= \text{out-deg}(w; G^O) \qquad ({}^{\forall} w \in V(G) - \{u, x\}). \end{aligned}$$

By the condition $\text{out-deg}(v) \leq \text{out-deg}(u) - 2$, it is verified that $\varphi(O_p) < \varphi(O)$ since $2^a + 2^b > 2^{a+1} + 2^{b-1}$ if $a \leq b - 2$, hence O is not optimal.

For the "if" part, assume an orientation O has no directed path from u to v for any $u, v \in V(G)$ with $\text{out-deg}(v) \leq \text{out-deg}(u) - 2$, and O^* is an optimal orientation with $\varphi(O) > \varphi(O^*)$. Let $G^{(O,O^*)}$ be the subgraph of G induced by the edges of G with different orientations in O and O^*, and $G^{(O,O^*)O}$ ($G^{(O,O^*)O^*}$) the graph $G^{(O,O^*)}$ oriented by O (by O^*). Here, we observe that we can choose O^* such that $G^{(O,O^*)O^*}$ has no directed cycles: if there is a directed cycle in $G^{(O,O^*)O^*}$, we can reverse the orientations of the edges in O^* along the cycle without changing the value of $\varphi(O^*)$, and we get a required O^* by continuing this. Further, we choose O^* such that the number of edges of $G^{(O,O^*)}$ is minimum. Since $G^{(O,O^*)O^*}$ is acyclic and thus $G^{(O,O^*)O}$ is also acyclic, we can find a path $q = x \rightsquigarrow y$ on $G^{(O,O^*)}$ such that, in $G^{(O,O^*)O}$, x is a source, y is a sink, and the path q is a directed path from x to y. Here, we have $\text{out-deg}(x; G^{O^*}) \leq \text{out-deg}(x; G^O) - 1$ and $\text{out-deg}(y; G^{O^*}) \geq \text{out-deg}(y; G^O) + 1$ since x is a source and y is a sink in $G^{(O,O^*)O}$. We have $\text{out-deg}(y; G^O) \geq \text{out-deg}(x; G^O) - 1$ by the assumption on O. Hence we have

$$\text{out-deg}(x; G^{O^*}) \leq \text{out-deg}(x; G^O) - 1 \leq \text{out-deg}(y; G^O) \leq \text{out-deg}(y; G^{O^*}) - 1.$$

Now let O_q^* be the orientation reversing the edges on the path q in O^*. If $\text{out-deg}(x;\, O^*) \leq \text{out-deg}(y;\, O^*) - 2$, we have $f(O_q^*) < f(O^*)$, a contradiction to the optimality of O^*. If $\text{out-deg}(x;\, O^*) = \text{out-deg}(y;\, O^*) - 1$, we have $f(O_q^*) = f(O^*)$ with $|E(G^{(O,O_q^*)}| < |E(G^{(O,O^*)}|$, a contradiction to the minimality of the number of edges of $G^{(O,O^*)}$. This completes the proof of Lemma 11. □

Theorem 8 *The problem (P5) can be solved in a polynomial time.*

Proof To solve (P5), Lemma 11 suggests the following easy algorithm. First, start from an arbitrary orientation of G. Then, find a directed path $u \rightsquigarrow v$ in the orientation such that out-deg$(v) \leq$ out-deg$(u) - 2$ and reverse the orientations of edges along the path. Continue this until there is no such a directed path found. The resulted orientation is an optimal solution of (P5). Since finding such a path in each repetition can be easily done in a polynomial time, what remains is to evaluate the number of repetitions in this algorithm. For this evaluation, consider a function

$$F(O) = \sum_{\{u,v\} \in \binom{V(G)}{2}} |\text{out-deg}(u; G^O) - \text{out-deg}(v; G^O)|.$$

When the orientations of the edges are reversed along a path $p = x \rightsquigarrow y$ with out-deg$(y) \leq$ out-deg$(x) - 2$, out-deg(x) decreases and out-deg(y) increases by one respectively, and thus we have the following.

- If $w \in V(G) - \{x, y\}$ has out-deg$(w; G^O) \leq$ out-deg$(y; G^O)$ or out-deg$(w; G^O) \geq$ out-deg$(x; G^O)$, then

$$\Big(|\text{out-deg}(x; G^O) - \text{out-deg}(w; G^O)| + |\text{out-deg}(y; G^O) - \text{out-deg}(w; G^O)|\Big) - \Big(|\text{out-deg}(x; G^{O_p}) - \text{out-deg}(w; G^{O_p})| + |\text{out-deg}(y; G^{O_p}) - \text{out-deg}(w; G^{O_p})|\Big) = 0.$$

- If $w \in V(G) - \{x, y\}$ has out-deg$(y; G^O) <$ out-deg$(w; G^O) <$ out-deg$(x; G^O)$, then

$$\Big(|\text{out-deg}(x; G^O) - \text{out-deg}(w; G^O)| + |\text{out-deg}(y; G^O) - \text{out-deg}(w; G^O)|\Big) - \Big(|\text{out-deg}(x; G^{O_p}) - \text{out-deg}(w; G^{O_p})| + |\text{out-deg}(y; G^{O_p}) - \text{out-deg}(w; G^{O_p})|\Big) = 2.$$

- We have $|\text{out-deg}(x; G^O) - \text{out-deg}(y; G^O)| - |\text{out-deg}(x; G^{O_p}) - \text{out-deg}(y; G^{O_p})| = 2$ $(*)$, and $|\text{out-deg}(w; G^O) - \text{out-deg}(z; G^O)|$ remains unchanged for $w, z \in V(G) - \{x, y\}$.

Thus, in total, we have $F(O) - F(O_p) \geq 2$ (from $(*)$). On the other hand, for any orientation O we have $0 \leq F(O) < n^3$, hence the number of repetition is bounded by $n^3/2$. This completes the proof of Theorem 8. □

Lemma 11 and Theorem 8 relies only on the convexity property of the function 2^x in the summand that $2^a + 2^b > 2^{a+1} + 2^{b-1}$ for $a \leq b - 2$. Likewise, the same holds if the objective function is a function ψ satisfying the condition that $\psi(O) - \psi(O') > 0$ if the out-degrees of the nodes are the same in O and O' except u and v, out-deg$(u; G^O) \leq$ out-deg$(v; G^O) - 2$, out-deg$(u; G^{O'}) =$ out-deg$(u; G^O) + 1$,

and out-deg$(v; G^{O'}) =$ out-deg$(v; G^O) - 1$. Also, we can apply the same algorithm for the problems (P2)-(P4) without acyclicity constraint starting from an orientation with out-deg$(\tau) = 1$ for all $\tau \in \mathcal{R}(\Gamma)$ and finding a directed path $\sigma \rightsquigarrow \sigma'$ with $\sigma, \sigma' \in \mathcal{F}(\Gamma)$ in each repetition. (Note that we have out-deg$(\tau) = 1$ for all $\tau \in \mathcal{R}(\Gamma)$ in the optimal orientation as same as remarked in the end of Sect. 3.2.2.)

To conclude this chapter, we list some open problems to be studied.

For the original optimization problem (P1), such a good property as Lemma 11 does not likely hold and this makes the problem difficult. As is remarked before, (P1) seems difficult even if we restrict the graph G to be k-regular with $k \geq 4$. To look for a nontrivial class of graphs for which optimization problems like (P1)-(P4) can be solved in a polynomial time is an interesting problem. For example, is (P1) efficiently solvable for 3-regular graphs?

On the other hand, we believe the problems (P1)-(P4) are difficult to solve in general, but we do not have NP-hardness results for these problems. To show NP-hardness of these problems is an important problem.

Our results in Sect. 3.2 are based on the fact that the optimization for problems (P2) or (P3) gives an acyclic partition of a given simplicial complex. Such a partition without acyclicity is called partitionability and have been an important topic of study, see Kleinschmidt and Onn [15], Stanley [18, Ch. III.2], etc. See also Duval, Goeckner, Klivans, and Martin [9] for recent progress. Signability, introduced by Kleinschmidt and Onn [15] as a generalization of partitionability, is very closely related to our discussion in Sects. 3.2 and 3.3. Lemma 1 is essentially equivalent to the relation between partitionability and signability shown in [15] where the orientations of edges $\sigma \to \tau$ and $\sigma \leftarrow \tau$ are replaced to the assignment of signs $+$ and $-$ to the covering relations between facets σ and ridges τ. Though partitionability is a property removing the acyclicity structure from shellability, unfortunately partitionability cannot be represented by the optimization problems just removing acyclicity constraints from (P2)-(P4) as is considered in (P5). To assure partitionability, $G^O(\Gamma)_{\supseteq \eta}$ should have exactly one source facet node for all faces $\eta \in \Gamma$. In Theorem 3, for this requirement, acyclicity assures that each $G^O(\Gamma)_{\supseteq \eta}$ has at least one source facet node, and optimization reduces it to exactly equal to one node. For partitionability, the lack of acyclicity makes it difficult to assure $G^O(\Gamma)_{\supseteq \eta}$ to have at least one source facet node. How to treat partitionability in a similar framework is a difficult problem.

Related to shellability and partitionability, Hachimori and Kashiwabara [12] introduced hereditary-shellability and hereditary-partitionability, which are properties requiring the restriction to any vertex subset has the property to be shellable and partitionable. (Other related hereditary properties are defined in the same way.) This is motivated by the notion of obstructions introduced by Wachs [19]. To treat these hereditary properties in the optimization setting is a quite open problem.

Finally, to look for other topics that can be formulated using optimizations on orientations of graphs will be an interesting problem.

References

1. Björner, A.: Posets, regular CW complexes and Bruhat order. Eur. J. Combin. **5**, 7–16 (1984)
2. Björner, A.: Topological methods. In: Graham, R., Grötschel, M., Lovász, L. (eds.) Handbook of Combinatorics, pp. 1819–1872. North-Holland (1995)
3. Björner, A., Wachs, M.: On lexicographically shellable posets. Trans. Am. Math. Soc. **277**, 323–341 (1983)
4. Björner, A., Wachs, M.: Shellable nonpure complexes and posets I. Trans. Am. Math. Soc. **348**, 1299–1327 (1996)
5. Björner, A., Wachs, M.: Shellable nonpure complexes and posets II. Trans. Am. Math. Soc. **349**, 3945–3975 (1997)
6. Blind, R., Mani, P.: On puzzles and polytope isomorphisms. Aequationes Math. **34**, 287–297 (1987)
7. Colbourn, C.J.: The Combinatorics of Network Reliability. Oxford University Press, Oxford (1987)
8. Danaraj, G., Klee, V.: A presentation of 2-dimensional pseudomanifolds and its use in the design of a linear-time shelling algorithm. Ann. Discrete Math. **2**, 53–63 (1978)
9. Duval, A.M., Goeckner, B., Klivans, C.J., Martin, J.L.: A non-partitionable Cohen-Macaulay simplicial complex. Adv. Math. **299**, 381–395 (2016)
10. Forman, R.: Morse theory for cell complexes. Adv. Math. **134**, 90–145 (1998)
11. Hachimori, M.: Orientations on simplicial complexes and cubical complexes, unpublished manuscript, 8 p. (http://infoshako.sk.tsukuba.ac.jp/~hachi/archives/cubic_morse4.pdf)
12. Hachimori, M., Kashiwabara, K.: Obstructions to shellability, partitionability, and sequential Cohen-Macaulayness. J. Combin. Theory Ser. A **118**(5), 1608–1623 (2011)
13. Hachimori, M., Moriyama, S.: A note on shellability and acyclic orientations. Discrete Math. **308**, 2379–2381 (2008)
14. Kalai, G.: A simple way to tell a simple polytope from its graph. J. Combin. Theory Ser. A **49**(2), 381–383 (1988)
15. Kleinschmidt, P., Onn, S.: Signable posets and partitionable simplicial complexes. Discrete Comput. Geom. **15**, 443–466 (1996)
16. Kaibel, V., Pfetsch, M.: Some algorithmic problems in polytope theory. In: Joswig, M., Takayama, N. (eds.) Algebra, Geometry and Software Systems, pp. 23–47. Springer, Berlin (2003)
17. Milnor, J.: Morse Theory. Princeton University Press, Princeton (1963)
18. Stanley, R.P.: Combinatorics and Commutative Algebra, 2nd edn. Birkhäuser, Boston (1996)
19. Wachs, M.: Obstructions to shellability. Discrete Comput. Geom. **22**, 95–103 (2000)
20. Williamson Hoke, K.: Completely unimodal numberings of a simple polytope. Discrete Appl. Math. **20**, 69–81 (1996)
21. Ziegler, G.: Lectures on Polytopes. Springer, Berlin (1994). Second revised printing 1998

Chapter 4
On Ideal Minimally Non-packing Clutters

Kenji Kashiwabara and Tadashi Sakuma

4.1 Introduction

4.1.1 Background and Motivation

In the celebrated paper [15] of Seymour, motivated by the pluperfect and (weak) perfect graph theorems for the set covering problem by Fulkerson and Lovász, he introduced the concept of so-called "the Max-Flow-Min-Cut property" of clutters, which is the packing counterpart of the totally dual integrality built in the perfection. That is, a clutter $\mathscr{C}$ has the *Max-Flow-Min-Cut property* (the MFMC property, for short) if, for its clutter matrix $M(\mathscr{C})$, the linear system $M(\mathscr{C})\mathbf{x} \geqq \mathbf{1}$, $\mathbf{x} \geqq \mathbf{0}$ is totally dual integral. A matrix inequality $A\mathbf{x} \geq \mathbf{b}$ (resp. to $A\mathbf{x} \leq \mathbf{b}$) is called *totally dual integral* if the linear program $\min\{\langle \mathbf{w}, \mathbf{x}\rangle | A\mathbf{x} \geq \mathbf{b}\}$ (resp. to $\max\{\langle \mathbf{w}, \mathbf{x}\rangle | A\mathbf{x} \leq \mathbf{b}\}$) has an integral optimal dual solution $\mathbf{y}$ for every integral cost vector $\mathbf{w}$ for which the above linear program has a finite optimum. In the case of the anti-blocking polytope of a clutter matrix, its integrality and the totally dual integrality of its linear system are coincident with the perfection. Seymour[15] also pointed out that this "obvious analog" of the set covering problem is false for the set packing problem, because there exists a non-MFMC clutter $Q_6 := \big\{\{1, 3, 5\}, \{1, 4, 6\}, \{2, 3, 6\}, \{2, 4, 5\}\big\}$ whose blocking polyhedron $\{\mathbf{x} \in \mathbb{R}^6 | \mathbf{0} \leq \mathbf{x}, M(Q_6)\mathbf{x} \geq \mathbf{1}\}$ is integral (i.e., *ideal*). On the other hand, he proved that this Q_6 is the only ideal binary clutter which is minimally non-MFMC as the meaning of clutter minor.

K. Kashiwabara
Department of General Systems Studies, University of Tokyo, 3-8-1 Komaba, Meguro-ku, Tokyo 153-8902, Japan
e-mail: kashiwa@idea.c.u-tokyo.ac.jp

T. Sakuma (✉)
Faculty of Science, Yamagata University, 1-4-12 Kojirakawa, Yamagata 990-8560, Japan
e-mail: sakuma@sci.kj.yamagata-u.ac.jp

S. K. Neogy et al. (eds.), *Mathematical Programming and Game Theory*, Indian Statistical Institute Series, https://doi.org/10.1007/978-981-13-3059-9_4

A clutter $\mathscr{C}$ has the *packing property* (resp. *packs*) if the both sides of the linear programming equation $\min\{\langle \boldsymbol{\omega}, \mathbf{x}\rangle | \mathbf{x} \geqq \mathbf{1}, M(\mathscr{C})\mathbf{x} \geqq \mathbf{1}\} = \max\{\langle \mathbf{y}, \mathbf{1}\rangle | \mathbf{y} \geqq \mathbf{0}, \mathbf{y}M(\mathscr{C}) \leqq \boldsymbol{\omega}\}$ have optimal solution integral vectors $\mathbf{x}$ and $\mathbf{y}$ for all cost vectors $\boldsymbol{\omega}$ with components equal to 0, 1 or ∞ (resp. when $\boldsymbol{\omega} = \mathbf{1}$). Lehman [12] proved that the packing property implies the idealness. However, the converse is false because the ideal clutter Q_6 does not pack and hence does not have the packing property. By definition, the MFMC property implies the packing property. But how about the converse? In 1993, Conforti and Cornuéjols [1] proposed the following famous conjecture.

Conjecture 1 *(Conforti and Cornuéjols 1993)* A clutter has the packing property if and only if it has the MFMC property.

Despite its natural appearance, this conjecture is very difficult and still open. The existing approaches can be classified roughly into two categories

The first category is to find a clutter class for which the conjecture is affirmative (or, if possible, false). Conjecture 1 holds for the binary clutters [15], the diadic clutters [4], the clutter of circuits of a digraph [8], the clutter of cycles in an undirected graph [6], the broken circuit clutter of two-dimension affine convex geometries [9], the Ehrhart clutters [13], and so on. However, for the almost all of them, except for the case of the binary clutters shown in the Seymour's initial paper [15], the MFMC property is coincident with not only the packing property but also the idealness. In other words, there are only minimally non-ideal excluded clutter minors for these classes to have the MFMC property. Of course, there are several known clutter classes, other than the binary clutters, on which the packing property is not coincident with the idealness. It is well known that the clutter of disjoins [3, 14, 16] falls into the case. See also [11] for another example. However, as far as the authors know, Conjecture 1 seems unsettled even if restricted to each of these classes. To begin with, there are so few clutter classes on which the packing property is characterized by the set of minimally excluded minors which inevitably includes some ideal clutters (again, see [11]).

The second category is to investigate "the packing property" itself and extract key nature of the concept by which we can prove or disprove the conjecture. The first essential step on this line was achieved by Cornuéjols, Guenin and Margot [4]. Starting with the discovery of Q_6 [15], there have been found numerous (and several infinite families of) ideal minimally non-packing clutters until today (e.g., [3, 4, 7, 11, 14, 16]). All of these existing clutters have the common property: The blocking number is 2 for all of them. Cornuéjols, Guenin and Margot [4] conjectured that the converse is also true.

Conjecture 2 *The blocking number of every ideal minimally non-packing clutter is* 2.

Furthermore, they proved that the above conjecture implies Conjecture 1.

In this chapter, the authors will provide a framework to attack Conjecture 2. A *tilde core* of an ideal minimally non-packing clutter $\mathscr{C}$ is the maximal set of hyperedges of $\mathscr{C}$ such that every minimum transversal of $\mathscr{C}$ has a unique common element

with each of the hyperedges. As the concept of the *core* has greatly developed the theory of minimally non-ideal clutters (see [2] for details), the concept of the tilde core may have similar impact to the theory of ideal minimally non-packing clutters. Actually, Cornuéjols, Guenin and Margot [4] proved that several key features of the ideal minimally non-packing clutters are controlled by their tilde cores. The authors will develop their idea to a framework to check whether a given clutter can be a tilde core of an ideal minimally non-packing clutter or not. This framework is useful not only for the search of counterexamples but also to prove the conjecture. We demonstrate this by applying our framework to the case of a special clutter, namely, the combinatorial affine planes. We show that every combinatorial affine plane whose blocking number is at least 3 cannot be a tilde core of any ideal minimally non-packing clutter (Theorem 8).

In connection with this, we should note that whether a combinatorial projective plane except for the Fano plane F_7 can be a core of a minimally non-ideal clutter or not is a famous open question of the theory of minimally non-ideal clutters (see Question 6 in [5]).

4.1.2 Overview of Our Results

We consider Conjecture 2 in this chapter. That is, we consider the (non-)existence problem of an ideal minimally non-packing clutter of blocking number at least 3. We propose a new framework to attack the conjecture.

Let E be a finite ground set of clutters throughout this chapter. $\tilde{\mathscr{C}}$ denotes the set of hyperedges in a clutter $\mathscr{C}$ each of which intersects every minimum transversal in exactly one element. A tilde clutter $\tilde{\mathscr{C}}$ was first introduced in Cornuéjols, Guenin and Margot [4]. That paper gave necessary conditions for $\mathscr{C}$ to be an ideal minimally non-packing clutter in terms of $\tilde{\mathscr{C}}$. In our chapter, we develop their idea. We contrive tractable necessary conditions for $\mathscr{C}$ to be an ideal minimally non-packing clutter in terms of $\tilde{\mathscr{C}}$. By our approach, clutters that we have to consider are restricted.

We divide the (non-)existence problem of an ideal minimally non-packing clutter $\mathscr{D}$ as in Conjecture 2 into two steps. In the first step (Sect. 4.3), we give necessary conditions for $\mathscr{C} = \tilde{\mathscr{D}}$ when $\mathscr{D}$ is an ideal minimally non-packing clutter. We call a clutter satisfying the conditions in the step 1 a precore clutter. In the second step (Sect. 4.4), for a precore clutter $\mathscr{C}$, we consider whether $\mathscr{C}$ has an ideal minimally non-packing clutter $\mathscr{D}$ with $\mathscr{C} = \tilde{\mathscr{D}}$. Since the necessary conditions in step 1 are rather strong, clutters that we have to consider are much confined. However, we found a several classes of precore clutters. When we try to find a counterexample or prove the conjecture, we have only to consider the problem for each precore clutter $\mathscr{C}$. That is, it is the problem for $\mathscr{C}$ to have an ideal non-minimally non-packing clutter $\mathscr{D}$ with $\mathscr{C} = \tilde{\mathscr{D}}$. Starting with a (rather vague) task to find a counterexample to Conjecture 2, here, we obtain a tractable concrete problem whether a given clutter $\mathscr{C}$ has an ideal non-minimally non-packing clutter $\mathscr{D}$ with $\mathscr{C} = \tilde{\mathscr{D}}$ or not.

Section 4.3 is devoted to step 1. For an ideal non-packing clutter $\mathscr{D}$, we present several necessary conditions of $\tilde{\mathscr{D}}$: the integral blocking condition, tilde-invariance, the integrality of $\mathrm{I}(\tilde{\mathscr{D}})$, and non-separability (Theorem 2). The integral blocking condition is defined as the coincidence of the fractional packing number and the blocking number. This condition is a fundamental condition as a premise of an argument. A clutter $\mathscr{C}$ satisfying the integral blocking condition is called tilde-invariant if $\mathscr{C} = \tilde{\mathscr{C}}$ holds. For an ideal clutter $\mathscr{C}$, $\tilde{\mathscr{C}}$ is tilde-invariant. A clutter is ideal if and only if the blocking polyhedron $\{x \in \mathbb{R}^E | \langle 1_H, x\rangle \geq 1 \text{ for all } H \in \mathscr{C}, x \geq 0\}$ is an integral polyhedron. The polyhedron $\mathrm{I}(\mathscr{C})$ is a face of the above blocking polyhedron defined by the equalities corresponding to minimum transversals. We show that, for an ideal clutter $\mathscr{C}$, not only $\mathrm{I}(\mathscr{C})$ but also $\mathrm{I}(\tilde{\mathscr{C}})$ is an integral polyhedron (Theorem 1). The minimum transversals define the affine hull of $\mathrm{I}(\mathscr{C})$ and some non-minimum transversals define facets of $\mathrm{I}(\mathscr{C})$. By observing these transversals carefully, we can derive useful information from them.

Cornuéjols, Guenin, and Margot [4] proved that deleting all the elements on a hyperedge of an ideal minimally non-packing clutter decreases the blocking number by at least two. We call such a condition hyperedge-non-separability. We also present a condition called non-separability, which is a generalization of hyperedge-separability.

We have the following implications among conditions on $\mathscr{C}$ under the conditions that its minimum transversals cover E and the integral blocking condition (Lemmas 6, 12 and 13).

Integrality of $\mathrm{I}(\mathscr{C})$ $\Rightarrow$ tilde-full condition+dimension condition $\Rightarrow$
tilde-full condition $\Rightarrow$ weak tilde-invariant clutter.

In Sect. 4.4, when a precore clutter $\mathscr{C}$ is given, we present several necessary conditions for an ideal minimally non-packing clutter $\mathscr{D}$ with $\mathscr{C} = \tilde{\mathscr{D}}$: Conditions IM, IF, H, and B (Theorems 4 and 5). Since these conditions for $\mathscr{D}$ are strong enough, the next condition for a precore $\mathscr{C}$ is derived. When a precore clutter $\mathscr{C}$ has an ideal minimally non-packing clutter, there must exist a clutter $\mathscr{D}$ satisfying Conditions IM, IF, H, and B (Corollary 3).

In Sect. 4.5, we consider the problem with an additional condition that the maximum fractional packing is unique. Many classes of precore clutters satisfy this condition as far as we know. In this case, $I(\tilde{\mathscr{C}})$ for a precore clutter is an integral simplex. This condition is characterized in terms of transversals and a condition about dimension (Theorem 6). We give an example of a precore clutter, namely, a combinatorial affine plane. The clutter $\mathscr{C}$ of a combinatorial affine plane is obtained by deleting one element from a combinatorial projective plane. We show that the clutter $\mathscr{C}$ of a combinatorial affine plane is a precore clutter (Theorem 7). Moreover, we show that the clutter $\mathscr{C}$ of a combinatorial affine plane cannot have a counterexample $\mathscr{D}$ to Conjecture 2 with $\mathscr{C} = \tilde{\mathscr{D}}$ (Theorem 8).

4.2 Preliminaries

Let E be a finite set. A family $\mathscr{C} \subseteq 2^E$ of sets is said to be a *clutter* if no member includes another member. A member of $\mathscr{C}$ is called a *hyperedge*. For details about clutters, please refer to [2].

For a clutter $\mathscr{C}$, a set on E is a *transversal* if it intersects every element of $\mathscr{C}$ and it is minimal with respect to inclusion in such sets.[1] $\mathrm{b}(\mathscr{C})$ denotes the clutter consisting of all the transversals of $\mathscr{C}$. A *minimum transversal* of $\mathscr{C}$ is a transversal of the minimum size. $\mathrm{minb}(\mathscr{C})$ denotes the set of minimum transversals of a clutter $\mathscr{C}$. Note that we assume that the word "transversal" always means a "minimal" transversal to avoid the confusion between a minimum transversal and a minimal transversal in our definition.

The *blocking number* $\mathrm{bn}(\mathscr{C})$ of a clutter $\mathscr{C}$ is the minimum size of a transversal in $\mathrm{b}(\mathscr{C})$. The *packing number* $\mathrm{pn}(\mathscr{C})$ of a clutter $\mathscr{C}$ is the maximum size of a family of hyperedges of a clutter such that any pair of them does not intersect. Clearly, $\mathrm{pn}(\mathscr{C}) \leq \mathrm{bn}(\mathscr{C})$ holds. When $\mathrm{pn}(\mathscr{C}) = \mathrm{bn}(\mathscr{C})$ holds, $\mathscr{C}$ is said to *pack*. When $\mathscr{C} \subseteq \mathscr{C}'$, $\mathrm{pn}(\mathscr{C}) \leq \mathrm{pn}(\mathscr{C}')$ and $\mathrm{bn}(\mathscr{C}) \leq \mathrm{bn}(\mathscr{C}')$ hold.

The *contraction* of A from $\mathscr{C}$ is $\mathscr{C}/A = \min(\{X - A | X \in \mathscr{C}\})$ where min is the operation of collecting minimal sets with respect to inclusion. The *deletion* of A from $\mathscr{C}$ is $\mathscr{C}\backslash A = \{X \in \mathscr{C} | X \cap A = \emptyset\}$. A *minor* of $\mathscr{C}$ is a clutter which is obtained by contractions and deletions iteratively from $\mathscr{C}$. A *proper minor* means a minor which is not equal to the original clutter. The *restriction* of $\mathscr{C}$ to A is $\mathscr{C}[A] = \mathscr{C}\backslash A^c$.

A clutter $\mathscr{C}$ is called *minimally non-packing* if it does not pack and every proper minor packs. A clutter $\mathscr{C}$ is called *minimally non-packing with respect to deletion* if it does not pack and every proper deletion minor packs. A clutter on E is called *minimum-transversal-covered* if its minimum transversals cover E.

Lemma 1 *For a minimally non-packing clutter with respect to deletion, it is minimum-transversal-covered.*

Proof Since $\mathscr{C}$ does not pack, $\mathrm{pn}(\mathscr{C}) < \mathrm{bn}(\mathscr{C})$ holds. For a minimally non-packing clutter $\mathscr{C}$ with respect to deletion and $a \in E$, the deletion $\mathscr{C}\backslash a$ packs. Therefore $\mathrm{pn}(\mathscr{C}\backslash a) = \mathrm{bn}(\mathscr{C}\backslash a)$. Since deleting one element decreases the blocking number by at most one, we have $\mathrm{bn}(\mathscr{C}\backslash a) = \mathrm{bn}(\mathscr{C}) - 1$. Recall that the deletion of a clutter corresponds to the contraction of the clutter of its transversals. If a is not covered by any minimum transversal, every minimum transversal of $\mathscr{C}\backslash a$ is also a minimum transversal of $\mathscr{C}$, a contradiction to $\mathrm{bn}(\mathscr{C}\backslash a) = \mathrm{bn}(\mathscr{C}) - 1$. □

For a clutter $\mathscr{C}$, $M(\mathscr{C})$ denotes a clutter matrix of $\mathscr{C}$, whose row vectors coincide with the incidence vectors of its hyperedges. We consider the following linear problem.

[1] In standard terminology, this concept normally would be called a "minimal transversal". However since we only treat minimal transversals and this term is repeatedly used throughout this chapter, we include minimality in our definition for convenience.

$$\max\left\{\sum_{H\in\mathscr{C}} y(H)|yM(\mathscr{C})\leq 1_E, y\in\mathbb{R}^{\mathscr{C}}\right\}=\min\left\{\sum_{a\in E} x(a)|M(\mathscr{C})x\geq 1_{\mathscr{C}}, x\in R^E\right\}.$$

Note that the above equality always holds because of the duality theorem of the linear programming. We call the maximizing problem of y the *primal problem* and the minimizing problem of x the *dual problem*.

A clutter $\mathscr{C}$ is *ideal* if $\{x\in\mathbb{R}^E|\langle x, 1_H\rangle\geq 1 \text{ for all } H\in\mathscr{C}, x\geq 0\}$ is an integral polyhedron where 1_H is the incidence vector of H. Note that $x\geq 0$ means $x(a)\geq 0$ for every $a\in E$. It is known that every minor of an ideal clutter is an ideal clutter again.

By the complementary slackness of the linear programming, we have that, for every maximum solution y of the primal problem, every minimum solution x of the dual problem and, for every $a\in E$, $x(a)>0$ implies $\sum_{H:a\in H\in\mathscr{C}} y(H)=1$.

A *maximum fractional packing* y of a clutter $\mathscr{C}$ is a function $\mathscr{C}\to\mathbb{R}_{\geq}$ maximizing the sum $\sum_{H\in\mathscr{C}} y(H)$ such that $\sum_{H:a\in H\in\mathscr{C}} y(H)\leq 1$ for every $a\in E$. Every maximum fractional packing is an optimal solution of the primal problem. The *support* of a maximum fractional packing y is the set of hyperedges H with $y(H)>0$. Define

$$\mathrm{F}(\mathscr{C})=\{z\in\mathbb{R}^{\mathscr{C}}|z \text{ is a maximum fractional packing}\}.$$

The *fractional packing number* is $\sum_{H\in\mathscr{C}} y(H)$ for $y\in\mathrm{F}(\mathscr{C})$, denoted by $\mathrm{fpn}(\mathscr{C})$. Note that $\mathrm{bn}(\mathscr{C})\geq\mathrm{fpn}(\mathscr{C})\geq\mathrm{pn}(\mathscr{C})$. For an ideal clutter $\mathscr{C}$, $\mathrm{fpn}(\mathscr{C})=\mathrm{bn}(\mathscr{C})$ holds. When a clutter $\mathscr{C}$ packs, $\mathrm{pn}(\mathscr{C})=\mathrm{fpn}(\mathscr{C})=\mathrm{bn}(\mathscr{C})$.

$\tilde{\mathscr{C}}$ denotes the set of hyperedges in a clutter $\mathscr{C}$ which intersect every minimum transversal in exactly one element. That is,

$$\tilde{\mathscr{C}}=\{H\in\mathscr{C}:|B\cap H|=1 \text{ for all } B\in\mathrm{minb}(\mathscr{C})\}.$$

We call $\tilde{\mathscr{C}}$ the *tilde clutter* of $\mathscr{C}$. A clutter $\tilde{\mathscr{C}}$, obtained by the tilde operation, plays a crucial role in this chapter.

4.3 Precore Conditions

In this section, we present several necessary conditions for $\tilde{\mathscr{D}}$ when $\mathscr{D}$ is ideal minimally non-packing: the integral blocking condition, the integrality of $\mathrm{I}(\mathscr{C})$, and non-separability.

4.3.1 Integral Blocking Condition

Definition 1 A clutter $\mathscr{C}$ satisfies the *integral blocking condition* if its fractional packing number $\mathrm{fpn}(\mathscr{C})$ is equal to its blocking number $\mathrm{bn}(\mathscr{C})$.

Lemma 2 *Assume that a minimum-transversal-covered clutter $\mathscr{C}$ satisfies the integral blocking condition. Then, for every $y \in F(\mathscr{C})$, $\sum_{H\in\mathscr{C}} y(H)1_H = 1_E$ holds.*

Proof By the integral blocking condition, we have $\sum_{H\in\mathscr{C}} y(H) = \mathrm{bn}(\mathscr{C})$ for a maximum fractional packing y. By the complementary slackness, we have $\sum_{H\in\mathscr{C}} y(H)1_H = 1_E$ since the minimum transversals cover E. Note that its minimum transversals are optimal solutions of the dual problem by the integral blocking condition. □

Lemma 3 *Assume that a minimum-transversal-covered clutter $\mathscr{C}$ satisfies the integral blocking condition. Then every hyperedge in the support of a maximum fractional packing of $\mathscr{C}$ intersects every minimum transversal in exactly one element. That is, every hyperedge in the support of some maximum fractional packing of $\mathscr{C}$ belongs to $\tilde{\mathscr{C}}$.*

Proof By definition of transversals, every hyperedge $H \in \mathscr{C}$ and every minimum transversal B satisfy $|H \cap B| \geq 1$. When there exist some hyperedge H and some minimum transversal B with $|H \cap B| > 1$, $\langle \sum_{H\in\mathscr{C}} y(H)1_H, 1_B\rangle = \sum_{H\in\mathscr{C}} y(H)\langle 1_H, 1_B\rangle > \sum_{H\in\mathscr{C}} y(H)$. Since $1_E = \sum_{H\in\mathscr{C}} y(H)1_H$ by Lemma 2, $\langle 1_E, 1_B\rangle = \langle \sum_{H\in\mathscr{C}} y(H)1_H, 1_B\rangle > \sum_{H\in\mathscr{C}} y(H) = \mathrm{bn}(\mathscr{C})$, which contradicts the fact that $\langle 1_E, 1_B\rangle = \mathrm{bn}(\mathscr{C})$. □

So we can regard $y \in \mathrm{F}(\mathscr{C})$ as $y \in \mathrm{F}(\tilde{\mathscr{C}})$.

Lemma 4 *Assume that a minimum-transversal-covered clutter $\mathscr{C}$ satisfies the integral blocking condition. For $y \in F(\mathscr{C})$, $\sum_{H\in\tilde{\mathscr{C}}} y(H)1_H = 1_E$ and $\sum_{H\in\tilde{\mathscr{C}}} y(H) = bn(\tilde{\mathscr{C}}) = bn(\mathscr{C})$. Moreover $F(\mathscr{C}) = F(\tilde{\mathscr{C}})$ holds.*

Proof By the integral blocking condition and covering by the minimum transversals, $\sum_{H\in\mathscr{C}} y(H) = \mathrm{bn}(\mathscr{C})$ holds for $y \in \mathrm{F}(\mathscr{C})$. We have $\sum_{H\in\mathscr{C}} y(H)1_H = 1_E$ by Lemma 2. Therefore $\sum_{H\in\tilde{\mathscr{C}}} y(H)1_H = \sum_{H\in\mathscr{C}} y(H)1_H = 1_E$ holds by Lemma 3. By taking the inner product between each side of the equality and a minimum transversal B of $\mathscr{C}$, we have $\sum_{H\in\tilde{\mathscr{C}}} y(H)\langle 1_H, 1_B\rangle = \langle 1_E, 1_B\rangle$. Since $\langle 1_H, 1_B\rangle = 1$ for $H \in \tilde{\mathscr{C}}$, $\sum_{H\in\tilde{\mathscr{C}}} y(H) = \mathrm{bn}(\mathscr{C})$. Since $\tilde{\mathscr{C}} \subseteq \mathscr{C}$, $\sum_{H\in\tilde{\mathscr{C}}} y(H) \leq \mathrm{fpn}(\tilde{\mathscr{C}}) \leq \mathrm{bn}(\tilde{\mathscr{C}}) \leq \mathrm{bn}(\mathscr{C})$. Therefore y attains a maximum fractional packing of $\tilde{\mathscr{C}}$. So $\mathrm{F}(\mathscr{C}) = \mathrm{F}(\tilde{\mathscr{C}})$. We have $\mathrm{fpn}(\tilde{\mathscr{C}}) = \mathrm{bn}(\mathscr{C}) = \mathrm{bn}(\tilde{\mathscr{C}})$. □

Corollary 1 *For a minimum-transversal-covered clutter $\mathscr{C}$ which satisfies the integral blocking condition, $\tilde{\mathscr{C}}$ is minimum-transversal-covered and also satisfies the integral blocking condition. Moreover, $minb(\mathscr{C}) \subseteq minb(\tilde{\mathscr{C}})$ holds.*

Proof By Lemma 4, $\mathrm{F}(\mathscr{C}) = \mathrm{F}(\tilde{\mathscr{C}})$ and $\mathrm{bn}(\mathscr{C}) = \mathrm{bn}(\tilde{\mathscr{C}})$. So $\mathrm{fpn}(\tilde{\mathscr{C}}) = \mathrm{fpn}(\mathscr{C}) = \mathrm{bn}(\mathscr{C}) = \mathrm{bn}(\tilde{\mathscr{C}})$.

By definition, $\tilde{\mathscr{C}} \subseteq \mathscr{C}$ holds. So every minimum transversal of $\mathscr{C}$ intersects every hyperedge of $\tilde{\mathscr{C}}$. Since $\mathrm{bn}(\mathscr{C}) = \mathrm{bn}(\tilde{\mathscr{C}})$, a minimum transversal of $\mathscr{C}$ is a minimum transversal of $\tilde{\mathscr{C}}$. So the minimum transversals of $\tilde{\mathscr{C}}$ cover E. □

Example 1 Let $\mathscr{C} = \{ac, bc, bd\}$ on $E = \{a, b, c, d\}$. Then $\mathrm{b}(\mathscr{C}) = \{ab, bc, cd\}$ and $\tilde{\mathscr{C}} = \{ac, bd\}$. Since $\mathrm{b}(\tilde{\mathscr{C}}) = \{ab, bc, cd, da\}$, this is an example whose minimum transversals are different between $\mathscr{C}$ and $\tilde{\mathscr{C}}$.

Corollary 2 *Let $\mathscr{C}$ be an ideal minimum-transversal-covered clutter. Then $\mathscr{C}$ satisfies the integral blocking condition. Moreover $\tilde{\mathscr{C}}$ satisfies the integral blocking condition.*

Proof By the duality theorem of linear programming, for every maximum fractional packing y on $\mathscr{C}$, there exists a minimum transversal $B \in \mathrm{minb}(\mathscr{C})$ with $yM(\mathscr{C})1_B = |B|$. Note that we can take an integral optimal solution 1_B since $\mathscr{C}$ is ideal. By Lemma 3, $yM(\mathscr{C})1_B = yM(\tilde{\mathscr{C}})1_B$. Since $M(\tilde{\mathscr{C}})1_B = 1_{\tilde{\mathscr{C}}}$, $yM(\tilde{\mathscr{C}})1_B = \sum_{H \in \tilde{\mathscr{C}}} y(H)$ is also the fractional packing number of $\mathscr{C}$. Therefore $\mathscr{C}$ satisfies the integral blocking condition.

Moreover, by Corollary 1, $\tilde{\mathscr{C}}$ also satisfies the integral blocking condition. □

Proposition 1 *When every minor of a clutter satisfies the integral blocking condition, the clutter is an ideal clutter.*

Proof When the clutter is not ideal, it has a minimally non-ideal clutter $\mathscr{C}'$ as a minor. Then $\mathrm{fpn}(\mathscr{C}') > \mathrm{bn}(\mathscr{C}')$ since any minimum transversal of $\mathscr{C}'$ cannot be an optimal solution of the dual problem. The clutter $\mathscr{C}'$ does not satisfy the integral blocking condition. □

Lemma 5 *For a minimum-transversal-covered clutter $\mathscr{C}$ which satisfies the integral blocking condition, the number of hyperedges in $\tilde{\mathscr{C}}$ is at least the blocking number.*

Proof There exists at least one maximum fractional packing. The number of hyperedges which belong to the support is at least the blocking number since, for each minimum transversal B, every element of B intersects a different hyperedge in $\tilde{\mathscr{C}}$. Therefore the statement follows from Lemma 3. □

4.3.2 Tilde-Invariant Clutters and Tilde-Full Condition

Definition 2 A clutter $\mathscr{C}$ is a *tilde-invariant clutter* if it satisfies $\mathscr{C} = \tilde{\mathscr{C}}$ and the integral blocking condition.

Definition 3 A clutter $\mathscr{C}$ is a *weak tilde-invariant clutter* if $\mathscr{C}$ satisfies the integral blocking condition and $\tilde{\mathscr{C}}$ is a tilde-invariant clutter.

By definition, every tilde-invariant clutter is a weak tilde-invariant clutter.

Example 2 Even if the minimum transversals of a clutter cover E, and if it satisfies the integral blocking condition, it may not be a weak tilde-invariant clutter. Let $\mathscr{C} = \{abc, de, ef, fd, af, bd, ce\}$ on $\{a, b, c, d, e, f\}$ shown in Fig. 4.1. We have $\mathrm{b}(\mathscr{C}) = \{ade, bef, cfd\}$, and $\tilde{\mathscr{C}} = \{af, bd, ce, abc\}$. Since $\tilde{\mathscr{C}}$ is not a tilde-invariant clutter, $\mathscr{C}$ is not a weak tilde-invariant clutter.

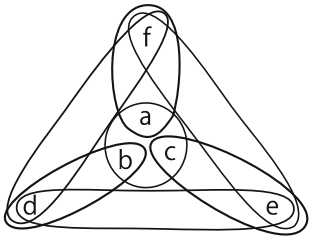


Fig. 4.1 An example of a non-weak tilde-invariant clutter

Definition 4 A clutter $\mathscr{C}$ satisfies the *tilde-full condition* when $\mathscr{C}$ satisfies the following conditions.

- It is minimum-transversal-covered.
- $\mathscr{C}$ satisfies the integral blocking condition.
- Every hyperedge in $\tilde{\mathscr{C}}$ belongs to the support of some maximum fractional packing.

Lemma 6 *If a clutter $\mathscr{C}$ satisfies the tilde-full condition, $\mathscr{C}$ is a weak tilde-invariant clutter.*

Proof Assume that $\mathscr{C}$ satisfies the tilde-full condition. Then every hyperedge H in $\tilde{\mathscr{C}}$ belongs to the support of some maximum fractional packing y. By Lemma 4, y is also a maximum fractional packing of $\tilde{\mathscr{C}}$. On $\tilde{\mathscr{C}}$, H and any minimum solution $x_0 \in \{x \in \mathbb{R}^E | M(\tilde{\mathscr{C}})x \geq 1_E, x \geq 0\}$ in the dual problem satisfy $\langle 1_H, x_0 \rangle = 1$. Since any minimum transversal $B \in \mathrm{minb}(\tilde{\mathscr{C}})$ is a minimum solution, we have $|B \cap H| = 1$. Therefore $\tilde{\mathscr{C}}$ is a tilde-invariant clutter. □

Lemma 7 *When a clutter $\mathscr{C}$ satisfies the tilde-full condition, $\tilde{\mathscr{C}}$ also satisfies the tilde-full condition.*

Proof By Corollary 1, $\tilde{\mathscr{C}}$ satisfies the integral blocking condition. For every hyperedge $H \in \tilde{\mathscr{C}}$, there exists a maximum fractional packing y of $\mathscr{C}$ whose support contains H since $\mathscr{C}$ satisfies the tilde-full condition. Since the support of y is contained in $\tilde{\mathscr{C}}$ by Lemma 4, y is also a maximum fractional packing of $\tilde{\mathscr{C}}$. Therefore $\tilde{\mathscr{C}}$ satisfies the tilde-full condition. □

4.3.3 Polytope I($\mathscr{C}$)

We define a polyhedron $\mathrm{I}(\mathscr{C})$ as follows.

$$\mathrm{I}(\mathscr{C}) = \{x \in \mathbb{R}^E : \langle x, 1_D \rangle = 1 \text{ for all } D \in \mathrm{minb}(\mathscr{C}), \langle x, 1_D \rangle \geq 1 \text{ for all } D \in \mathrm{b}(\mathscr{C}), x \geq 0\}.$$

Note that this polyhedron $I(\mathscr{C})$ is a face of the blocking polyhedron $\{x \in \mathbb{R}^E : \langle x, 1_D \rangle \geq 1 \text{ for all } D \in \mathrm{b}(\mathscr{C}), x \geq 0\}$. This polyhedron plays a central role in this chapter.

Lemma 8 *For a clutter $\mathscr{C}$ which satisfies the integral blocking condition, $I(\mathscr{C})$ is a non-empty polyhedron. For a clutter $\mathscr{C}$ which satisfies the integral blocking condition, $I(\mathscr{C})$ is a polytope if and only if $\mathscr{C}$ is minimum-transversal-covered.*

Proof We show the first statement. Since $\mathscr{C}$ satisfies the integral blocking condition, there exists at least one maximum fractional packing whose support intersects every minimum transversal in exactly one element. Therefore $\tilde{C}$ contains some hyperedge H. Then $1_H \in \mathrm{I}(\mathscr{C})$.

We show the second statement.

Assume that $\mathscr{C}$ is minimum-transversal-covered. By Lemma 5, $\tilde{\mathscr{C}}$ is non-empty. Since the incidence vector of every hyperedge of $\tilde{\mathscr{C}}$ belongs to $\mathrm{I}(\mathscr{C})$, $\mathrm{I}(\mathscr{C})$ is non-empty.

Assume $x \in \mathrm{I}(\mathscr{C})$. For any $a \in E$, there exists $B \in \mathrm{minb}(\mathscr{C})$ with $a \in B$ since the minimum transversals cover E. Therefore $x(a)$ is at most 1 since $x \geq 0$ and $\langle x, 1_B \rangle = 1$. Since $x \geq 0$, $\mathrm{I}(\mathscr{C})$ is bounded.

Conversely, assume that there exists $a \in E$ which is covered with no minimum transversals. Since $\mathscr{C}$ satisfies the integral blocking condition, $\tilde{C}$ contains some hyperedge H. Then $1_H \in \mathrm{I}(\mathscr{C})$. $1_H + k1_a \in \mathrm{I}(\mathscr{C})$ for any $k \geq 0$. So $\mathrm{I}(\mathscr{C})$ is not bounded. □

Lemma 9 *For an ideal minimum-transversal-covered clutter $\mathscr{C}$, $I(\mathscr{C})$ is an integral polytope.*

Proof Since $\mathscr{C}$ is ideal, $\{x \in \mathbb{R}^E | \langle x, 1_B \rangle \geq 1 \text{ for any } B \in \mathrm{b}(\mathscr{C}), x \geq 0\}$ becomes an integral polyhedron. Since $\mathrm{I}(\mathscr{C})$ is a face of it, $\mathrm{I}(\mathscr{C})$ is also an integral polyhedron. Note that every face of an integral polyhedron is integral. $\mathrm{I}(\mathscr{C})$ is a polytope since its minimum transversals cover E and Lemma 8. □

Lemma 10 *For a minimum-transversal-covered clutter $\mathscr{C}$, the set of integral points in $I(\mathscr{C})$ coincides with the set of incidence vectors of $\tilde{\mathscr{C}}$. And hence every integral point in $I(\mathscr{C})$ is an integral extreme point of $I(\mathscr{C})$.*

Proof Every incidence vector of $\tilde{\mathscr{C}}$ satisfies all the inequalities defining $\mathrm{I}(\mathscr{C})$. Conversely, consider an integral point x in $\mathrm{I}(\mathscr{C})$. Note that such an integral point x is a 01-vector because the minimum transversals cover E. Let H be the set with $1_H = x$. We show that H is a hyperedge of $\tilde{\mathscr{C}}$. By the definition of $\mathrm{I}(\mathscr{C})$, H intersects every transversal and intersects every minimum transversal in exactly one element. Next we show that such H is minimal. If there exists another integral point $1_{H'}$ such that $H' \subsetneq H$. Then there exists $a \in H - H'$. Since the minimum transversals cover E, there exists a minimum transversal B containing a. But $B - a$ also intersects every hyperedge, which contradicts the minimality of a hyperedge. So such a set is a hyperedge in $\tilde{\mathscr{C}}$. And hence x is also an extreme point of the polytope $\mathrm{I}(\mathscr{C})$. □

Lemma 11 *For a minimum-transversal-covered clutter $\mathscr{C}$ which satisfies the integral blocking condition, the point consisting of all $1/bn(\mathscr{C})$ is contained in the relative interior of $I(\mathscr{C})$.*

Proof Since every transversal defining a facet of $\mathrm{I}(\mathscr{C})$ has a size of at least $\mathrm{bn}(\mathscr{C}) + 1$, the inner product of a transversal defining a facet of $\mathrm{I}(\mathscr{C})$ and the point in the statement is more than 1. On the other hand, the inner product of a minimum transversal of $\mathrm{I}(\mathscr{C})$ and the point in the statement is exactly 1. Therefore the point consisting of $1/\mathrm{bn}(\mathscr{C})$ is contained in the relative interior of $\mathrm{I}(\mathscr{C})$. □

Lemma 12 *Assume that a clutter $\mathscr{C}$ which satisfies the integral blocking condition. When $I(\mathscr{C})$ is an integral polytope, $\mathscr{C}$ satisfies the tilde-full condition.*

Proof Since $\mathrm{I}(\mathscr{C})$ is a polytope, $\mathscr{C}$ is minimum-transversal-covered by Lemma 8. By the integrality of the polytope $\mathrm{I}(\mathscr{C})$ and Lemma 10, there exists no extreme point other than such incidence vectors of $\tilde{\mathscr{C}}$. Therefore the point in Lemma 11 is expressed as a positive combination of the extreme points of $\mathrm{I}(\mathscr{C})$ because the point is in the relative interior of $\mathrm{I}(\mathscr{C})$. By multiplying such a coefficient by $\mathrm{bn}(\mathscr{C})$, the sum of all the components of the vector attains $\mathrm{bn}(\mathscr{C})$, which is the fractional packing number. So there is a maximum fractional packing such that all the coefficients are positive. □

Theorem 1 *For an ideal minimum-transversal-covered clutter $\mathscr{C}$, $I(\tilde{\mathscr{C}})$ is an integral polytope with $I(\mathscr{C}) = I(\tilde{\mathscr{C}})$. The extreme points of $I(\mathscr{C})$ consist of the incidence vectors of $\tilde{\mathscr{C}}$.*

Proof $\mathrm{I}(\mathscr{C})$ is an integral polytope by Lemma 9. By Corollaries 1 and 2, $\mathscr{C}$ and $\tilde{\mathscr{C}}$ are minimum-transversal-covered and satisfy the integral blocking condition. Since $\tilde{\mathscr{C}} \subseteq \mathscr{C}$, there exists $B' \in \mathrm{b}(\tilde{\mathscr{C}})$ with $B' \subseteq B$ for any $B \in \mathrm{b}(\mathscr{C})$. Therefore $I(\tilde{\mathscr{C}}) \subseteq I(\mathscr{C})$ holds. Therefore $\mathrm{I}(\mathscr{C}) = \mathrm{I}(\tilde{\mathscr{C}})$ follows from Lemma 10. □

Definition 5 A clutter $\mathscr{C}$ satisfies the *dimension condition* if

$$(\text{affine dimension of } \tilde{\mathscr{C}}) + (\text{affine dimension of } \mathrm{minb}(\tilde{\mathscr{C}})) = |E| - 1.$$

The affine dimension of $\tilde{\mathscr{C}}$ means the dimension of the affine hull of all the incidence vectors of $\tilde{\mathscr{C}}$.

Lemma 13 *For a tilde-invariant clutter $\mathscr{C} = \tilde{\mathscr{C}}$ such that $I(\mathscr{C})$ is an integral polytope, $\mathscr{C}$ satisfies the dimension condition.*

Proof Since $\mathrm{I}(\mathscr{C})$ is a polytope, $\mathscr{C}$ is minimum-transversal-covered by Lemma 8. Since $\mathrm{I}(\mathscr{C})$ is an integral polytope, the extreme points of $\mathrm{I}(\mathscr{C})$ consist of the incidence vectors of $\tilde{\mathscr{C}}$ by Lemma 10. So the dimension of $\mathrm{I}(\mathscr{C})$ is equal to the dimension of the affine hull of $\tilde{\mathscr{C}}$. We have only to show that the dimension of $I(\mathscr{C})$ is not affected by transversals other than $\mathrm{minb}(\tilde{\mathscr{C}})$. If the dimension of $I(\mathscr{C})$ is affected by a non-minimum transversal, such a non-minimum transversal intersects every hyperedge in $\mathscr{C}$ in exactly one element. For a maximum fractional packing y, $\sum_{H \in \tilde{\mathscr{C}}} y(H) = \sum_{H \in \tilde{\mathscr{C}}} y(H)\langle 1_H, 1_B\rangle = \sum_{H \in \tilde{\mathscr{C}}} y(H)\langle 1_H, 1_B\rangle = \langle 1_E, 1_B\rangle = |B| > \mathrm{bn}(\mathscr{C})$, a contradiction to Lemma 4. □

4.3.4 Non-separability

Separability is a necessary condition for a clutter $\mathscr{C}$ to have an ideal minimally non-packing clutter $\mathscr{D}$ with $\mathscr{C} = \tilde{\mathscr{D}}$.

Definition 6 A clutter $\mathscr{C}$ is *separable* if there exists a nontrivial partition $\{E_1, E_2\}$ of E and $\mathrm{bn}(\mathscr{C}) = \mathrm{bn}(\mathscr{C}[E_1]) + \mathrm{bn}(\mathscr{C}[E_2])$ where $\mathscr{C}[E_1] := \mathscr{C} \backslash E_1^c$ with $\mathscr{C}[E_2] := \mathscr{C} \backslash E_2^c$. Otherwise it is called *non-separable*.

Lemma 14 *A minimally non-packing clutter with respect to deletion is non-separable.*

Proof Since a clutter $\mathscr{C}$ is minimally non-packing, $\mathscr{C}$ does not pack. Assume that it is separable with a partition $\{E_1, E_2\}$ of E.

Consider the case where both $\mathscr{C}[E_1]$ and $\mathscr{C}[E_2]$ pack. Then $\mathrm{fpn}(\mathscr{C}[E_1]) = \mathrm{pn}(\mathscr{C}[E_1]) = \mathrm{bn}(\mathscr{C}[E_1])$ and $\mathrm{fpn}(\mathscr{C}[E_2]) = \mathrm{pn}(\mathscr{C}[E_2]) = \mathrm{bn}(\mathscr{C}[E_2])$ and $\mathrm{fpn}(\mathscr{C}) \geq \mathrm{fpn}(\mathscr{C}[E_1]) + \mathrm{fpn}(\mathscr{C}[E_2])$. By separability, $\mathrm{pn}(\mathscr{C}) \geq \mathrm{pn}(\mathscr{C}[E_1]) + \mathrm{pn}(\mathscr{C}[E_2]) = \mathrm{bn}(\mathscr{C}[E_1]) + \mathrm{bn}(\mathscr{C}[E_2]) = \mathrm{bn}(\mathscr{C})$. Since $\mathrm{pn}(\mathscr{C}) \leq \mathrm{bn}(\mathscr{C})$ generally, $\mathrm{pn}(\mathscr{C}) = \mathrm{bn}(\mathscr{C})$ holds. So $\mathscr{C}$ packs. This contradicts the fact that $\mathscr{C}$ does not pack.

Consider the case where either of them does not pack. This contradicts the assumption that $\mathscr{C}$ is minimally non-packing. □

Lemma 15 *Assume that a minimum-transversal-covered clutter $\mathscr{C}$ satisfies the integral blocking condition and non-separability. Then $\tilde{\mathscr{C}}$ is non-separable.*

Proof Assume that $\tilde{\mathscr{C}}$ is separable with a partition $\{E_1, E_2\}$ such that $\mathrm{bn}(\tilde{\mathscr{C}}[E_1]) + \mathrm{bn}(\tilde{\mathscr{C}}[E_2]) = \mathrm{bn}(\tilde{\mathscr{C}})$. We have $\mathrm{bn}(\tilde{\mathscr{C}}[E_1]) \leq \mathrm{bn}(\mathscr{C}[E_1])$ and $\mathrm{bn}(\tilde{\mathscr{C}}[E_2]) \leq \mathrm{bn}(\mathscr{C}[E_2])$ since $\tilde{\mathscr{C}}[E_1] \subseteq \mathscr{C}[E_1]$ and $\tilde{\mathscr{C}}[E_2] \subseteq \mathscr{C}[E_2]$. By Lemma 4 and the integral blocking condition on $\mathscr{C}$, we have $\mathrm{bn}(\mathscr{C}) = \mathrm{bn}(\tilde{\mathscr{C}})$. Since $\mathrm{bn}(\mathscr{C}[E_1]) + \mathrm{bn}(\mathscr{C}[E_2]) \leq \mathrm{bn}(\mathscr{C})$ generally, we have $\mathrm{bn}(\tilde{\mathscr{C}}) = \mathrm{bn}(\tilde{\mathscr{C}}[E_1]) + \mathrm{bn}(\tilde{\mathscr{C}}[E_2]) \leq \mathrm{bn}(\mathscr{C}[E_1]) + \mathrm{bn}(\mathscr{C}[E_2]) \leq \mathrm{bn}(\mathscr{C})$. Therefore we have $\mathrm{bn}(\mathscr{C}[E_1]) + \mathrm{bn}(\mathscr{C}[E_2]) = \mathrm{bn}(\mathscr{C})$, which contradicts the fact that $\mathscr{C}$ is non-separable. □

Definition 7 A minimum-transversal-covered clutter $\mathscr{C}$ satisfying the integral blocking condition is *hyperedge-separable* if there exists a hyperedge $H \in \tilde{\mathscr{C}}$ such that $\mathrm{bn}(\mathscr{C} \backslash H) = \mathrm{bn}(\mathscr{C}) - 1$. Otherwise, that is, if $\mathrm{bn}(\mathscr{C} \backslash H) < \mathrm{bn}(\mathscr{C}) - 1$ holds for every hyperedge $H \in \tilde{\mathscr{C}}$, the clutter $\mathscr{C}$ is called *hyperedge-non-separable*.

Lemma 16 *When a minimum-transversal-covered clutter $\mathscr{C}$ satisfying the integral blocking condition is hyperedge-separable, it is separable.*

Proof When a clutter $\mathscr{C}$ is hyperedge-separable at $H \in \tilde{\mathscr{C}}$, we take a partition $\{H, H^c\}$ of E. Then we have $\mathrm{bn}(\mathscr{C}[H]) + \mathrm{bn}(\mathscr{C} \backslash H) = \mathrm{bn}(\mathscr{C})$ because of $\mathrm{bn}(\mathscr{C}[H]) = 1$. □

4.3.5 Summarizing the Conditions in Step 1

Definition 8 For a clutter $\mathscr{C}$, a clutter $\mathscr{D}$ is called a *solution clutter* of $\mathscr{C}$ if $\mathscr{C} = \tilde{\mathscr{D}}$.

As a weaker problem, we first consider the problem for a clutter $\mathscr{C}$ to have an ideal clutter $\mathscr{D}$ with $\mathscr{C} = \tilde{\mathscr{D}}$. That is, we discard the condition of "minimally non-packing".

We have considered necessary conditions for the tilde-invariant clutter $\mathscr{C}$ to have an ideal minimally non-packing clutter $\mathscr{D}$ with $\mathscr{C} = \tilde{\mathscr{D}}$. In this subsection, we integrate these results.

Theorem 2 *Assume that a clutter $\mathscr{C}$ has an ideal minimally non-packing solution $\mathscr{D}$. Then $\mathscr{C}$ satisfies the following conditions.*

- *$\mathscr{C}$ satisfies the integral blocking condition.*
- *$I(\mathscr{C})$ is an integral polytope.*
- *$\mathscr{C}$ is non-separable.*

Proof The clutter $\mathscr{C}$ is minimum-transversal-covered by Lemma 1. The clutter $\mathscr{C}$ is integral blocking by Corollary 2. So the integrality of $I(\mathscr{C})$ follows from Theorem 1. The non-separability condition follows from Lemmas 14 and 15. □

In other words, for any ideal minimally non-packing solution $\mathscr{D}$, $\tilde{\mathscr{D}}$ satisfies the conditions in Theorem 2.

We call a clutter $\mathscr{C} = \tilde{\mathscr{C}}$ satisfying the conditions in Theorem 2 a *precore clutter*. When we consider Conjecture 2, we have only to consider the precore clutters. We give an example of a precore clutter in Sect. 4.5.2.

Theorem 3 *Assume that $\mathscr{C}$ satisfies the integral blocking condition and $I(\mathscr{C})$ is an integral polytope. Then $\tilde{\mathscr{C}}$ is minimum-transversal-covered and tilde-invariant, and satisfies the tilde-full condition, and the dimension condition.*

Proof The clutter $\tilde{\mathscr{C}}$ is minimum-transversal-covered by Lemma 8. The integral blocking condition follows from Corollary 2. The clutter $\tilde{\mathscr{C}}$ satisfies the tilde-full condition by Lemmas 12 and 7. So $\tilde{\mathscr{C}}$ is a tilde-invariant clutter by Lemma 6. The dimension condition follows from Lemma 13. □

4.4 Conditions in the Second Step

After we find a clutter $\mathscr{C}$ satisfying the conditions in step 1, we have to discuss whether it has an ideal clutter $\mathscr{D}$ which is minimally non-packing with $\mathscr{C} = \tilde{\mathscr{D}}$ further.

For a precore clutter $\mathscr{C}$, we discuss necessary conditions for an ideal solution clutter $\mathscr{D}$. The difference between the conditions in Theorem 2 and those in this section is whether $\mathscr{D}$ appears in the conditions directly or not.

Condition I: $\mathrm{I}(\mathscr{C}) = \mathrm{I}(\mathscr{D})$ holds.

We can divide Condition I into Conditions IM and IF.

Condition IM: The affine space generated by the incidence vectors of the minimum transversals of $\mathscr{C}$ is equal to be the affine space generated by the incidence vectors of the minimum transversals of $\mathscr{D}$.

Condition IF: If a facet F of $\mathrm{I}(\mathscr{C})$ is defined by a transversal of $\mathscr{C}$, there exists at least one element $B \in \mathrm{b}(\mathscr{D})$ defining the facet F. Moreover, every $B \in \mathrm{b}(\mathscr{D})$ intersects every $H \in \mathscr{C}$.

Condition H: $\mathscr{C} \subseteq \mathscr{D}$ must hold. For any $H \in \mathscr{D} - \mathscr{C}$, there exists $B \in \mathrm{minb}(\mathscr{C})$ with $|H \cap B| \geq 2$.

Theorem 4 *For a precore clutter $\mathscr{C}$, every ideal solution clutter $\mathscr{D}$ to $\mathscr{C}$ satisfies Conditions IM, IF, and H.*

Proof By Theorem 1, Condition I holds. Since the affine hull of $I(\mathscr{C})$ is defined by $\mathrm{minb}(\mathscr{C})$, Condition IM holds. Since the facets of $I(\mathscr{C})$ are defined by $\mathrm{b}(\mathscr{C})$ and $x \geq 0$, Condition IF holds. Note that every $B \in \mathrm{b}(\mathscr{D})$ intersects every $H \in \mathscr{C}$ since $\mathscr{C} \subseteq \mathscr{D}$.

Assume $H \in \mathscr{D} - \mathscr{C}$. Since $H \notin \tilde{\mathscr{C}} = \mathscr{C}$, there exists $B \in \mathrm{minb}(\mathscr{C})$ with $|H \cap B| \geq 2$. Therefore Condition H holds. □

We consider necessary conditions for the tilde-invariant clutter $\mathscr{C}$ to have an ideal minimally non-packing solution clutter $\mathscr{D}$.

Condition B: For any disjoint sets $A, B \subseteq E$ with $A \cup B \neq \emptyset$, $\mathrm{bn}(\mathscr{C}/A\backslash B) \leq \mathrm{pn}(\mathscr{D}/A\backslash B)$ holds.

Theorem 5 *For a precore clutter $\mathscr{C}$, every ideal minimally non-packing solution clutter $\mathscr{D}$ to $\mathscr{C}$ satisfies Condition B.*

Proof Since every proper minor of $\mathscr{D}$ has the packing property, $\mathrm{bn}(\mathscr{D}/A\backslash B) = \mathrm{pn}(\mathscr{D}/A\backslash B)$ holds. Since $\mathscr{C} \subseteq \mathscr{D}$, $\mathrm{bn}(\mathscr{C}/A\backslash B) \leq \mathrm{bn}(\mathscr{D}/A\backslash B)$. Therefore Condition B holds. □

Corollary 3 *When a precore clutter $\mathscr{C}$ has an ideal minimally non-packing solution, there must exist a clutter $\mathscr{D}$ satisfying Conditions IF, IM, H, and B.*

We have not found a precore clutter $\mathscr{C}$ with $\mathscr{D}$ satisfying the above conditions yet. If we can prove that there exist no such precore clutters, then Conjecture 2 will be affirmative. Actually, the conditions in Corollary 3 are effectively used in Sect. 4.5.2.

4.5 Unique Maximum Fractional Packing

In this section, we consider the problem under an additional condition that the maximum fractional packing is unique. Many important classes of precore clutters satisfy this condition. Section 4.5.1 is concerned with step 1 in the case that the maximum

fractional packing is unique. In Sect. 4.5.2, we consider an example of a precore clutter. Moreover we show that there exists no counterexample to Conjecture 2 in that class (Theorem 8).

4.5.1 Unique Maximum Fractional Packing

In this subsection, we consider a clutter which has a unique maximum fractional packing. For example, the clutter $Q_6 = \{abc, cde, efa, bdf\}$ has a unique maximum fractional packing.

Lemma 17 *Consider a clutter $\mathscr{C}$ which satisfies the tilde-full condition. The incidence vectors of $\tilde{\mathscr{C}}$ are affinely independent if and only if its maximum fractional packing is unique.*

Proof Assume that a maximum fractional packing is unique. Then $y(H) > 0$ for all $H \in \tilde{\mathscr{C}}$ by the tilde-full condition. The support of the maximum fractional packing y of $\mathscr{C}$ consists of the hyperedges of $\tilde{\mathscr{C}}$ by Lemma 3. When these incidence vectors are affinely dependent, the maximum fractional packing can be moved slightly so that it is still a maximum fractional packing, a contradiction.

When a maximum fractional packing is not unique, by taking two maximum fractional packings y_1 and y_2, they are affinely dependent since $\sum_{H\in\mathscr{C}} y_1(H) = \sum_{H\in\mathscr{C}} y_2(H)$ and $\sum_{H\in\mathscr{C}} y_1(H)1_H = \sum_{H\in\mathscr{C}} y_2(H)1_H = 1_E$ by Lemma 4. □

Generally, when a polyhedron P is not full dimensional, its facet-defining inequality is not unique. Here, we call a linear inequality $\langle 1_B, x\rangle \geq 0$ a *facet-defining inequality* of P to a facet F when $\{x \in P | \langle 1_B, x\rangle = 0\} = F$.

Lemma 18 *Consider a clutter $\mathscr{C}$ such that $I(\mathscr{C})$ satisfies the integral blocking condition and is an integral polytope. Its maximum fractional packing is unique if and only if $I(\mathscr{C})$ is a simplex.*

Proof Since $I(\mathscr{C})$ is a polytope, the minimum transversals of $\mathscr{C}$ cover E by Lemma 8. By Lemma 12, $\mathscr{C}$ satisfies the tilde-full condition. By Lemma 17, its maximum fractional packing is unique if and only if the incidence vectors of $\tilde{\mathscr{C}}$ are affinely independent. □

We call an integral polytope which is simplex an *integral simplex*. For $H \in \tilde{\mathscr{C}}$, we call a transversal $B \in b(\tilde{\mathscr{C}})$ a *facet transversal* of H if $|H \cap B| > 1$ and $|H' \cap B| = 1$ for $H' \in \tilde{\mathscr{C}} - \{H\}$.

Theorem 6 *Let $\mathscr{C}$ be a minimum-transversal-covered tilde-invariant clutter which satisfies the integral blocking condition. The polytope $I(\mathscr{C})$ is an integral simplex and the clutter $\mathscr{C}$ is hyperedge-non-separable if and only if $\mathscr{C}$ satisfies the dimension condition and, for each hyperedge of H of $\mathscr{C}(= \tilde{\mathscr{C}})$, there exists a facet transversal $B \in b(\mathscr{C})$ of H.*

Proof First, let us assume that the clutter $\mathscr{C}$ is tilde-invariant and hyperedge-non-separable and that the polytope $\mathrm{I}(\mathscr{C})$ is an integral simplex. From Lemma 10, we have that, for every extreme point x of the simplex $\mathrm{I}(\mathscr{C})$, there exists a hyperedge H of the clutter $\tilde{\mathscr{C}}(=\mathscr{C})$ such that $x = 1_H$ holds. Let F_H be the unique facet of the simplex $\mathrm{I}(\mathscr{C})$ which does not contain 1_H. If there exists a transversal $B \in \mathrm{b}(\mathscr{C})$ defining F_H, then, by definition, it will be a facet transversal of H. On the contrary, suppose that there exists no facet transversal $B \in \mathrm{b}(\mathscr{C})$ defining F_H. Then the facet F_H must be defined by a linear inequality of a nonnegative constraint $x(a) \geq 0$ for some element a of E. And hence every hyperedge in $\mathscr{C}$ except for the hyperedge H satisfies $x(a) = 0$, that is, it does not contain the element a. The point 1_E is attained by the nonnegative combination of $\mathscr{C}$ by Lemma 4. Therefore when 1_E is represented as a nonnegative combination y of $\mathscr{C}$, the coefficient $y(H)$ to H is 1. Since the deletion of all the elements reduces the fractional packing number by exactly one, it is hyperedge-separable. Therefore any facet-defining inequality is defined by a facet transversal. The dimension condition follows from Lemma 13.

Next, suppose conversely that, for every hyperedge H of the clutter $\mathscr{C}$, there exists a transversal $B \in \mathrm{b}(\mathscr{C})$ such that $|H \cap B| > 1$ holds and that $|H' \cap B| = 1$ holds for every $H' \in \mathscr{C} - \{H\}$. Then the incidence vector 1_H of every hyperedge H in $\mathscr{C}(=\tilde{\mathscr{C}})$ is an extreme point of the polytope $\mathrm{I}(\mathscr{C})$ by Lemma 10. For each $1_H \in \mathrm{I}(\mathscr{C})$, the facet $\langle 1_B, x\rangle = 1$ contains all the integral extreme points except 1_H. And hence $\mathrm{I}(\mathscr{C})$ has a $(|\mathscr{C}| - 1)$-dimensional simplicial face F whose extreme points coincide with the incidence vectors of the hyperedges of the clutter $\mathscr{C}(=\tilde{\mathscr{C}})$. By the dimension condition, the dimension of $\mathrm{I}(\mathscr{C})$ is the size $|\mathscr{C}| - 1$. Therefore the simplicial face F is coincident with the polytope $\mathrm{I}(\mathscr{C})$ itself. Since its extreme points are expressed as $\mathscr{C}$, it is an integral simplex. Since $|H \cap B| \geq 2$ holds, there exists a hyperedge X of $\mathscr{C} - \{H\}$ such that $X \cap (H \cap B) \neq \emptyset$ holds. And hence the clutter $(\mathscr{C} \setminus H) \cup \{H\}$ cannot contain the hyperedge X. Since the clutter $\mathscr{C}$ is tilde-invariant, the hyperedge X is also a hyperedge of $\tilde{\mathscr{C}}$ and hence the incidence vector 1_X of X forms an extreme point of the integral simplex $\mathrm{I}(\mathscr{C})$. On the other hand, from Lemma 18, we have that the clutter $\mathscr{C}$ has a unique maximum fractional packing and that it inevitably uses the incidence vector 1_X of X as an element of its convex combination. Thus, for every hyperedge H of the clutter $\mathscr{C}$, we have that the unique maximum fractional packing of $\mathscr{C}$ is different from any maximum fractional packing of the clutter $(\mathscr{C} \setminus H) \cup \{H\}$, which means that the clutter $\mathscr{C}$ is hyperedge-non-separable. □

We give an example of a precore clutter. A graph G is called a *brick* if it is 3-connected and $G - \{u, v\}$ has a perfect matching for all pairs of distinct $u, v \in V(G)$. For a graph G, the *vertex cut clutter* $\mathscr{C}(G)$ is the clutter $\{\{\{a, b\} \in E(G)\} | a \in V(G)\}$.

Example 3 For a brick G, the vertex cut clutter $\mathscr{C}(G)$ of G is an example of a precore clutter. In fact, since G is non-bipartite, the maximum fractional packing is unique and the integral blocking condition is satisfied. Since G is matching covered, $\mathscr{C}(G)$ satisfies the minimum-transversal-covered. Moreover, $\mathscr{C}(G) = \widetilde{\mathscr{C}(G)}$ holds. Since the dimension of the matching polytope of a brick G is $|E(G)| - |V(G)|$, the dimension condition is satisfied. For any vertex $x \in V(G)$, there exists a factor of a

brick G such that vertex x has degree 3 and the other vertices have degree 1. Such a factor becomes a facet transversal. Therefore by Theorem 6, $I(\mathscr{C}(G))$ is an integral simplex and hyperedge-non-separable. We can show that $\mathscr{C}(G)$ is also non-separable by the definition of a brick.

4.5.2 Combinatorial Affine Planes

A clutter $\mathscr{C}$ on a finite set E is called a *combinatorial projective plane* if the following three conditions are satisfied.

(1) For any two distinct elements, there exists a unique hyperedge containing the two elements.

(2) Any two distinct hyperedges intersect in exactly one element.

(3) There are four elements such that no hyperedge contains more than two of them.

On a combinatorial projective plane $\mathscr{C}$, each hyperedge is also called a *point* and each element is also called a *line*. (The inverse correspondence between point and line is possible but the reason why we adopt this correspondence is due to a combinatorial affine plane appeared later.) For any combinatorial projective plane $\mathscr{C}$, there exists a natural number n such that $n^2 + n + 1 = |\mathscr{C}| = |E|$. Every element $a \in E$ is contained in $(n + 1)$ hyperedges, and every hyperedge has size $n + 1$.

By deleting one element $a \in E$ from the clutter $\mathscr{C}$, we obtain another clutter $\mathscr{C}\backslash a$. This clutter $\mathscr{C}\backslash a$ becomes a clutter of a combinatorial affine plane.

Definition 9 A clutter $\mathscr{C}$ on a finite set E is called a *combinatorial affine plane* if the following three conditions are satisfied.

(1) For any two distinct hyperedges H and H', $|H \cap H'| = 1$.

(2) Given an element a and a hyperedge $H \in \mathscr{C}$ with $a \notin H$, there exists a unique element $b \in H$ such that a and b are not contained in the same hyperedge.

(3) There exist three hyperedges which do not contain the same element.

For example, a combinatorial projective plane on seven elements induces a combinatorial affine plane on six elements. It is Q_6, which is an ideal clutter of blocking number 2.

The following proposition is folklore (for example, see [10]).

Proposition 2 *Let the size of a hyperedge of a combinatorial affine plane be $n + 1$. Then $|E| = n^2 + n$, the size of its minimum transversal is n, the number of its minimum transversals is $n + 1$. Each element is contained in exactly n hyperedges. Its minimum transversals form a partition of E. Any two elements which belong to different minimum transversals are included in exactly one hyperedge. Each hyperedge is also a transversal.*

Lemma 19 *A combinatorial affine plane $\mathscr{C}$ satisfies the integral blocking condition, and $\mathscr{C} = \tilde{\mathscr{C}}$ holds. Therefore it is tilde-invariant.*

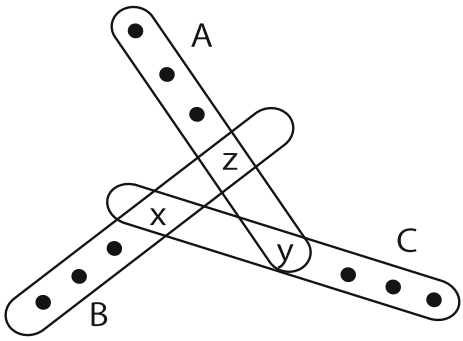


Fig. 4.2 An example of $\mathscr{D}[X]$

Proof Let $n+1$ be the size of its hyperedge. We first show the integral blocking condition. The blocking number of $\mathscr{C}$ is n. For each element in E, there exist n hyperedges of $\mathscr{C}$ containing the element. Therefore the sum of the incidence vectors of all the hyperedges of $\mathscr{C}$ is $n1_E$, which is a fractional packing of $\mathscr{C}$. Since they form a maximum fractional packing, $\mathscr{C}$ satisfies the integral blocking condition.

Since every minimum transversal and every hyperedge of $\mathscr{C}$ intersect in exactly one element by Proposition 2, we have $\mathscr{C} = \check{\mathscr{C}}$. □

Lemma 20 *The maximum fractional packing of a combinatorial affine plane is unique.*

Proof By calculating the determinant of the clutter matrix, the incidence vectors of hyperedges of the clutter are affinely independent. So the statement follows from Lemma 17. □

Theorem 7 *For a combinatorial affine plane $\mathscr{C}$, $I(\mathscr{C})$ is an integral simplex and $\mathscr{C}$ is non-separable. Therefore $\mathscr{C}$ is a precore clutter.*

Proof We can take the hyperedges on $\mathscr{C}$ as facet transversals by Proposition 2. The number of the hyperedges is n^2 and the number of the minimum transversals is $n+1$. Since they are affinely independent, we have the dimension condition. By Theorem 6, $I(\mathscr{C})$ is an integral simplex.

By deleting all the points on a hyperedge, all the hyperedges disappear. So the clutter of a combinatorial affine plane is non-separable. □

Theorem 8 *Every combinatorial affine plane $\mathscr{C}$ of blocking number at least 3 has no ideal minimally non-packing solution clutter.*

Proof Assume that $\mathscr{C}$ has an ideal minimally non-packing solution clutter $\mathscr{D}$.

Consider distinct hyperedges A, B, and C in $\mathscr{C}$ with $A \cap B \cap C = \emptyset$. Let z be a unique point in $A \cap B$. Similarly, let x be a unique point in $B \cap C$, and y be a unique point in $C \cap A$ (Fig. 4.2).

Then consider the restriction $\mathscr{C}[X]$ where X is the union of the three hyperedges A, B, and C. Note that $X \neq E$ since the blocking number is at least 3. Then since

such a clutter has exactly three hyperedges, its blocking number is 2. Since $\mathscr{D}[X]$ must pack, there exists a packing of size 2 in $\mathscr{D}[X]$ (Condition B). By Condition IF, any facet transversal of $I(\mathscr{C})$ is also a facet transversal of $I(\mathscr{D})$. Therefore any hyperedge $H \in \mathscr{C}$ is also a transversal in $\mathscr{D}$. Moreover any minimum transversal of $\mathscr{C}$ is a minimum transversal of $\mathscr{D}$ by Condition IM. Therefore the two elements consisting of x and any one element of $A - \{y, z\}$ form a transversal of $\mathscr{D}[X]$. Similarly, two elements consisting of y and any one element of $B - \{z, x\}$ form a transversal, and two elements consisting of z and any one element of $C - \{x, y\}$ form a transversal. Such two elements are included in some minimum transversal or included in some hyperedge which is also a transversal in $\mathrm{b}(\mathscr{D})$ since the deletion of elements from a clutter corresponds to the contraction of them from the clutter of transversals. Therefore X is covered by transversals of size 2. By regarding such two elements as an edge of a graph, such a graph has three connected components and each of them is a star. A packing of size 2 becomes a partition on X consisting of two hyperedges of size $|X|/2$ as in Fig. 4.3. For a packing of size 2 in $\mathscr{D}[X]$, two elements as a transversal belong to different hyperedges in the packing of size 2 on $\mathscr{D}[X]$. Therefore we can take four types of packings of size 2.

In three types out of the four types of packings, one hyperedge in packings of size 2 is either of A, B, and C, the other hyperedge is included in the complement of the hyperedge in X. These cases contradict the fact that A, B, C themselves are transversals because every hyperedge must intersect every transversal. We discuss the remaining type of the packings, that is, one hyperedge is included in $\{x, y, z\}$, and other hyperedge is included in $X - \{x, y, z\}$. Since $\{x, y, z\}$ intersects any minimum transversal in exactly one element, $\{x, y, z\}$ cannot be a hyperedge of $\mathscr{D}$ by Condition H, a contradiction. □

We should note that whether a combinatorial projective plane except for the F_7 can be a core of a minimally non-ideal clutter or not is a famous open question of the theory of minimally non-ideal clutters (see Question 6 in Cornuéjols, Guenin and Tunçel [5]). The following conjecture asserts that, except for the F_7, there is no combinatorial projective plane which is a core of some minimally non-ideal clutter:

Conjecture 3 *A clutter of a combinatorial affine plane of blocking number at least 3 has no ideal solution clutter.*

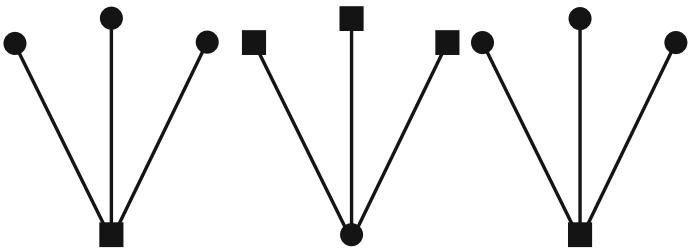


Fig. 4.3 Sets of vertices indicated by circles and squares represent hyperedges

Acknowledgements The second author's research is supported by Grant-in-Aid for Scientific Research (C) (26400185).

References

1. Conforti, M., Cornuéjols, G.: Clutters that pack and the max flow min cut property: a conjecture. In: Pulleyblank, W.R., Shepherd, F.B. (eds.) The Fourth Bellairs Workshop on Combinatorial Optimization (1993)
2. Cornuéjols, G.: Combinatorial Optimization - Packing and Covering. CBMS-NFS Regional Conference Series in Applied Mathematics. SIAM, Philadelphia (2001)
3. Cornuéjols, G., Guenin, B.: On Dijoins. Discrete Math. **243**, 213–216 (2002)
4. Cornuéjols, G., Guenin, B., Margot, F.: The packing property. Math. Program. Ser. A **89**, 113–126 (2000)
5. Cornuéjols, G., Guenin, B., Tuncel, L.: Lehman matrices. J. Comb. Theory Ser. B **99**, 531–556 (2009)
6. Ding, G., Zang, W.: Packing cycles in graphs. J. Comb. Theory Ser. B **86**, 381–407 (2002)
7. Guenin, B.: On Packing and Covering Polyhedra, Ph.D. Dissertation, Carnegie Mellon University, Pittsburgh, PA (1998)
8. Guenin, B.: Circuit Mengerian directed graphs. In: Aardal, K., Gerards, B. (eds.) Integer Programming and Combinatorial Optimization. Proceedings 8th International IPCO Conference, Utrecht. Lecture Notes in Computer Science, vol. 2081, pp. 185–195 (2001)
9. Hachimori, M., Nakamura, M.: Rooted circuits of closed-set systems and the max-flow min-cut property. Discrete Math. **308**, 1674–1689 (2008)
10. Hall, M.: Projective planes. T. Am. Math. Soc. **54**, 229–277 (1943)
11. Kashiwabara, K., Sakuma, T.: The positive circuits of oriented matroids with the packing property or idealness. Electron. Notes Discrete Math. **36**, 287–294 (2010)
12. Lehman, A.: On the width-length inequality. Math. Program. **17**, 403–417 (1979)
13. Martinez-Bernal, J., O'Shea, E., Villarreal, R.: Ehrhart clutters: regularity and max-flow min-cut. Electron. J. Comb. **17**, # R52 (2010)
14. Schrijver, A.: A counterexample to a conjecture of Edmonds and Giles. Discrete Math. **32**, 213–214 (1980)
15. Seymour, P.D.: The Matroids with the max-flow min-cut property. J. Comb. Theory Ser. B **23**, 189–222 (1977)
16. Williams, A.M., Guenin, B.: Advances in packing directed joins. Electron. Notes Discrete Math. **19**, 249–255 (2005)

Chapter 5
Symmetric Travelling Salesman Problem

Some New Algorithmic Possibilities

Tiru Arthanari and Kun Qian

5.1 Introduction

Starting from humble beginnings in a travelling manual in 1832 [51], the Travelling Salesman Problem (TSP) has grown to become the iconic problem in combinatorial optimization [20]. The difficulty of the TSP was first brought to the attention of the mathematics community by Austrian mathematician Karl Menger in 1930 [20].

The underlying objective of the TSP is to find an optimal tour that visits every node in a finite set of nodes and returns to the origin node on a graph, given the matrix of distances between any two nodes. It is easy to define yet hard to find an optimal solution [35]. There are a few variations of the TSP, such as the standard Asymmetrical Travelling Salesman Problem (ATSP) [46]; the time-dependent travelling salesman problem [27, 28]. In this chapter, our focus is on the Symmetric Travelling Salesman Problem (STSP) where the distance between two nodes is the same in either direction.

In 1954, George Dantzig, Ray Fulkerson and Selmer Johnson from the RAND Corporation presented a proof for the optimal path they had found for a tour through 49 cities in the United States of America (USA) [22]. They used a combination of linear programming and some heuristics to calculate the shortest tour manually for the 12,345-mile tour through 49 cities.

Initially, most of the work done on the TSP is motivated by the wide range of applicability of TSP algorithms on other discrete optimisation problems [20]. Over time the algorithms developed for the TSP have been used in industry to solve many

T. Arthanari (✉) · K. Qian
Department of ISOM, Faculty of Business and Economics, University of Auckland, Auckland, New Zealand
e-mail: t.arthanari@auckland.ac.nz

K. Qian
e-mail: kqia040@aucklanduni.ac.nz

S. K. Neogy et al. (eds.), *Mathematical Programming and Game Theory*,
Indian Statistical Institute Series, https://doi.org/10.1007/978-981-13-3059-9_5

practical problems. Some of the problems include the following: vehicle routing [24], genome mapping [1], guiding industrial machines [20] and organizing data [38].

Although there had been some progress in solving specific instances of the TSP problem, researchers were starting to wonder whether or not there exists an efficient algorithm to solve the TSP problem. Jack Edmonds famously stated in 1967 that he 'conjectures that there is no good algorithm for the travelling salesman problem' [24]. The 'good algorithm' Edmonds was referring to is an algorithm where the time taken to solve the problem will increase at an acceptable rate and not grow exponentially as the problem size grows. The nature of the TSP, however, requires the brute force algorithm to compare and rank all the combinations of possible paths through the nodes, and therefore, n nodes in an instance will require $(n-1)!$ operations, taking exponential time with respect to the size of the problem.

Both exact methods and heuristic algorithms have been developed to find solutions to the TSP. Exact methods include dynamic programming algorithms like the Held Karp algorithm [32]; the branch and bound algorithm by Little et al. [41] and polyhedral approaches [39] like the branch and cut, used in Concorde—the program that solved the largest $STSP$ problem to date having 85900 cities [2]. Heuristics include algorithms such as Christofides algorithm [18], Lin–Kernighan algorithm [40] and Lin–Kernighan–Helsgaun algorithm [34]. Metaheuristics are heuristic methods for developing heuristics to solve general problems. Some examples of metaheuristics are as follows: local search and hill climbing; simulated annealing [36]; genetic algorithm [45].

The research of this chapter departs from the standard formulation Dantzig, Fulkerson and Johnson [22] and other formulations of the $STSP$ such as Bellman [13]; Carr [16]; Claus [19]; Fox, Gavish and Graves [25]; Gavish and Graves [26]; Held and Karp [33]; Lawler et al. [37]; and Miller, Tucker and Zemlin [43]. And considers for study a new multistage insertion (MI) formulation of the $STSP$ [5, 11]. *Insertion* is a local search heuristic commonly employed to generate a tour involving $k+1$ cities from a tour that involves k cities, where k varies from 3 to $n-1$ [23, 37]. The sequence of insertion decisions made to insert city $k+1$ in an edge available in the k-tour resulting from the earlier insertion decisions starting with the unique 3-city tour, (1, 2, 3, 1) was formulated in Arthanari [5] as an integer programming problem (MI-formulation), solving which yields the best tour.

Naddef [44] succinctly summarizes the comparative strengths of the different models for $STSP$. One can paraphrase it as follows: Among all the known integer linear programming models, three models emerge and attain the same value for their linear relaxations. These are as follows: the MI-formulation, the *cycle-shrink* formulation of Carr [16] and the standard formulation or the *subtour elimination* formulation of Dantzig, Fulkerson and Johnson [22]. The multistage insertion is inspired by dynamic programming recursion—building up a tour step-by-step. Cycle shrink does the opposite—going from a tour to a node. Not surprisingly these two formulations are equivalent (see proof in [12]).

Haerian [30] as part of her doctoral thesis has compared different formulations of the symmetric travelling salesman problem, including the standard formulation or DFJ formulation with respect to (i) the LP relaxation of the formulation and integral-

ity gap, (ii) number of simplex iterations and (iii) CPU time used to find an optimal continuous solution. It turns out that *MI*-formulation has emerged more often than not as the winner as far as the gap is concerned. Also, it has shown superiority among those formulations with similar gap by needing less number of simplex iterations. Gubb [29] compared 19 formulations of *STSP* that model the problem from different perspectives, namely, flows, insertions and subtours (some of these are in the list of references [19, 25, 26, 43, 50, 52, 53]). He concludes that even among those formulations that are similar in polytope-wise implications, they vary in computational efficiency. *MI*-formulation stands out in this experimental comparison as well. Unfortunately, in both Haerian [30] and Gubb [29], the sizes of the instances from *TSPLIB* [49] considered are less than 300. The reason for this limitation arises primarily from the capacity of the commercial LP solver software used. *MI*-formulation has $\frac{n(n-1)}{2} + (n-3)$ constraints and $\tau_n = \sum_{k=4}^{n} \frac{(k-1)(k-2)}{2}$ variables. In order to solve larger size problems, one needs to abandon using general purpose LP algorithms to solve the LP instances that are sparse, with $0, \pm 1$-matrices, with the non-zero elements occurring in specific positions that can be given by a formula. In this chapter, we consider the constraint matrix of the *MI*-formulation for further clue to devise special purpose LP algorithms to exploit completely the structure of the problem matrix.

Rest of the chapter is structured as follows: Sect. 5.2 provides notations used and some definitions and concepts from graph theory. Section 5.3 gives a brief account of the different formulations of STSP and their comparisons. Sections 5.4–5.6 develop the concepts and algorithms that are required to solve the *MI*-relaxation as a hypergraph minimum cost flow problem. The last two sections provide details of our four-phase research and concluding remarks.

5.2 Preliminaries

In this chapter, we define some terminologies used in graph theory and give an introduction to linear programming and the simplex algorithm. The definitions and algorithms are taken from Cunningham [21], Bondy et al. [14] and Cook [20].

5.2.1 Graph Theory

Let V be a finite non-empty set of elements called nodes or vertices and E be a subset of $V \times V$ with its elements called edges be defined by $e = (v, u) \in E$ such that $u, v \in V$ are the end points of edge $e \in E$. An edge $e \in E$ is called an incident edge to some node $v \in V$, if node v is an end point of e. For an edge $e \in E$, if the edge is directed from one end point to the other, it is called a directed edge; otherwise, it is called an undirected edge.

Definition 1 A graph G is defined using a set of nodes V and a set of edges E and is denoted by $G = (V, E)$. Graphs are generally categorized as undirected graphs or directed (digraphs). If all $e \in E$ are undirected edges, then G is an undirected graph; if all $e \in E$ are directed edges, then G is a directed graph, otherwise G is a mixed graph.

Given a graph $G = (V, E)$, an incident edge of node $v \in V$ are all edges that have node $v \in V$ as an end point. The set of incident edges is denoted by $\delta(v)$.

Definition 2 The degree of a node is the number of edges incident to that node.

Definition 3 Two nodes in a graph are called adjacent to each other if there exists an edge joining them.

Definition 4 A node sequence $(v_0, \ldots, v_k)$ in G is called a path if there are no repetitions for all $i = 1, \ldots, k$ in the sequence and $e = (v_{i-1}, v_i) \in E$. A node sequence $(v_0, \ldots, v_k)$, for some $3 \leq k \leq |V|$ is called a cycle, if $v_0 = v_k$, and the sequence $(v_0, \ldots, v_{k-1})$ is a path. A cycle that contains all the node in V is called a Hamiltonian cycle.

Definition 5 Let $E(U) = \{(i; j) \in E \mid i, j \in U\}$; for some $U \subseteq V$ in graph G. A graph $G' = (V', E')$ is called a subgraph of G if $V' \subseteq V$. A subgraph $G' = (V', E')$ of G is called a component of G, if and only if there is a path between any two nodes in V' and not between any of the nodes from V' and $V \backslash V'$.

Definition 6 If a graph has only one component, it is called a connected graph. A connected graph with no cycles is called a tree. Given a digraph, a strongly connected component of the graph is a subset of V such that for any given set of vertices u and v in the component, there is a path from u to v.

Let *R, Q, Z, N* denote the set of reals, respectively, and B stands for the binary set of $\{0, 1\}$. Let R^d denote the set of d-tuples of R. Similarly, the superscript ^{d} is applied the same to the rationals, integers and natural numbers.

Let $K_n = (V_n, E_n)$ be the complete graph of $n \geq 4$ vertices, where $'V_n = \{1, \ldots, n\}$ is the set of vertices labelled in some order, and $E_n = \{e = (i, j) \mid i, j \in V_n, i < j\}$ is the set of edges.

Definition 7 A subset HC[1] of E_n is called a Hamiltonian cycle in K_n if it is the edge set of a simple cycle in K_n, of length n. We also call such a Hamiltonian cycle an n-tour in K_n.

Definition 8 A combinatorial optimization problem (COP) aims to find a $X \in F$ that minimizes $c(X)$, where

1. E be a finite set called the ground set.

[1] We use HC to represent a Hamiltonian cycle instead of H because in later sections, we use H to represent a hypergraph.

2. F is a collection of subsets of E.
3. $c : F \rightarrow R$ denotes a cost function.

Let $\{0, 1\}^{|E|}$ denote the set of all $0-1$ vectors indexed by E. Since any subset of E can be given by a $0-1$ vector, called the incidence vector, the collection F can be equivalently given by a subset F of $\{0, 1\}^{|E|}$.

We can make the following observations:

1. The convex hull of F, denoted by $conv(F)$, is a $0-1$ polytope.
2. The set of vertices of the polytope can be seen as F.
3. We can create a combinatorial optimization problem by establishing (E, F, c).

For example, the STSP is equivalent to the problem of finding a Hamiltonian cycle that minimizes a linear objective function over the set of all Hamiltonian cycles (or $n-tours$) in K_n is a COP. For this problem, E is the set of edges in a complete graph on n vertices, E_n. F is the set of incidence vectors of $HC \in HC_n$. And we are given the cost function $c \in R^{|E_n|}$. Let Q_n denote the polytope $conv(F)$.

5.3 Formulations for the TSP

Over the last century, there have been various formulations suggested for the TSP. These formulations usually trade off between the number of constraints for an increasing number of variables. However, the goal remains the same—to find the most optimal values for the variables which satisfy the constraints and minimizes the total cost. In the following sections, we present and compare a few formulations of the TSP and discuss the strength of the LP relaxation of these formulations.

5.3.1 Dantzig, Fulkerson and Johnson

The most well-known TSP formulation is the formulation proposed by Dantzig, Fulkerson and Johnson (DFJ) in 1954.

$$\min \sum_{j=1}^{n} \sum_{i}^{n} c_{ij} x_{ij}$$

Subject to

$$\sum_{i=1}^{n} x_{ij} = 1, \quad \forall j = 1, \ldots n \tag{5.1}$$

$$\sum_{j=1}^{n} x_{ij} = 1, \quad \forall i = 1, \ldots n \tag{5.2}$$

$$\sum_{i,j \in S} x_{ij} \leq |S| - 1, \quad \forall S \subseteq V, 2 \leq |S| \leq n - 1 \tag{5.3}$$

$$x_{ij} \in \{0, 1\}, \forall i, j. \tag{5.4}$$

Constraints (5.1) and (5.2) make sure that every node is visited. Constraint (5.3) is known as the subtour elimination constraint and makes sure that the output, in the end, is a Hamiltonian cycle. The algorithm partitions the set V into two groups: nodes that have already been visited and nodes that have yet to be visited while constraining the sum of the values of the edges that are connected between the two groups. The DFJ formulation has $2^{n-1} + n - 1$ constraints and $n(n - 1)$ variables. The exponential number of subtour elimination constraints creates a barrier for implementing this formulation efficiently. However, Dantzig et al. [22] solved the LP relaxation of the formulation with subtour elimination constraints relaxed, this would allow outputs of non-Hamiltonian cycles to be produced. To compensate for the relaxation, Dantzig et al. would let the algorithm run until a complete tour was produced [3].

5.3.2 Cycle Shrink

Carr [17] proposed the cycle-shrink relaxation which is an LP formulation that models the LP relaxation of the DFJ formulation. For a given node some $k \in V$, let $V_k = \{k + 1, \ldots, n\}$ and let $G_k = (V_k, E_k)$ be a subgraph of the complete graph G that is induced by V_k. For each edge $e \in E_k$, we define a decision variable x_e^k.

Let x^0 be an indicator vector of a Hamiltonian cycle $H^0(x^0)$ in G. Let $H^1(x^0)$ be a Hamiltonian cycle in G^1 that is formed by removing vertex 1 from $H^0(x^0)$ and connected its neighbours with an edge. Similarly, let $H^k(x^0)$ be a Hamiltonian cycle in G^k that is obtained by removing vertex k from $H^{k-1}(x^0)$ in G^{k-1} and connecting its neighbours.

The incidence vector of H^0 is indicated as $x^0 = (x_e^0 \,|e \in E)$ and for any $k \in \{1, \ldots, n - 3\}$ the incidence vector of H^k is $x^k = (x_e^k \,|e \in E_k)$. The solution to Carr's formulation are sequences of nodes that have been removed from initial Hamiltonian cycles in G and are represented by $x = (x^0, x^1, \ldots, x^{n-3})$. The cycle-shrink relaxation formulation by Carr is

$$\min \sum_{e \in E} c_e x_e^0.$$

Subject to

$$x_e^0 \geq 0, \forall e \in E \tag{5.5}$$

$$\sum_{e \in \delta(\{j\}) \cap E_k} x_e^k = 2, \forall k \in \{0, \ldots, n - 3\}, \forall j \in V_k \tag{5.6}$$

$$x_e^{k-1} - x_e^k \leq 0, k \in \{1, \ldots, n-3\}, e \in E_k. \tag{5.7}$$

Carr [17] has shown that all the subtour elimination constraints would be satisfied by a feasible solution for the cycle-shrink model. Let $\tau_n = \sum_{k=4}^{n}(k-1)(k-2)/2$. The cycle-shrink model has $(\tau_{n+1} + (n+3)(n-2)/2)$ constraints and τ_{n+1} variables.

5.3.3 The Multistage Insertion Formulation for the STSP

Arthanari [5] proposed the Symmetric Travelling Salesman Problem (STSP) as a multistage dynamic programming problem. He presented the mathematical programming formulation and showed that the slack variables from this formulation are the edge-tour incidence vectors. This formulation was called the Multistage Insertion formulation (MI formulation) and uses n^3 variables and n^2 constraints.

Let $K_n = (V_n, E_n)$ be a complete graph with $n \geq 4$ vertices, where $V_n \in \{1, \ldots, n\}$ is a set of vertices labelled in an arbitrary order, and $E_n \in \{e = (i, j) \mid i, j \in V_n, i < j\}$ is a set of edges. The cardinality of E_n denoted by p_n is $n(n-1)/2$ as K_n is a complete graph. We assign a unique edge label $l_{ij} = p_{j-1} + i$ to each edge $e = (i, j) \in E_n$. For a subset $F \subseteq E_n$ the characteristic vector of F is represented by $x_F \in \mathbb{R}^{p_n}$. Assuming that edges in E_n are ordered in increasing order according to their edge labels, the characteristic vector is defined as follows:

$$x_F(e) = \begin{cases} 1, & \text{if } e \in F, \\ 0, & \text{otherwise.} \end{cases}$$

For a subset $S \subset V_n$, we define $E(S) = \{(i, j) \in E_n \mid i, j \in S\}$. The set $\delta(S)$ denote the set of edges with one node in S and one node in $V_n \setminus S$.
Let $T_k = [v_1, v_2, \ldots, v_k, v_1]$ be an STSP tour of size k also called a $k - tour$ corresponding to a Hamiltonian cycle in a graph $K_k = (V_k, E_k)$, where $1 \leq k \leq n$. Let $v_i \in V_k$ for $3 \leq i \leq k$ indicate that the ith node in the k-$tour$ T_k.

The MI formulation is based on $n-3$ iterations of node insertions into the 3-tour $T_3 = [1, 2, 3, 1]$. This tour is eventually expanded to an $n - tour$ as the nodes from 4 to n are inserted successively into the tour. The decision of choosing an edge for insertion at state $k-3$ for $4 \leq k \leq n$ is represented by the variable x_{ijk}, for all $1 \leq i < j < k$, such that

$$x_{ijk} = \begin{cases} 1, & \text{if node } k \text{ is inserted between nodes } i \text{ and } j, \\ 0, & \text{otherwise.} \end{cases}$$

The first stage of the insertion starts with the decision of inserting node 4 into one of the edges in T_3, i.e. node 4 is inserted between one of the edges in the set $\{(1, 2)(1, 3)(2, 3)\}$. Suppose the edge which is chosen is labelled as $(i_4, j_4) \in E_3$ then the available edges in the next stage would be $\{(1, 2)(1, 3)(2, 3)\} \cup$

$\{(i_4\,,4),\ (j_4\,,4)\}\backslash\{(i_4\,,\,j_4)\}$. Generally, the tour that is constructed at stage k, depends on available edges from the $(k-1)^{\text{th}}$ stage and also on the choice of edge $(i_{k-1}\,,\,j_{k-1})$ for the insertion of the node $k-1$. The set of available edges in each stage A_k can be shown as $A_k = A_{k-1} \cup \{(i_{k-1}\,,k),\ (j_{k-1}\,,k)\}\backslash\{(i_{k-1}\,,\,j_{k-1})\}$.

Since by the end of the $n-3$ stages, each node $4 \le k \le n$ is inserted into one edge only, we have the condition as a constraint

$$\sum_{1\le i<j<k} x_{ijk} = 1, \forall\, 4 \le k \le n. \tag{5.8}$$

For each edge of the initial 3-tour, namely, the edges $\{(1,\ 2),\ (1,\ 3),\ (2,\ 3)\}$ can be used for the insertion of at most one node $4 \le k \le n$. This condition can be shown as a constraint

$$\sum_{k=1}^{n} x_{ijk} \le 1, \forall\, 1 \le i,\ j \le 3. \tag{5.9}$$

At the $k-3$ stage of insertion, the edge $(i\,,\,j)$ that is needed for insertion of node k is required to have existed at stage $k-3$, implying that the edge $(i\,,\,j)$ must have been created in one of the stages prior to stage $k-3$. Additionally, no node other than k should be inserted between the edge $(i\,,\,j)$.

Moreover, if the edge $(i\,,\,j) \notin E_3$, in some stage prior to $k-3$, edge $(i\,,\,j)$ needs to be constructed by inserting j into either edge $(r\,,i)$ or $(i\,,s)$, where $1 \le r < i$ and $i < s < j$. This requires that the sum $\sum_{r=1}^{i-1} x_{rij} + \sum_{s=i+1}^{j-1} x_{isj} = 1$. Second, this edge could be used for insertion by only one node $k > i$. These two conditions are combined to create the constraint

$$-\sum_{r=1}^{i-1} x_{rij} - \sum_{s=i+1}^{j-1} x_{isj} + \sum_{k=j+1}^{n} x_{ijk} \le 0, \quad \forall\, 4 \le j < n, 1 \le i < j. \tag{5.10}$$

Let c_{ij} denote the cost of an edge $(i\,,\,j) \in E_n$. Insertion of node k into edge $(i\,,\,j)$, would replace edge $(i\,,\,j)$ with two new edges $(i\,,k)$ and $(j\,,k)$. This replacement increases the total cost of the tour by $C_{ijk} = c_{ik} + c_{jk} - c_{ij}$.

The MI formulation minimizes the total incremental cost of the tour that is made by the node insertions at each stage. Since the initial cost of the 3-tour $c_{12} + c_{13} + c_{23}$ is the same in all of the tours of a given instance, it is not included in the objective function of the MI formulation. The complete MI formulation is given below [5]:

$$\min \sum_{k=4}^{n} \sum_{1\le i<j<k} C_{ijk}\, x_{ijk}.$$

Subject to

$$\sum_{1 \le i < j < k} x_{ijk} = 1, \forall 4 \le k \le n \tag{5.8}$$

$$\sum_{k=1}^{n} x_{ijk} \le 1, \forall 1 \le i\ j \le 3 \tag{5.9}$$

$$-\sum_{r=1}^{i-1} x_{rij} - \sum_{s=i+1}^{j-1} x_{isj} + \sum_{k=j+1}^{n} x_{ijk} \le 0, \quad \forall 1 \le i < j,\ 4 \le j < n. \tag{5.10}$$

The number of constraints for the MI formulation is $p_n + n - 3$, and the number of variables is $\tau_n = \sum_{k=4}^{n} p_{k-1}$. By relaxing the integrality constraint from the MI formulation and adding the following constraint, the MI-relaxation problem is defined.

$$-\sum_{r=1}^{i-1} x_{rin} - \sum_{s=i+1}^{n-1} x_{isn} \le 0 \quad , i = 1, \ldots, n-1 \tag{5.11}$$

Although constraint (5.11) is non-binding, it is added as a constraint to the model because of its corresponding slack variables. The slack variables of the formulation determine the edges that are chosen for the tour. The polytope given by the LP relaxation of the MI formulation is denoted by $P_{MI}(n)$. Let u_{ij}, $1 \le i < j \le n$, be the slack variables corresponding to the inequalities in the MI formulation. Arthanri and Usha [11] uses slack variables of the inequalities (5.9)–(5.11) to define the corresponding tour as given in (5.12)

$$u_{i\,j} = \begin{cases} 1, \text{ if edge (i,j) is present in the tour,} \\ 0, \text{ otherwise.} \end{cases} \tag{5.12}$$

Definition 9 ([11]) Let e_k be a vector of size $1 \times k$ with all its coordinates equal to one. The matrix corresponding to Eq. (5.8) is denoted E_n and is constructed as follows.

For $n = 4$, we have

$$E_4 = e_{\frac{3\times 2}{2}} = (1, 1, 1).$$

For $n = 5$, we have

$$E_5 = \begin{pmatrix} e_{\frac{3\times 2}{2}} & 0 \\ 0 & e_{\frac{4\times 3}{2}} \end{pmatrix} = \begin{pmatrix} 1\,1\,1\,0\,0\,0\,0\,0\,0 \\ 0\,0\,0\,1\,1\,1\,1\,1\,1 \end{pmatrix}.$$

In general, E_n can be constructed recursively as shown below:

$$\begin{pmatrix} e_{\frac{3\times 2}{2}} & 0 & 0 & \cdots & 0 & 0 \\ 0 & e_{\frac{4\times 3}{2}} & 0 & \cdots & \vdots & 0 \\ 0 & 0 & \ddots & & 0 & 0 \\ 0 & 0 & & \ddots & 0 & 0 \\ 0 & 0 & & & e_{\frac{(n-2)\times(n-3)}{2}} & 0 \\ 0 & 0 & & & 0 & e_{\frac{(n-1)\times(n-2)}{2}} \end{pmatrix} = \begin{pmatrix} E_{n-1} & 0 \\ 0 & e_{\frac{(n-1)\times(n-2)}{2}} \end{pmatrix}.$$

Let

$$A^{(n)} = \begin{pmatrix} I_{p_{n-1}} \\ -M_{n-1} \end{pmatrix},$$

where M_i corresponds to coefficients of constraints (5.9)–(5.11), the matrix corresponding to constraints (5.9)–(5.11) is denoted as A_n and it is constructed as follows:

$$A_n = \begin{pmatrix} A_4 & & & \\ 0 & A_5 & & \\ \vdots & \vdots & \ddots & \\ 0 & 0 & 0 & A_n \end{pmatrix} = \begin{pmatrix} A_{n-1} & \\ 0 & A_n \end{pmatrix}.$$

Let U denote the vector of slack variables in the MI formulation and let $C^T = (c_{124}, c_{134}, c_{125}, \ldots, c_{(n-2)(n-1)n})$. Based on the definition of A_n and E_n and some manipulation by letting $C^T = -c^T A_n$, the MI formulation can also be defined as problem.

Problem 1

$$\min C^T X$$

s.t.

$$\begin{pmatrix} E_n & 0 \\ A_n & I \end{pmatrix} \begin{pmatrix} X \\ U \end{pmatrix} = \begin{pmatrix} e_{n-3} \\ e_3 \\ 0 \end{pmatrix}, \; X, \; U \geq 0.$$

5.3.4 The Pedigree Polytope

The integer solution to the MI formulation for the STSP has a 1-1 correspondence with a combinatorial object called *pedigree*. In the following section, we will define the *pedigree* and give an example of using the MI formulation to solve an STSP problem of size 5.

Let HC_n be the set of all Hamiltonian cycles of $K_n = (V_n, \; E_n)$ and let $HC^k \in HC_k$ for all $3 \leq k \leq n$. Let $e = (i, \; j) \in E_{k-1}$, by inserting k in e is equivalent as replacing e by $\{(i, \; k), (j, \; k)\}$.

Definition 10 ([7]) Edge Generators: Given $e = (i,\ j) \in E_n$, $G(e)$ is called the set of generators of e

$$G(e) = \begin{cases} \delta(i) \cap E_{j-1}, & \text{if } j \geq 4 \\ E_3 \backslash \{e\}, & \text{otherwise.} \end{cases}$$

As edge $e = (i,\ j)$ for when $j > 3$ is created through inserting j into any existing edge in a stage prior to this one, all the edges that are replaced at each stage of the MI formulation are added to the G(e).

Let $n = 5$, $e = (1,\ 4)$. In this example, $j = 4$ which is greater than 3, so $G(e) = \delta(i) \cap E_{j-1}$ and as $i = 1$ therefore using the definition of $\delta(i)$, $\delta(1) = \{(1, 2), (1, 3), (1, 4), (1, 5)\}$. $E_{j-1} = E = \{(1, 2), (1, 3), (2, 3)\}$. Therefore, hence $G(e) = \{(1, 2), (1, 3)\}$.

Definition 11 ([7]) Given n, considering $W = (e_4, \ldots e_n)$, where $e_k = (i_k,\ j_k)$. For $1 \leq i_k < j_k \leq k - 1, 4 \leq k \leq n$. W is called a pedigree if and only if

1. $e_k,\ 4 \leq k \leq n$ are all distinct,
2. $e_k \in E_{k-1},\ 4 \leq k \leq n$ and
3. for every $k,\ 5 \leq k \leq n$, there exists a $e' \in G(e_k)$ such that $e_q = e'$, where $q = \max\{4,\ j_k\}$.

Let P_n denote the set of all pedigrees for a given $n > 3$. For any $4 \leq k \leq n$, given an edge $e \in E_{k-1}$, with edge label l, we can associate a 0-1 vector, $x(e) \in B^{\tau_n}$ such that $x(e)$ has a 1 in the lth coordinate, and zeros everywhere else. That is, $x(e)$ is an indicator vector of e.

Let $E = E_3 \times E_4 \cdots \times E_{n-1}$ be the ground set. Let B^{τ_n} denote the set of all binary vectors with τ_n coordinates. That is, here $\{0,\ 1\}^{|E|} = B^{\tau_n}$. Then, we can associate an $X = (x_4, \ldots x_n) \in B^{\tau_n}$, the characteristic vector of the pedigree W, where $(W)_k = e_k$, the $(k - 3)$rd component of W, $4 \leq k \leq n$ and x_k is the indicator of e_k.

Let $P_n = \{X \in B^{\tau_n} : X$ is the characteristic vector of W a pedigree$\}$. Consider the convex hull of P_n. We call this the *pedigree polytope*, denoted by $conv(P_n)$.

Given a cost vector $C \in \mathbb{R}^{\tau_n}$ the goal is to find pedigree X^* in P_n that minimizes CX^*.

We illustrate this with a 5-city example, with the cost matrix C, using the MI formulation to solve for the most optimal tour and formulating it using Problem 1.

$$c = \begin{array}{c} \\ 1 \\ 2 \\ 3 \\ 4 \\ 5 \end{array} \begin{array}{c} \begin{array}{ccccc} 1 & 2 & 3 & 4 & 5 \end{array} \\ \begin{pmatrix} & 30 & 26 & 50 & 50 \\ & & 24 & 40 & 50 \\ & & & 24 & 26 \\ & & & & 30 \\ & & & & \end{pmatrix} \end{array}$$

We wish to solve

$$\min C^T X$$

$$\begin{pmatrix} E_5 & 0 \\ A_5 & I \end{pmatrix} \begin{pmatrix} X \\ U \end{pmatrix} = \begin{pmatrix} e_{n-3} \\ e_3 \\ 0 \end{pmatrix}, X, U \geq 0.$$

Expanding

$$\begin{pmatrix} E_5 & 0 \\ A_5 & I \end{pmatrix}$$

we get

	4	4	4	5	5	5	5	5	5	0	0	0	0	0	0	0	0	0	0
	1,2	13	23	12	1,3	23	14	24	34	1,2	1,3	2,3	1,4	2,4	3,4	1,5	2,5	3,5	4,5
4	1	1	1																
5				1	1	1	1	1	1										
1,2	1			1						1									
1,3		1			1						1								
2,3			1			1						1							
1,4	−1	−1					1						1						
2,4	−1		−1					1						1					
3,4		−1	−1						1						1				
1,5				−1	−1		−1									1			
2,5				−1		−1		−1									1		
3,5					−1	−1			−1									1	
4,5							−1	−1	−1										1

Finally, using any *pedigree* for K_5 as a basic feasible solution, we can run an LP solver to solve for X as in the formulation in Problem 1. The final result is the 5 *tour*, $T_5 = \{(1,2), (2,3), (3,4), (4,5), (1,5)\}$ with a total distance of 148, this corresponds to the *pedigree*: ((1,3), (1,4)).

Research on pedigree polytopes and their connection to tour polytopes are of interest. Pedigree polytope and its properties are studied by Arthanari [6–8] and Ardekani and Arthanari [31]. Arthanari [9] proves that a sufficiency condition for non-adjacency in tour polytope is non-adjacency of the corresponding pedigrees in the pedigree polytope.

Makkeh, Pourmoradnasseri and Theis [42] modelled [8] characterization of adjacency for the Pedigree polytope using an 'adjacency game'. The authors show that the minimum degree for a vertex in a pedigree graph is asymptotically equal to the number of vertices—i.e. asymptotically, the graph is almost complete.

The main result of that paper: The minimum degree of a vertex on the Pedigree polytope for n cities is $(1 - o(1)) * (n-1)!/2$ for $n \to \infty$.

And in rephrasing the theorem: $\forall \varepsilon > 0$ there is an integer N such that $\forall n \geq N$ and all cycles A_n, with node set $\{n\}$, if B_n is drawn uniformly at random from all cycles with node set $\{n\}$ then the probability that the pedigree graph corresponding to the pedigrees A_n, and B_n is complete larger than or equal to $1 - \varepsilon$.

The next section reports some computational comparisons made by Haerian [30] and Gubb [29].

Table 5.1 Table comparing number of constraints and variables for the different formulations of TSP

Formulation	Constraints	Integer variables	Continuous variables
DFJ [22]	$2^{n-1} + n - 1$	$n(n-1)$	
FLOOD [24]	n^2	$n(n+1)(n-1)$	
MTZ [43]	$n^2 - n + 2$	$n(n-1)$	$(n-1)$
FGG [25]	n	$n(n-1)$	$n(n-1)(n+1)$
WONG [53]	$2(n^3 + n^2 + 1)$	$n(n-1)$	$2n(n-1)^2$
CLAUS [19]	$n^3 + n^2 + 3n$	$2n^2 + 2n$	
CARR [17]	$\tau_{n+1} + (n+3)(n-2)/2$	τ_{n+1}	
MI [5]	$n(n-1)/2 + (n-3)$	n^3	

5.3.5 Comparisons

The biggest disadvantage of the DFJ formulation is the exponential number of subtour elimination constraints. This has motivated researchers to suggest more compact formulations which have polynomial number of constraints. There are many other formulations such as Bellman [13]; Carr [16]; Claus [19]; Fox, Gavish and Graves [25]; Gavish and Graves [26]; Held and Karp [33]; Lawler et al. [37]; and Miller, Tucker and Zemlin [43]. We show the number of constraints and the number variables of the different TSP formulations in Table 5.1.

To compare the different formulations when used in LP-based solutions methods, Padberg and Sung [48] have used a special transformation technique to map polytopes given by other formulations into the DFJ formulation variable space. After finding the projection of different formulations into the DFJ variable space, the sizes of the projected polytopes were compared with that of the DFJ formulation. They showed that the DFJ formulation gives the tightest polytope for the TSP.

Although most of these formulations have a polynomial number of constraints, they have taken some sort of trade-off in the quality of their LP relaxation.

Arthanari and Usha [12] show that the multistage insertion and Carr's cycle-shrink formulations are equivalent and that the MI formulation is as tight as the DFJ formulation. Haerian [30] compared different formulations of the symmetric travelling salesman problem with respect to:

1. the LP relaxation of the formulations,
2. the integrality gap of the formulations,
3. number of simplex iterations taken to reach the solution and
4. CPU time used to find an optimal continuous solution.

She found that the continuous solution solved by the *MI*-formulation has one of the best integrality gaps and is among the formulations that take lesser simplex iterations.

Unfortunately, in Haerian's [30] research, she was not able to solve instances from TSPLIB [49] which have size greater than 300 cities. The reason for this limitation arises primarily from the capacity of the commercial LP solver software used.

In order to solve larger size problems, one needs to abandon using general purpose LP algorithms to solve the LP instances and take advantage of the special structure that arises from the MI relaxation. In the following sections, we present ideas from Arthanari [10] for the solution to this problem and details of the implementation of a prototype which acts as a proof of concept.

5.4 Hypergraphs

Hypergraphs are a generalization of a graph where an edge can join any number of vertices, while a normal graph edge consists of a pair of nodes, hyperedges or hyperarcs contain an arbitrary number of nodes.

The following definitions are from Cambini et al. [15].

Definition 12 A directed hypergraph is a pair of $H = (V, E)$, where $V = \{v_1, \ldots, v_n\}$ is the set of vertices and $E = \{e_1, \ldots, e_k\}$ is the set of hyperarcs. Therefore, E is a subset of $P(V) \setminus \{\emptyset\}$ where $P(V)$ is the power set of v. A hyperarc e is a pair $(T_e,\ h_e)$ where $T_E \subset V$ is the tail of e and $h_e \in V \setminus T_e$ is its head. A hyperarc that is headless, $(T_e, \emptyset)$ is called a sink and a tailless hyperarc $(\emptyset, h_e)$ is called a source.

Definition 13 The size of a hypergraph can be defined by

$$size(H) = \sum_{e_i \in E} |e_i|.$$

Given a hypergraph $H = (V, E)$, a positive real multiplier $\mu_v(e)$ associated with each $v \in T_e$, a real demand vector b associated with V, and a non-negative capacity vector w, a flow on H is a function $f : E \to \mathbb{R}$ which satisfies

$$\sum_{v=h_e} f(e) - \sum_{v=T_e} \mu_v(e) f(e) = b(v), \quad \forall v \in V \qquad (Conservation) \qquad (5.13)$$

$$0 \le f_e, \forall e \in E, \qquad (Feasibility) \qquad (5.14)$$

$$f_e \le w(e), \forall e \in E, \qquad (Capacity). \qquad (5.15)$$

Problem 2 Minimum cost hypergraph flow problem
Let c(e) be the cost associated with the hyperarc $e, \forall e \in E$. Find f^* such that $\sum_{e \in E} c(e) f^*(e)$ is a minimum over all f satisfying the flow constraints stated above.

A directed path P_{st} from s to t in H is a sequence $P_{st} = (v_1 = s, e_1, v_2 \ldots, e_q, v_{q+1} = t)$ where $s \in T_{e1}$, $h_{eq} = t$ and $v_i \in T_e \cap h_{e_{i-1}}$ for $i = 2, \ldots, q$.

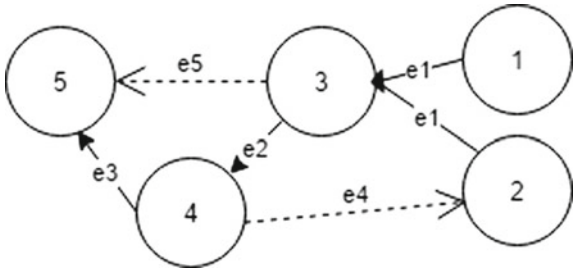


Fig. 5.1 Showing an example of a spanning hypertree. Let $T_R = (\{1, 2\}, e_1, 3, e_2, 4, e_3, 5)$. In this example, $R = \{1, 2\}$, $N = \{3, 4, 5\}$, $E_T = \{e_1, e_2, e_3\}$, $E_X = \{e_4, e_5\}$

If $s = t$, then P_{st} is a directed cycle and when no directed cycle exists in the graph, then H is called a cycle-free hypergraph.

Definition 14 A directed hyperpath $\prod_{st}$ from the source set S to the sink node t, such that each node with the exception of the nodes in S has exactly one entering hyperarc. A hyperarc e' is said to be a permutation of a hyperarc e if $T_e \cup \{h_e\} = T_{e'} \cup \{h_{e'}\}$. A hypergraph H' is a permutation of a hypergraph H if its hyperarcs are permutations of the hyperarcs of H.

Definition 15 A directed hypertree with root set R and a set of hyperarcs E_T called tree arcs is a hypergraph $T_R = (R \cup N, E_T)$ such that

1. T_R has no isolated nodes and does not contain any directed cycles.
2. $R \cap N = \emptyset$.
3. Each node $v \in N$ has exactly one entering hyperarc.
4. No hyperarc has a vertex of R as its head.

Remark 1 T_R is a directed hypertree with root set R and has a set of nodes N which are non-root nodes. Any non-root node not contained in the tail of any tree arc is called a leaf. Any permutation of a directed hypertree rooted at R yields an undirected hypertree rooted at R.

It can be shown that T_R is a directed hypertree rooted at R if and only if T_R has no isolated nodes, with $R \cap N = \emptyset$ and $|N| = |E_T| = q$ and an ordering $(v_1, \ldots, v_q)$ and $(e_1, \ldots, e_q)$ exists for the elements of N and of E_T such that $h_{e_j} = v_j$ and $R \cup \{v_1, \ldots, v_{j-1}\} \supseteq T_{ej}, \forall e_j \in E_T$.

Definition 16 An undirected hypertree rooted at R is any permutation of a directed hypertree rooted at R. In the case of undirected hypertrees, a leaf is a non-root node which belongs to exactly one hyperarc.

Definition 17 A spanning hypertree of $H = (V, E)$ is an undirected hypertree $T_R = (V, E_T)$ such that $E_T \subseteq E$ and $(T_e \cup \{h_e\}) \nsubseteq R$, $\forall e \in E \backslash E_T$. Figure 5.1 gives a hypergraph with spanning hypertree T_R.

Definition 18 E_X is a subset of $E \backslash E_T$ and columns corresponding to this set of hyperarcs that form a linearly independent set with the columns corresponding to the hyperarcs of the spanning tree.

5.5 Hypergraph Simplex

The simplex algorithm, developed by George Dantzig in 1947, solves linear programming (LP) problems starting with a basic feasible solution, first tests whether the optimality conditions are satisfied by the current basis and if satisfied stops. Otherwise, it selects a suitable variable not in the basis to enter the basis (reduced costs for the non-basic variables are computed for this purpose) and a corresponding basic variable leaves the basis, to form a new basis. The algorithm continues until an optimal solution is found.

One can specialize the simplex method to solve the minimum cost flow problem which can be efficiently solved using the network simplex method. The network simplex method adopts the simplex algorithm and finds the optimal solution by pushing flow through a network [47].

Like the simplex algorithm, the hypergraph simplex algorithm proposed by Cambini et al. [15] is a generalized formulation of the network simplex algorithm which solves the minimum cost flow problem on a hypergraph instead of a regular graph. Starting with a spanning tree of the hypergraph which is a basic feasible solution, the FLOW method (explained below) can be used to determine the optimal amount of flow to be pushed through each hyperarc. The POTENTIAL method is used to calculate the reduced cost for each hyperarc that is not in the solution. Just like the simplex algorithm, the hypergraph simplex algorithm selects a hyperarc that is violating the optimality condition and enters that hyperarc into the basis and forces out of the basis a corresponding hyperarc. In the minimum cost hypergraph flow problem, each basis M corresponds to a pair (T_R, E_X) where T_R is a spanning hypertree of the sub-hypergraph H^* corresponding to M. And E_X is the set of external hyperarcs, that is, the basic hyperarcs outside the spanning hypertree.

Flows and Potential

Flow

For any $|N|$ vector $d(N)$ and any $|E_X|$ vector $f(E_X)$, there exists unique vectors $d(R)$ and $f(T)$ such that $f = (f(T), f(E_X))$ is a flow which satisfies the conservation constraints at the nodes, with $d = (d(R), d(N))$ as the demand vector. Both $f(T)$ and $d(R)$ can be determined in $O(size(H))$ time through the flow algorithm shown below which was adapted from the flow algorithm proposed in Cambini et al. [15].

The flow algorithm takes four input parameters. They are as follows:

1. A hypergraph $H = (V, E)$.
2. A hypertree $T_R = (R, e_1, v_1, e_2, v_2 \ldots e_q, v_q)$.
3. Demand vector for non-root nodes $d(N)$.
4. Flow vector for edges not in the hypertree $f(E_X)$.

The outputs of the flow algorithm are as follows:

1. Demand vector for root nodes $d(R)$.
2. Flow vector for edges in the hypertree $f(T)$.

Potential

The *reduced cost* of hyperarc e is

$$c(e) + \sum_{v \in T_e} \mu_v(e)\pi(v) - \pi(h_e),$$

where c_e is the cost of e, and $\pi(v)$ is the potential of node v.

For any $|E_T|$ cost vector $c(T)$ and any $|R|$ vector $\pi(R)$, there exist a unique potential vector, $\pi(N)$, and cost vector, $c(E_X)$, such that the reduced cost of each basic hyperarc is equal to zero. The running time for this algorithm is $O(size(H))$.

The potential algorithm takes four input parameters. They are as follows:

1. A hypergraph $H = (V, E)$.
2. A hypertree $T_R = (R, e_1, v_1, e_2, v_2 \ldots e_q, v_q)$.
3. Cost vector for edges in the hypertree $c(T)$.
4. Potential vector for edges in root nodes of the hypertree $\pi(R)$.

The outputs of the Potential algorithm are as follows:

1. Cost vector for edges not in the hypertree, $c(E_X)$.
2. Potential vector for the non-root nodes in the hypertree, $\pi(N)$.

The Root Matrix

Definition 19 ([15]) Let $T_R = (V, E_T)$ be one of the spanning hypertrees of H, rooted at R, and

$$A = \begin{bmatrix} B & C \\ U & D \end{bmatrix}$$

be the incidence matrix of H in canonical form with respect to T_R.

The root matrix of H is the $|R| \times |E_X|$ matrix $A_R = (C - BU^{-1}D)$. Each column of A_R corresponds to one of the external hyperarcs, while each of its rows corresponds to one of the roots.

Let $A_R(*, e)$ be the column of A_R corresponding to hyperarc e and $A_R(v, *)$ be the row of A_R corresponding to the root node v.

This root matrix can be calculated by the flow and potential algorithm defined previously [15].

Definition 20 Let M be the incidence matrix of a sub-hypergraph with $|V|$ nodes and $|V|$ hyperarcs of a hypergraph H. If M is non-singular, then the sub-hypergraph cannot have isolated nodes, otherwise, M should have a zero row, and as a consequence, it has a spanning hypertree, T_R since $|R| = |E_X|$. M can be converted in canonical form with C and the root matrix M_R being square matrices.

The rooted spanning trees which characterize the basis matrices in the case of standard graphs, are particular spanning hypertrees, where the root set is a singleton.

$M_R{}^{-1}$, the inverse of the root matrix M_R can be calculated in terms of flows and potentials.

Primal, Dual and Hypergraph Simplex

The algebraic equivalent problem being solved by the hypergraph simplex algorithm is of type $Mf = \bar{b}$ and $\pi M = \bar{c}$, where M is an $|V| \times |V|$ basis of H and $\bar{b} = (\bar{b}(R), \bar{b}(N))$ and $\bar{c} = (\bar{c}(T), \bar{c}(E_X))$ are vectors of length $|V|$ and needs to be solved.

Cambini et al. [15] show that the first system can be interpreted as the problem of finding on the sub-hypergraph (whose incidence matrix M) a flow f satisfying a given demand vector $\bar{b}$. The solution to this system $Mf = \bar{b}$ can be obtained as the sum of a flow and of a circulation.

The primal algorithm first calculates the flow which satisfies the flow requirements at the non-roots and the relative root demand vector $d(R)$. Then it computes the circulation which yields a flow vector $f(E_X) = M_R{}^{-1}(\bar{b}(R) - d(R))$ on the external hyperarcs, and adds this circulation to the previously computed flow.
On the other hand, consider the second system $\pi M = \bar{c}$, where π_0 and c_0 are the potential vector and the cost vector on the external hyperarcs returned by the *Potential* algorithm. When $\pi(R) = 0$, let π_1 be the vector returned by the *Potential* algorithm when $\bar{c}(T) = 0$ and $\pi(R) = (\bar{c}(E_X) - c_0)M_R{}^{-1}$.

Similar to the PRIMAL algorithm, the DUAL algorithm first calculates the potential of the non-root nodes and then proceeds to calculate the reduced costs of all the hyperarcs which are not part of the current basis.

Optimality Testing

Let M be the current feasible basis. H^* the corresponding hypergraph and T_R one of the corresponding spanning hypertrees. Given the inverse of the root matrix $M_R{}^{-1}$, we can use the primal and dual algorithms from the previous section to carry out the computation of the primal basic solution $f = M^{-1}b^*$ and the corresponding dual vector $\pi = c^*M^{-1}$, where b^* is the demand vector induced on the nodes by the flows on the non-basic hyperarcs, while c^* is the cost vector relative to the basic hyperarcs.

The optimality conditions we check are, based on the reduced costs: the non-basic hyperarcs must have reduced costs ≥ 0, if their flow is zero, and if their flow is at the upper bound, then the reduced costs must be ≤ 0. If these conditions are satisfied, M is optimal and the algorithm terminates. Otherwise, the algorithm selects a hyperarc e' not in the basis which violates the optimality conditions and forces it into the basis.

After the algorithm determines which the entering and leaving hyperarc of the spanning tree are, the spanning tree needs to be updated. Cambini et al., [15] provide a detailed description of this method.

5.6 MI formulation in Hypergraph

Hyperflow and MI Relaxation

We can convert the MI-relaxation problem into a minimum cost hypergraph flow problem using the MI formulation of the STSP [10]. This is then solved using a

specialized version of the procedures used in the hypergraph simplex algorithm of Cambini et al. [15]. The details of these algorithms are provided in Appendix A.

We consider the following hypergraph $H = (V, E)$ corresponding to the MI formulation.

We have: $V = \{4, \ldots, n\} \cup \{(i,\ j) | \ 1 \leq i \leq j \leq n\}$ and $E = \{(\emptyset\ , (i,\ j)) | \ (i,\ j) \in V\} \cup_{k=4}^{n} \{(k : (i,\ j)) | \ 1 \leq i < j < k\}$ where $(k : (i,\ j))$ denotes the hyperarc, $(\{(i,\ k), (j,\ k),\ k\}, (i,\ j))$ for $1 \leq i < j < k,\ \forall k \in V$.

Theorem 1 *The hypergraph $H = (V, E)$ corresponding to MI formulation is cycle free.*

Proof Vertices in S = {4,…n} have no hyperarcs entering any of $k \in S$. So any directed path starting from k cannot end in k. So there are no cycles involving $k \in S$. Consider any (i, j) for any $1 \leq i < j \leq n$.

Case 1: $(i, j) \in V, 1 \leq i < j \leq 3$. Since none of these vertices is in the tail set of any hyperarc, a directed cycle involving such an (i, j) is not possible.

Case 2: $(i, j) \in V, 4 \leq i < j \leq n$. Suppose for some (i_0, j_0) there is a directed cycle, then (i_0, j_0) is the head of a hyperarc e and is in the tail set of another hyperarc e'.

Therefore, e' has to be an arc $(j_0 : (u, v))$ for some $u < v < j_0$ with u or $v = i_0$ and e is either $(\emptyset : (i_0, j_0))$ or $(r : (i_0, j_0))$ for $n \geq r > j_0$.

First we show that $e = (\emptyset : (i_0, j_0))$ is not possible. In any directed path, if e appears as e_i for some $1 \leq i \leq q$, then the vertex v_i is required to belong to $\{h_{e_{i-1}}\} \cup T_{e_i}$. But $T_{e_i} = T_e = \emptyset$ implies e cannot appear in any such directed path.

So $e = (r : (i_0, j_0))$ for some $n \geq r > j_0$. Now we have the directed path,

$$\ldots, (r : (i_0, j_0)), (i_0, j_0), (j_0 : (u, v)), \ldots$$

with $u < v < j_0$ and one of u or $v = i_0$.

Thus, any directed path $P_{st} = (v_1 = s = (i, j), e_1, v_2, \ldots e_q, v_{q+1} = t)$ in H is such that $h_{e_q} = t$ and that $t = (a, b)$ with $\max\{a, b\} < j$. Therefore $t = s$ is not possible, and hence H is cycle free. □

Theorem 2 *Given any pedigree P, we have a spanning hypertree of $H = (V, E)$ given by $H = (R \cup N, E_T)$ with*

$$E_T = \{(k : (i, j)) | (i, j) \in E(P)\} \cup \{(\emptyset : (i, j)) | (i, j) \in E_{n-1} \backslash E(P)\},$$

where $E(P) = \{(i_k, j_k) | 4 \leq k \leq n, \ni P = ((i_4, j_4), \ldots (i_n, j_n))\}$ is the given pedigree.

Proof Since $E_T \subset E$ and (V, E_T) is a sub-hypergraph of H and is cycle free. Let $R = \{k | 4 \leq k \leq n\}$, and $N = V \backslash R$. So (V, E_T) satisfies the requirement $R \cap N = \emptyset$. No hyperarc in E_T has a vertex in R as head. Every vertex $v = (i, j) \in N$ either has a unique hyperarc $(k : (i, j))$ entering (i, j) if $(i, j) \in E(P)$ or has a unique hyperarc $(\emptyset : (i, j))$ if $(i, j) \notin E(P)$. Thus (V, E_T) is a hypertree. Notice that it is a spanning hypertree as well, as all vertices in V are spanned. □

Since we have $m = |V|$ which is the number of vertices and we have only $|E_T|$ hyperarcs, we need to add $m - |E_T|$ other hyperarcs such that the incidence matrix corresponding to the spanning hypertree is extended to a basis of size $m \times m$.

Using this traversal of the hypertree, we can rearrange the incidence matrix of the spanning tree.

Remark 2

1. Observe that $\forall e \in E \backslash E_T$, $T_e \cup \{h_e\}$ is not a subset of R as every hyperarc not in E_T has $h_e \in E_n$ and so not in R.
2. E_X is a subset of $E \backslash E_T$ and columns corresponding to this set of hyperarcs form a linearly independent set with the columns corresponding to the hyperarcs of the spanning tree.
3. In the MI-hypergraph flow problem, we have a feasible basis given by the spanning tree corresponding to any pedigree, so we can start the hypergraph simplex algorithm without phase I using artificial variables.
4. The set of linearly independent columns corresponding to the hyperarcs in E_T needs to be expanded to a basis of size $m \times m$.
5. The basis is used by the primal and dual algorithms and the basis is changed based on the reduced cost of non-basis hyperarcs.

Consider $n = 6$ and the pedigree in P_6 given by $P = ((1, 3), (1, 4), (2, 3))$. The corresponding traverse, T_R of the spanning hypertree with $R = \{4, 5, 6\}$ and $E_T = \{((4 : (1, 3)), (5 : (1, 4)), (6 : (2, 3)))\} \cup \{(\emptyset, (i,\ j))|\ (i,\ j) \in E_5$ and $(i,\ j) \notin P\}$ is as follows:

$$T_R = (R, (\emptyset, (2, 6)), (2, 6), (\emptyset, (3, 6)), (3, 6), (6, (2, 3)), (2, 3), (\emptyset, (1, 5)), (1, 5),$$
$$(\emptyset, (4, 5)), (4, 5), (5, (1, 4)), (1, 4), (\emptyset, (3, 4)), (3, 4), (4, (1, 3)), (1, 3),$$
$$(\emptyset, (1, 2)), (1, 2), (\emptyset, (2, 4)), (2, 4), (\emptyset, (2, 5)), (2, 5), (\emptyset, (3, 5)), (3, 5),$$
$$(\emptyset, (1, 6)), (1, 6), (\emptyset, (4, 6)), (4, 6), (\emptyset, (5, 6)), (5, 6)).$$

We expand this to a basis by adding $(6 - 3) + (6 \times 5)/2 - 15 = 3$ more external hyperarcs $((4 : (1, 2)), (5 : (3, 4)), (6 : (3, 5))$ that are linearly independent of the set of columns corresponding to the 15 hyperarcs in E_T. This initial basis is shown in Table 5.2. We observe that it is an upper triangular matrix.

Now we can apply the hypergraph simplex algorithm discussed previously (shown in the appendix) to solve the *MI*-relaxation problem.

5.7 Implementation of Hypergraph Approach

From the previous section, we outlined the details from Arthanari [10] discussing how to convert the MI-relaxation problem of the STSP into a hypergraph flow problem and how to find the solution using the HySimplex methods from Cambini et al. [15]. The advantage of the HySimplex is that it only requires to perform the inverse of the root matrix of size $n \times n$. Therefore, in theory, it can be solved quicker than the standard MI relaxation on a normal graph which requires a matrix of size $n^2 \times n^2$ to

Table 5.2 Initial basis corresponding to a spanning hypertree generated by $P = ((1, 3), (1, 4), (2, 3))$

Hyperarc			06			05		04								4	5	6
Nodes	26	36	23	15	45	14	34	13	12	24	25	35	16	46	56	12	34	35
4								−1								−1		
5						−1											−1	
6			−1															−1
26	1		−1															
36		1	−1															−1
23			1															
15				1		−1												
45					1	−1											−1	
14						1		−1								−1		
34							1	−1									1	
13								1										
12									1							1		
24										1						−1		
25											1							
35												1					−1	1
16													1					
46														1				
56															1			−1

be inverted. Based on the reported superior performance of the hypergraph simplex algorithm by Cambini et al. [15], we are encouraged to conduct this computational comparison for the hypergraph MI formulation of the STSP.

We have designed our research in four phases:

1. Adapt the HySimplex methods from Cambini et al. [15] for the MI-relaxation problem of the STSP.
2. Implementation of a prototype from the algorithms in phase 1.
3. Optimizing the prototype.
4. Computational experiments.

In phase 1 of the research, we specialize the HySimplex method in order to fully exploit the MI-relaxation's recursive structure. Our version of the algorithm uses the fact that the tail of a hyperarc in the MI-relaxation problem is either $\emptyset$ or $\{(i, k), (j, k), k\}$ where the head is (i, j), and therefore cuts down a huge number of operations needed to be carried out by either algorithm.

Specifically, the for loop starting at line 4 of our flow algorithm (shown in Appendix A) iterates through every hyperarc in the set E_X. For each hyperarc in this loop, the algorithm makes a constant number of computations as the number of nodes in the set $T_e \cup h_e$ is fixed in the MI relaxation. In the flow algorithm of Cambini

et al. [15], the number of computations in the loop is not constant as it depends on the number of nodes in the set $T_e \cup h_e$ which is not fixed. This can be seen again on line 23 where the loop makes a constant number of iterations as the set w is fixed. We adapt the potential algorithm in a similar fashion. This can be seen on line 6 and line 27 of our adapted Potential algorithm (shown in Appendix A).

Cambini et al. [15] use the technique of inserting artificial hyperarcs with infinite capacity and large cost to create the first initial feasible basis. However, we have shown earlier that an initial feasible basis can be constructed from a pedigree. We have designed the CreateBasis algorithm (shown in Appendix A) which generates an initial feasible basis by constructing a pedigree and extending it to form a basis.

Phase 2 of the research has been carried out by the second author in his unpublished Master's dissertation. He was able to build a prototype which solves small problems and found that cycling was a major obstacle and occurred more frequently as the size of the problem increased. Phase 3 of this research will consist of solving the cycling issue found in Phase 2 and further optimize the prototype. Finally, in phase 4 we will conduct comparative computational experiments. We plan to test the minimum cost hypergraph flow simplex algorithms on two sets of problems: first STSP problems in the TSPLIB and second randomly generated Euclidean STSP problem instances with varying sizes. We will collect data such as the integrality gap, CPU time, number of iterations and other performance statistics. The main aim of the experiments is to estimate the compression in computational time achieved by the new algorithms compared to solving these LP instances using commercial as well as open source generic LP solvers.

5.8 Concluding Remarks

In this chapter, we introduced the TSP problem, followed by some preliminaries in graph theory. We then compare the DFJ, cycle-shrink and the MI formulation given by Arthanari [5]. Various advantages of the MI formulation were discussed in the previous sections. With the same LP relaxation values as the classic DFJ formulation, the MI formulation has only n^3 variables and n^2 constraints, compared to the DFJ with $n(n-1)$ variables and $2^{n-1} + n - 1$ constraints. Using Cplex, a commercial LP solver, the MI formulation has shown to be competitive compared to other formulations of the TSP by Ardekani [4] and Gubb [29].

We introduced the structure of a hypergraph and defined the hypergraph minimum cost flow problem. We considered how to interpret the MI formulation as a hypergraph minimum cost flow problem and presented some theoretical computational complexity results on the algorithms involved in solving the hypergraph minimum cost flow problem, namely, the flow and potential algorithm [10].

We presented our four-phase approach for our research and why we expect to solve larger instances of MI-relaxation problem using the hypergraph flow approach than that is possible with commercially available LP solvers. Finally, we outline the plans for future computational experiments to verify the efficacy of the suggested approach.

Appendix A—Hypergraph Algorithms

The four algorithms used to solve the minimum cost flow problem on the hypergraph are presented here.

Algorithm 1 Calculate Flow

Input: $H = (V, E)$; $T_R = (R, e_1, v_1, e_2, v_2 \ldots e_q, v_q)$; $d(N)$; $f(X)$;
Output: $d(R)$; $f(T)$;

procedure FLOW(H; T_R; $d(N)$; $f(X)$;)
  for $v \in R$ **do** $d(v) = 0$
  end for
  for $e = (k : (i, j)) \in E_X$ **do**
    $d(i, k) \leftarrow d(i, k) + f(e)$
    $d(j, k) \leftarrow d(j, k) + f(e)$
    $d(k) \leftarrow d(k) + f(e)$
    $d(i, j) \leftarrow d(i, j) - f(e)$
  end for
  for $v \in V$ **do**
    $unvisited(v) =$ number of hyperarcs incident into v
  end for
  $Queue = \{v \mid v$ is a leaf of $T_R\}$
  while $Queue \neq \emptyset$ **do**
    Select $v \in Queue$
    $Queue \leftarrow Queue \backslash \{v\}$
    Let $e_v = (k' : (i', j'))$
    if $v = (i, j)$ **then**
      $f(e_v) \leftarrow d(v)$
    else
      $f(e_v) \leftarrow -d(v)$
    end if
    for $w \in \{(i', j'), (i', k'), (j', k'), k'\} \backslash \{v\}$ **do**
      if $w = (i', j')$ **then**
        $d(w) \leftarrow d(w) - f(e_v)$
      else
        $d(w) \leftarrow d(w) + f(e_v)$
      end if
      $unvisited(w) \leftarrow unvisited(w) - 1$
      if $unvisited(w) = 1$ AND $w \notin R$ **then**
        $Queue \leftarrow Queue \cup \{w\}$
      end if
    end for
  end while
  for $v \in V$ **do**
    $d(v) = -d(v)$
  end for
  return demand at root nodes $d(R)$ and flows on SHT arcs $f(T)$
end procedure

Algorithm 2 Calculate Potential

Input: $H = (V, E)$; $T_R = (R, e_1, v_1, e_2, v_2 \ldots e_q, v_q)$; $c(T)$; $\pi(R)$;
Output: $c(X)$; $\pi(N)$;

procedure POTENTIAL(H; T_R; $c(T)$; $\pi(R)$;)
  for $e \in E_x$ **do**
    $c(e) = 0$
  end for
  for $v \in R$ **do**
    for $e \in E$ such that $v \in \{(i', j'), (i', k'), (j', k'), k'\}$ **do**
      if $e \in E_T$ and e corresponds to x_{ijk} **then**
        $c(e) \leftarrow c(i, j, k) + \pi(v)$
      else if $e \in E_T$ and e corresponds to u_{ij}
        $c(e) \leftarrow d(ij) - \pi(v)$
      end if
    end for
  end for
  for $e \in E$ **do**
    $unvisited(e)$ =number of nodes of N incident into e
  end for
  $Queue = \{e \mid e \in T_R$ and $unvisited(e) = 1\}$
  while $Queue \neq \emptyset$ **do**
    Select $e \in Queue$
    $Queue \leftarrow Queue \backslash \{e\}$
    Let v be a unique unvisited node of N incident to e
    if $v = (i, j)$ **then**
      $\pi(v) \leftarrow c(e)$
    else
      $\pi(v) \leftarrow -c(e)$
    end if
    for $e' \in E \backslash \{e\}$ such that $v \in \{(i', j'), (i', k'), (j', k'), k'\}$ **do**
      if $e = (i, j : k)$ **then**
        $c(e) \leftarrow c(e) - \pi(v)$
      else
        $c(e) \leftarrow c(e) + \pi(v)$
      end if
      $unvisited(e) \leftarrow unvisited(e) - 1$
      if $unvisited(e) = 1$ AND $e \notin E_X$ **then**
        $Queue \leftarrow Queue \cup \{e\}$
      end if
    end for
  end while
  for $e \in E_X$ **do**
    $c(e) = -c(e)$
  end for
  return potential at non-root nodes $\pi(N)$ and cost for arcs not in the SHT $c(X)$
end procedure

Algorithm 3 Calculate Primal

Input: $H = (V, E)$; $T_R = (R, e_1, v_1, e_2, v_2 \ldots e_q, v_q)$; ${M_R}^{-1}$; $\bar{b}$
Output: f: flow vector which satisfies $f = M^{-1}\bar{b}$
1: **procedure** PRIMAL(H; T_R; ${M_R}^{-1}$; $\bar{b}$)
2: $\{d(R), f(T)\} = Flow(H, T_R, \bar{b}(N), 0)$
3: $f(X) = {M_R}^{-1}(\bar{b}(R) - d(R))$
4: $\{\bar{b}(R), f(T)\} = Flow(H, T_R, \bar{b}(N), f(X))$
5: **return** $f(T)$ as f
6: **end procedure**

Algorithm 4 Calculate Dual

Input: $H = (V, E)$; $T_R = (R, e_1, v_1, e_2, v_2 \ldots e_q, v_q)$; ${M_R}^{-1}$; $\bar{c}$
Output: π: potential vector which satisfies $\pi M = \bar{c}$
1: **procedure** DUAL(H; T_R; ${M_R}^{-1}$ $\bar{c}$)
2: $\{c_0, \pi_0(N)\} = Potential(H, T_R, \bar{c}(T), 0)$
3: $\pi(R) = (\bar{c}(X) - c_0){M_R}^{-1}$
4: $\{\bar{c}(X), \pi(N) = Potential(H, T_R, \bar{c}(T), \pi(R))$
5: **return** $\pi(N)$ as π
6: **end procedure**

Algorithm 5 Create Initial Basis

Input: n, N
Output: E_B: A list of hypertree arcs and a list of external arcs.
1: **procedure** CREATEBASIS(n, N)
2: $Initial_N = \{(1, 2), (1, 3), (2, 3)\}$, $E_T = \{\}$, $E_X = \{\}$
3: **for** k from 4 to $n + 1$ **do**
4: $v = (i, j)$ Be a random node from N
5: $E_T := E_T \cup \{(v, i)\}$, $N := N \setminus \{v\}$
6: $v' = (i', j')$ Be a random node from N
7: $E_X := E_X \cup \{(v, i)\}$, $N := N \setminus \{v'\}$
8: **if** $k > i$ **then**
9: $N := N \cup (i, k)$
10: **else**
11: $N := N \cup (k, i)$
12: **end if**
13: **if** $k > j$ **then**
14: $N := N \cup (j, k)$
15: **else**
16: $N := N \cup (k, j)$
17: **end if**
18: **end for**
19: **for** $v \in N$ **do**
20: **if** $v \notin \{h_e\} \forall e \in E_T$ **then**
21: $E_T := E_T \cup ((v, \emptyset))$
22: **end if**
23: **end for**
24: **return** E_T, E_X
25: **end procedure**

References

1. Agarwala, R.: A fast and scalable radiation hybrid map construction and integration strategy. Genome Res. **10**(3), 350–364 (2000)
2. Applegate, D., et al. Concorde Home. http://www.tsp.gatech.edu/concorde/index.html
3. Applegate, D.L., Bixby, R.E., Chvatal, V., Cook, W.J.: The Traveling Salesman Problem: A Computational Study. Princeton University press (2006)
4. Ardekani, L.H., Arthanari, T.S.: Traveling salesman problem and membership in pedigree polytope-a numerical illustration. Modelling, Computation and Optimization in Information Systems and Management Sciences, pp. 145–154. Springer, Berlin (2008)
5. Arthanari, T.S.: On the traveling salesman problem. Mathematical Programming - The State of the Art. Springer, Berlin (1983)
6. Arthanari, T.S.: Pedigree polytope is a combinatorial polytope. In: Mohan, S.R., Neogy, S.K. (eds.) Operations Research with Economic and Industrial Applications: Emerging Trends, pp. 1–17. Anamaya Publishers, New Delhi (2005)
7. Arthanari, T.S.: On pedigree polytopes and Hamiltonian cycles. Discret. Math. **306**(14), 1474–1492 (2006)
8. Arthanari, T.S.: On the membership problem of pedigree polytope. In: Neogy, S.K., et al. (eds.) Mathematical Programming and Game Theory for Decision Making. World Scientific, Singapore (2008)
9. Arthanari, T.S.: Study of the pedigree polytope and a sufficiency condition for nonadjacency in the tour polytope. Discret. Optim. **10**(3), 224–232 (2013)
10. Arthanari, T.S.: Symmetric traveling salesman problem and flows on hypergraphs - new algorithmic possibilities. In: Atti della Accademia Peloritana dei Pericolanti- Classe di Scienze Fisiche, Matematiche e Naturali, under consideration (2017)
11. Arthanari, T.S., Usha, M.: An alternate formulation of the symmetric traveling salesman problem and its properties. Discret. Appl. Math. **98**(3), 173–190 (2000)
12. Arthanari, T.S., Usha, M.: On the equivalence of the multistage-insertion and cycle shrink formulations of the symmetric traveling salesman problem. Oper. Res. Lett. **29**(3), 129–139 (2001)
13. Bellman, R.E.: Dynamic programming treatment of the traveling salesman problem. J. Assoc. Comput. Mach. **9**(1), 61–63 (1962)
14. Bondy, J., Murthy, U.S.R.: Graph Theory and Applications. Springer, Berlin (2008)
15. Cambini, R., Gallo, G., Scutellà, M.G.: Flows on hypergraphs. Math. Program. **78**(2), 195–217 (1997)
16. Carr, R.D.: Polynomial separation procedures and facet determination for inequalities of the traveling salesman polytope. Ph.D. thesis, Carnegie Mellon University (1995)
17. Carr, R.D.: Separating over classes of TSP inequalities defined by 0 node-lifting in polynomial time. In: International Conference on Integer Programming and Combinatorial Optimization, pp. 460–474. Springer (1996)
18. Christofides, N.: Worst-case analysis of a new heuristic for the travelling salesman problem. Technical report DTIC Document (1976)
19. Claus, A.: A new formulation for the travelling salesman problem. SIAM J. Algebr. Discret. Methods **5**(1), 21–25 (1984)
20. Cook, W.: In Pursuit of the Traveling Salesman: Mathematics at the Limits of Computation. Princeton University Press, Princeton (2012)
21. Cunningham, W.H.: A network simplex method. Math. Program. **11**(1), 105–116. ISSN: 1436-4646 (1976). https://doi.org/10.1007/BF01580379
22. Dantzig, G.B., Fulkerson, D.R., Johnson, S.M.: Solution of a large-scale traveling-saesman problem. Oper. Res. **2**(4), 393–410 (1954)
23. Delí Amico, M., Maffioli, F., Martello, S. (eds.): Annotated Bibliographies in Combinatorial Optimization. Wiley-Interscience, Wiley, New York (1997)
24. Flood, Merrill M.: The traveling-salesman problem. Oper. Res. **4**(1), 61–75 (1956)

25. Fox, K.R., Gavish, B., Graves, S.C.: An n-constraint formulation of the time-dependent travelling salesman problem. Oper. Res. **28**(4), 1018–1021 (1980)
26. Gavish, B., Graves, S.C.: The travelling salesman problem and related problems. Working Paper, OR-078-78, Operations Research Center, MIT, Cambridge
27. Godinho, M.T., Gouveia, L., Pesneau, P.: Natural and extended formulations for the time-dependent traveling salesman problem. Discret. Appl. Math. **164**, 138–153 (2014)
28. Gouveia, L., Voß, S.: A classification of formulations for the (timedependent) traveling salesman problem. Eur. J. Oper. Res. **83**(1), 69–82 (1995)
29. Gubb, M.: Flows, Insertions and Subtours Modelling the Travelling Salesman. Project Report, Part IV Project 2011. Department of Engineering Science, University of Auckland (2011)
30. Haerian, A.L.: New insights on the multistage insertion formulation of the traveling salesman problem- polytopes, experiments, and algorithm. Ph.D. thesis, University of Auckland, New Zealand (2011)
31. Haerian, A.L., Arthanari, T.S.: Traveling salesman problem and membership in pedigree polytope - a numerical illustration. In: Le Thi, H.A., Bouvry, P., Tao, P.D. (eds.) Modelling, Computation and Optimization in Information Systems and Management Science, pp. 145–154. Springer, Berlin (2008)
32. Held, M., Karp, R.M.: A dynamic programming approach to sequencing problems. J. Soc. Ind. Appl. Math. **10**(1), 196–210 (1962)
33. Held, M., Karp, R.M.: The travelling salesman problem and minimum spanning trees. Oper. Res. **18**(6), 1138–1162 (1970)
34. Helsgaun, K.: An effective implementation of the Lin-Kernighan traveling salesman heuristic. Eur. J. Oper. Res. **126**(1), 106–130 (2000)
35. Karp, R.M.: Combinatorics, complexity, and randomness. Commun. ACM **29**(2), 97–109 (1986)
36. Kirkpatrick, S., Gelatt, C.D., Vecchi, M.P., et al.: Optimization by simmulated annealing. Science **220**(4598), 671–680 (1983)
37. Lawler, E.L., et al.: The Traveling Salesman Problem: A Guided Tour of Combinatorial Optimization. Wiley, Hoboken (1985)
38. Lenstra, J.K.: Technical noteclustering a data array and the travelingsalesman problem. Oper. Res. **22**(2), 413–414 (1974)
39. Letchford, A.N., Lodi, A.: Mathematical programming approaches to the traveling salesman problem. Wiley Encycl. Oper. Res. Manag. Sci. (2011)
40. Lin, S., Kernighan, B.W.: An effective heuristic algorithm for the traveling-salesman problem. Oper. Res. **21**(2), 498–516 (1973)
41. Little, J.D., Murty, K.G., Sweeney, D.W., Karel, C.: An algorithm for the traveling salesman problem. Oper. res. **11**(6), 972–989 (1963)
42. Makkeh, A., Pourmoradnasseri, M., Theis, D.O.: The graph of the pedigree polytope is asymptotically almost complete (2016). arXiv:1611.08419
43. Miller, C., Tucker, A., Zemlin, R.: Integer programming formulations and traveling salesman problems. J. Assoc. Comput. Mach. **7**(4), 326–329 (1960)
44. Naddef, D.: The Hirsch conjecture is true for (0;1)-polytopes. Math. Program. B **45**, 109–110 (1989)
45. Nagata, Y.: New EAX crossover for large TSP instances. Parallel Problem Solving from Nature-PPSN IX, pp. 372–381. Springer, Berlin (2006)
46. Öncan, T., Altınel, İ.K., Laporte, G.: A comparative analysis of several asymmetric traveling salesman problem formulations. Comput. Oper. Res. **36**(3), 637–654 (2009)
47. Orlin, J.B., Plotkin, S.A., Tardos, Éva: Polynomial dual network simplex algorithms. Math. Program. **60**(1), 255–276 (1993)
48. Padberg, M., Sung, T.Y.: An analytical comparison of different formulations of the travelling salesman problem. Math. Program. **52**(1–3), 315–357 (1991)
49. Reinlet, G.: TSPLIB - a traveling salesman problem library. ORSA J. Comput. **3**, 376–384 (1991)

50. Sarin, S.C., Sherali, H.D., Bhootra, A.: New tighter polynomial length formulations for the asymmetric traveling salesman problem with and without precedence constraints. Oper. Res. Lett. **33**(1), 62–70 (2005)
51. Schrijver, A.: On the history of combinatorial optimization (till 1960). Handbooks in Operations Research and Management Science, vol. 12, pp. 1–68. Elsevier, Amsterdam (2005)
52. Sherali, H.D., Driscoll, P.J.: On tightening the relaxations of miller-tucker-zemlin formulations for asymmetric traveling salesman problems. Oper. Res. **50**(4), 656–669 (2002)
53. Wong, R.T.: Integer programming formulations of the traveling salesman problem. In: Proceedings of the IEEE International Conference of Circuits and Computers, pp. 149–152 (1980)

Chapter 6
About the Links Between Equilibrium Problems and Variational Inequalities

D. Aussel, J. Dutta and T. Pandit

6.1 Introduction and Motivation

In the recent decades, a huge number of papers of the literature of optimization have been dedicated to equilibrium problem. In the community of optimizers, this terminology is used to describe the following problem: given a subset $C \subset \mathbb{R}^n$ and a (bi)function $f : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$, the *equilibrium problem* consists in

$$EP(f, C) \qquad \text{find } x \in C \text{ such that } f(x, y) \geq 0, \quad \text{for all } y \in C.$$

This understanding of the term 'equilibrium' seems to be quite far to its usual sense in game theory. It is actually not really the case as we will see in the example described in the forthcoming Sect. 6.4.

It was Oettli who in 1994 [4] first coined the term *equilibrium problem* during the annual conference of the Indian Mathematical Society and his paper was published in the journal *Mathematics Student* of the Indian Mathematical Society. It is one of the most cited papers in optimization theory.

The power of this formulation is that it allows to include, in a common framework, a large set of problems. For example, consider $f(x, y) = \varphi(y) - \varphi(x)$ and a subset C of $\mathbb{R}^n$. Then the solution set of the problem $EP(f, C)$ coincides with the set of global minimizers of the function φ over C. Now if one consider $f(x, y) = \varphi(x) - \varphi(y)$, then, symmetrically, the solutions of the equilibrium problem $EP(f, C)$ are the

D. Aussel
Lab. PROMES, UPR CNRS 8521, University of Perpignan, Perpignan, France

J. Dutta (✉)
Department of Economic Sciences, IIT Kanpur, Kanpur, India
e-mail: jdutta@iitk.ac.in

T. Pandit
Department of Mathematics and Statistics, Indian Institute of Technology, Kanpur, India

S. K. Neogy et al. (eds.), *Mathematical Programming and Game Theory*,
Indian Statistical Institute Series, https://doi.org/10.1007/978-981-13-3059-9_6

global maximizers of the function φ over C. Thus, the concept of an equilibrium problem seems to unify both minimization and maximization problems.

On the other hand if the objective function φ is assumed to be differentiable over a closed convex set C, then it is a well-known fact (and simple to prove) that if $\bar{x}$ is a local minimizer of φ over C, then

$$\langle \nabla\varphi(\bar{x}), y - \bar{x}\rangle \geq 0, \qquad \forall y \in C. \tag{6.1}$$

The above inequality expresses the necessary optimality condition in the so-called *variational inequality* form $VI(\nabla\varphi, C)$. Of course, if additionally f is convex, then the above expression is both necessary and sufficient for global optimality and thus, in context of convex optimization, a first relationship between equilibrium problem and variational inequality occurs since

$$EP(f_\varphi, C) = \arg\min_C \varphi = VI(\nabla\varphi, C) \qquad \text{where } f_\varphi(x, y) = \varphi(y) - \varphi(x), \tag{6.2}$$

where the notations EP and VI are both used for the problem itself and its solution set.

Our aim in this short note is to make a synthesis/state of art of the relationships (inclusions, equality) of equilibrium problems and variational inequalities, that is, to give sufficient conditions ensuring that one is included in the other one or that they coincide. Then in Sect. 6.4, we also emphasize through an example that the variational inequality is possibly the most general form of an equilibrium problem arising in applications.

6.2 State of the Art of Relationships

6.2.1 A First Step: VI and EP Generated by an Optimization Problem

Before going further into the relationship between equilibrium problems and variational inequalities, let us continue the reformulation process started above with the reformulation of optimization problems in terms of variational inequalities. Indeed, the link stated in (6.2) still holds true, under slight modifications, even if the objective function φ is not differentiable and/or not convex.

If φ is a lower semi-continuous proper convex function which is not assume to be differentiable, then one can use both the concepts of (convex) subdifferential and set-valued variational inequality in order to obtain a relation similar to (6.2). Let us recall that the subdifferential of the convex function φ at a point is given by $\partial\varphi(x) = \{v \in \mathbb{R}^n : \langle v, y - x\rangle \leq \varphi(y) - \varphi(x),\ \forall\, y \in \mathbb{R}^n\}$ and that the general framework of variational inequalities is the following: given a set-valued map $F : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ and a subset C of $\mathbb{R}^n$, the (somehow called *generalized*) variational inequality $VI(F, C)$ consists in:

$$\text{find } \bar{x} \in C \text{ such that there exists } \bar{x}^* \in F(\bar{x}) \text{ with } \langle \bar{x}^*, y - \bar{x} \rangle \geq 0, \quad \forall x \in C.$$

Thus taking these notations into account, it is well known that Eq. (6.2) extends in

$$EP(f_\varphi, C) = VI(\partial\varphi, C) \qquad \text{where } f_\varphi(x, y) = \varphi(y) - \varphi(x). \tag{6.3}$$

Now if φ is not assumed to be convex but only quasi-convex, then thanks to some recent developments (see, e.g. [1]), it is nevertheless possible to achieve the perfect reformulation of the minimization of φ over a convex set C in terms of a related variational inequality. To be more precise, let us first recall some definitions:

A function $\varphi : X \to \mathbb{R} \cup \{+\infty\}$ is said to be

- *quasi-convex* on K if,

$$\text{for all } x, \ y \in K \text{ and all } t \in [0, 1], \qquad \varphi(tx + (1-t)y) \leq \max\{\varphi(x), \varphi(y)\},$$

or equivalently

$$\text{for all } \lambda \in \mathbb{R}, \text{ the sublevel set } S_\lambda = \{x \in X \ : \ \varphi(x) \leq \lambda\} \text{ is convex.}$$

- *semi-strictly quasi-convex* on K if, φ is quasi-convex and for any $x, y \in K$,

$$\varphi(x) < \varphi(y) \Rightarrow \varphi(z) < \varphi(y), \quad \forall z \in [x, y[.$$

Clearly, any convex function is semi-strictly quasi-convex while semi-strict quasi-convexity implies quasi-convexity. Roughly speaking, a semi-strictly quasi-convex function is a quasi-convex function that has no 'full dimensional flat part' except eventually at arg min.

Some years ago, a new concept of sublevel set called *adjusted sublevel set* has been defined in [1]: for any $x \in \text{dom } f$, we define

$$S^a_\varphi(x) = S_{\varphi(x)} \cap \overline{B}(S^<_{\varphi(x)}, \rho_x),$$

where $S^>_\lambda = \{x \in X \ : \ \varphi(x) < \lambda\}$ stands for the strict sublevel set of φ at point x and moreover $\rho_x = dist(x, S^<_{\varphi(x)})$, if $S^<_{\varphi(x)} \neq \emptyset$

and $S^a_\varphi(x) = S_{\varphi(x)}$ if $S^<_{\varphi(x)} = \emptyset$.

Note that actually $S^a_\varphi(x)$ coincides with $S_{\varphi(x)}$ if $\text{cl}(S^>_{\varphi(x)}) = S_{\varphi(x)}$. It is, for example, the case whenever f is semi-strictly quasi-convex.

Based on this concept of sublevel sets, one can naturally define the following set-valued map called *adjusted normal operator* N^a_φ defined by

$$N^a_\varphi(x) = \{x^* \in \mathbb{R}^n \ : \ \langle x^*, y - x \rangle \leq 0, \ \ \forall y \in S^a_\varphi(x)\}.$$

Now following [2, Prop. 5.1], a necessary and sufficient optimality conditions can be proved for the minimization of a quasi-convex function over a convex set.

Proposition 6.2.1 *Let C be a closed convex subset of X, $\bar{x} \in C$ and $\varphi : \mathbb{R}^n \to \mathbb{R}$ be continuous semi-strictly quasi-convex such that* $\text{int}(S^a_\varphi(\bar{x})) \neq \emptyset$ *and* $\varphi(\bar{x}) > \inf_X \varphi$. *Then the following assertions are equivalent:*

(i) $\varphi(\bar{x}) = \min_C \varphi$.
(ii) $\bar{x} \in VI(N^a_\varphi \setminus \{0\}, C)$.

Let us observe that the notation $N^a_f \setminus \{0\}$ means that at any point x, 0 is dropped from the cone $N^a_\varphi(x)$. It is an essential technical point for the above equivalence since it allows to avoid any 'trivial solution' of the variational inequality.

As a consequence if C is a closed convex subset of X such that $C \cap \arg\min_{\mathbb{R}} f = \emptyset$, then one has an analogous of the extremely important equivalence (6.2) and it can be proved in the context of quasi-convex optimization.

$$EP(f, C) = \arg\min_C \varphi = VI(N^a_\varphi \setminus \{0\}, C) \qquad \text{where } f(x, y) = \varphi(y) - \varphi(x). \tag{6.4}$$

The table below summarizes the interrelations stated above between equilibrium problems and variational inequalities in the very particular case where they are defined through an optimization problem.

Initial problem	EP reformulation	Hypothesis	VI reformulation	Hypothesis
$\min_C \varphi$	$\arg\min_C \varphi = EP(f_\varphi, C)$ with $f_\varphi(x, y) = \varphi(y) - \varphi(x)$	none	$\arg\min_C \varphi = VI(\nabla\varphi, C)$	φ diff. convex C convex non-empty
			$\arg\min_C \varphi = VI(\partial\varphi, C)$	φ lsc proper convex C convex non-empty
			$\arg\min_C \varphi = VI(N^a_\varphi \setminus \{0\}, C)$	φ continuous and semi-strictly quasi-convex C convex non-empty $C \cap \arg\min_{\mathbb{R}^n} f = \emptyset$

6.2.2 The More General Case

Based on the interrelations recalled in the previous subsection, we will now explore the relations that can be stated between equilibrium problem $EP(f, C)$ and variational inequalities whenever the function f is not coming from an optimization problem.

Given a subset C of $\mathbb{R}^n$ and a set-valued map $F : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ and the associated variational inequality $VI(F, C)$, an immediate link with an equilibrium problem can be stated by simply considering a dedicated bifunction f_F:

$$VI(F, C) = EP(f_F, C) \qquad \text{where } f_F(x, y) = \langle F(x), y - x \rangle.$$

This equality being valid without any hypothesis one can thus consider that variational inequality problem are actually particular cases of the class of equilibrium problems. Let us now explore the reverse question that is under which conditions an equilibrium problem $EP(f, C)$ can be seen as a variational inequality problem.

For an equilibrium problem in the general framework to yield nice results, it is needed to fulfil some assumption on the data. One the most common assumptions in the literature (see, for example [4–13]) is the following:

(H1) $f(x, x) = 0$ for all $x \in \mathbb{R}^n$ (or for just $x \in C$).
(H2) For any $x \in \mathbb{R}^n$, the function $y \mapsto f(x, y)$ is a convex function.

The first condition shows that if x^* is a solution of the equilibrium problem then x^* minimizes the function $f(x^*, y)$ over C. Now assume that f is a differentiable convex function in y and C is non-empty and convex. Then we can write down the necessary and sufficient optimality condition as

$$\langle \nabla_y f(x^*, x^*), y - x^* \rangle \geq 0, \qquad \forall y \in C.$$

This shows that x^* solves the variational inequality $VI(F_f, C)$ where, for each $x \in \mathbb{R}^n$, $F_f(x) = \nabla_y f(x, x)$. Further if x^* solves $VI(F_f, C)$ then by (6.1) and the convexity of f in the second variable it is clear that x^* minimizes $f(x^*, .)$ over C and since $f(x^*, x^*) = 0$ we conclude that x^* solves $EP(f, C)$. Thus, the solution set of $EP(f, C)$ coincides with the solution set of $VI(F_f, C)$ once we assume that f is differentiable and convex in the second variable that is

$$EP(f, C) = VI(F_f(x), C) \qquad \text{where } F_f(x) = \nabla_y f(x, x).$$

Looking to the developments of Sect. 6.2.1, one can wonder if the above relation (6.2.2) can actually be generalized to the case where f is not differentiable and/or not convex in the second variable. First if, for any $x \in C$, the function $f(x, \cdot)$ is convex lower semi-continuous then, using the same proof as above one obtains

$$EP(f, C) = VI(F_f(x), C) \qquad \text{where } F_f(x) = \partial_y f(x, x).$$

Finally if $(H1)$ holds true, C is convex and the function is only assumed to be continuous and semistricly quasi-convex in the second variable then, as previously explained, x^* is a solution of $EP(f, C)$ if and only if x^* minimizes $f(x^*, .)$ and therefore, using Proposition 6.2.1, one immediately have

$$EP(f, C) = VI(F_f(x), C) \qquad \text{where } F_f(x) = N^a_{f(x,\cdot)}(x) \setminus \{0\}.$$

Thus, as a conclusion, even if it is true in a full generality, we often have that an equilibrium problem $EP(f, C)$ can be seen as a variational inequality.

The above stated interrelations are summarize in the table below, where assumption $(H1)$ and $(H2)$ is assumed to hold.

Initial problem	EP reformulation	Hypothesis	VI reformulation	Hypothesis
$VI(F, C)$	$VI(F, C) = EP(f_F, C)$ with $f_F(x, y) = \langle F(x), y - x\rangle$	none		
$EP(f, C)$			$EP(f, C) = VI(F_f, C)$ with $F_f(x) = \nabla_2 f(x, \cdot)(x)$	$f(x, .)$ diff. convex, $\forall x$ C convex non-empty $f(x, x) = 0, \ \forall x$
			$EP(f, C) = VI(F_f, C)$ with $F_f(x) = \partial_2 f(x, \cdot)(x)$	$f(x, .)$ lsc proper convex, $\forall x$ C convex non-empty $f(x, x) = 0, \ \forall x$
			$EP(f, C) = VI(F_f, C)$ with $F_f(x) = N^a f(x, \cdot)(x) \setminus \{0\}$	$f(x, .)$ continuous and semi-strictly quasi-convex, $\forall x$ C convex non-empty $C \cap \arg\min_{\mathbb{R}^n} f = \emptyset$

6.3 Existence Results for EP Through VI

Here, we present some existence results for both equilibrium problem and the variational inequality problem which are well established in the literature. We can see that the relation between EP and VI mentioned in the previous table implies the interrelation between the existence results of these two classes of problems.

Theorem 6.3.1 ([14, 15]) *Let C is a non-empty, convex and compact subset of $\mathbb{R}^n$ and let $F : \mathbb{R}^n \to \mathbb{R}^n$ be a continuous mapping. Then there exists a solution to the problem $VI(F, C)$.*

Theorem 6.3.2 *Let $C \subset \mathbb{R}^n$ be non-empty, convex and compact. Also let that $f : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ is bifunction such that $f(x, .)$ is convex, differentiable and $f(x, x) = 0$ for any $x \in X$. Then the solution set of the problem $EP(f, C)$ is non-empty*

Keeping in view of the relation between EP and VI as presented in the previous section it is clear that Theorem 6.3.2 follows in a straightforward fashion from Theorem 6.3.1 The following existence result for VI with set-valued function is a particular case of Theorem 3.1 [16].

Theorem 6.3.3 *Let C be a non-empty, convex, compact subset of $\mathbb{R}^n$ and $F : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ be an upper semi-continuous set-valued map with convex, compact values. Then VI(F,C) has a solution.*

A similar theorem is present in the literature by Ky Fan [17] for the equilibrium problem.

Theorem 6.3.4 (Theorem 1, [17]) *Let C is a non-empty, convex, compact subset of $\mathbb{R}^n$. If a continuous bifunction $f : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ satisfies the following properties:*
- *$f(x, .) : \mathbb{R}^n \to \mathbb{R}$ is convex for each $x \in C$.*
- *$f(x, x) = 0$ for any $x \in C$.*

Then the equilibrium problem $EP(f, C)$ has a solution.

Again looking at the relationship between EP and the VI with set-valued map as we have presented in the previous section, it is clear that Ky Fan's result can be deduced from Theorem 6.3.3.

There are some results in the literature about the existence of the solutions of $EP(f, C)$ and $VI(F, C)$, when C is closed but unbounded. But these results were developed independently. Here we show that the link between EP and VI problems leads to those existence results of EP once we assume the same for the VI.

Theorem 6.3.5 (Prop 2.2.3 [19]) *Let $C \subset \mathbb{R}^n$ be closed convex and $F : \mathbb{R}^n \to \mathbb{R}^n$ be continuous. If there exists $u \in \mathbb{R}^n$ such that the set*

$$V_< := \{x \in C : \langle F(x), x - u\rangle < 0\}$$

is bounded (possibly empty), then $VI(F, C)$ has a solution.

The next theorem is an existential result for the equilibrium problem developed by Iusem et al. (Theorem 4.2 [18]). Here, we show that the same result is obtained using the last theorem which ensures the existence of a solution of a VI problem.

Theorem 6.3.6 *Let $C \subset \mathbb{R}^n$ is closed convex and $f : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ is a bifunction such that $f(x, .) : \mathbb{R}^n \to \mathbb{R}^n$ is differentiable convex and $f(x, x) = 0$ for each $x \in C$. If there exists $u \in C$ such that the set*

$$L_> := \{x \in C : f(x, u) > 0\}$$

is bounded (possibly empty), then $EP(f, C)$ has a solution.

Proof As $f(x, .)$ is convex and differentiable function for all $x \in C$, we already know that $EP(f, C) = VI(F_f, C)$; where $F_f(x) = \nabla_2 f(x, .)(x)$. Also for any $x \in C$,

$$f(x, u) \geq f(x, x) + \langle \nabla_2 f(x, .)(x), u - x\rangle.$$

By the given hypothesis, we get

$$f(x, u) \geq \langle F_f(x), u - x\rangle. \tag{6.5}$$

From (6.5), it is clear that $\{x \in C : \langle F_f(x), x-u\rangle < 0\} \subseteq \{x \in C : f(x,u) > 0\} = L_>$. Now the boundedness of $L_>$ (possibly empty) implies that $\{x \in C : \langle F_f(x), x-u\rangle < 0\}$ is bounded (possibly empty). Then by Theorem 6.3.5, $VI(F_f, C)$ has a solution, implying that $EP(f, C)$ also has solution.

Remark 6.3.1 With the similar assumptions on F and C as Theorem 6.3.5 for VI, if we assume that there exists $u \in C$ and $\zeta \geq 0$ such that

$$\liminf_{\|x\|\to\infty} \frac{\langle F(x), x-u\rangle}{\|x\|^\zeta} > 0, \tag{6.6}$$

then $VI(F, C)$ has a solution (Prop. 2.2.7, [19]). Note that the coercivity condition (6.6) implies the boundedness of the set $V_<$ in Theorem 6.3.5.

Similar thing happens with the equilibrium problem also. The boundedness condition of $L_>$ can be replaced by the coercivity condition of f,

$$\liminf_{\|x\|\to\infty} \frac{-f(x,u)}{\|x\|^\zeta} > 0.$$

6.4 Examples and Counterexamples

In the previous section, it was shown that under some natural assumptions the solution set of an equilibrium problem coincides with the solution set of an associated variational inequality problem. Given the problem $EP(f, C)$, where C is nonempty and convex, f is differentiable and $(H1)$ holds. We shall call the problem $VI(F_f, C)$, with $F_f(x) = \nabla_y f(x, .)(x) = \nabla_y f(x, x)$ as the variational inequality associated with the equilibrium problem $EP(f, C)$. This is because if x^* is a solution of $EP(f, C)$, then x^* solves $VI(F_f, C)$, though the converse need not be true. Taking a clue from an example taken from Muu et al. [3], we show an equilibrium problem which can not be solved by solving the associated variational inequality.

Example 6.4.1 Consider the following equilibrium problem. Find $x \in C$ such that

$$f(x, y) \geq 0 \quad \text{for all} \quad y \in C,$$

where $f(x, y) = \langle x, y-x\rangle + x^2 - y^2$ and $C = [-1, 1]$. Since there does not exist any such $x \in [-1, 1]$, this equilibrium problem does not have any solution. Here $f_y(x, y) = x - 2y$, which implies that $\nabla f_y(x, x) = -x$. Hence, the variational inequality associated with the above mentioned equilibrium problem is given as follows. Find x such that

$$\langle -x, y-x\rangle \geq 0 \quad \text{for all } y \in [-1, 1].$$

Note $x = 0$ satisfies the above inequality for all $y \in [-1, 1]$, implying that the associated variational inequality has a solution when the equilibrium problem does not.

The above example shows that in general an equilibrium problem may not be related to its associated variational inequality. The above example might appear artificial. Thus, it is natural to ask if there is an example of an equilibrium problem which is drawn from some application where its solution set does not coincide with the solution of its associated variational inequality. While trying to search for such an example, we came across the work of Muu et al. [3], where they have studied the profit maximization problem in the setting of an oligopolistic market. They showed that the existence of Nash equilibrium in such a market is equivalent to a hemivariational inequality. However, they assumed that cost function which tells us the cost of producing a given amount of a good is concave and increasing. Under this assumption, the Nash equilibrium problem cannot be solved by solving the hemivariational inequality problem. However, this assumption is flawed from the economic point of view. It is common knowledge in microeconomics that the function relating the cost of producing a given good with the quantity to be produced is a strictly(or strongly) convex function. We show below that if we consider the correct economic assumption on the cost function, the problem discussed by Muu et al. [3] is indeed equivalent to a variational inequality. We describe the problem in considerable detail.

Let us begin by considering an oligopolistic market. In an oligopolistic market, there are more than one firm produces the same commodity and compete among themselves. Thus, the unit price of the commodity fixed by one firm does not depend only on its own level of production but depends also on the amount of production achieved by other forms. More precisely, consider that there are n firms and let x_i be the amount of the commodity produced by the ith firm and let p_i be the price of the commodity given by the ith firm. In fact, we should write the price as $p_i(x_1, x_2, \ldots, x_n)$. Let h_i be the cost function associated with the firm and thus for producing the amount x_i, the firm i needs to spend $h(x_i)$. Thus, the profit or the pay-off function for the ith firm is a function $f_i : \mathbb{R}^m \to \mathbb{R}$ is s given as

$$f_i(x) = f_i(x_1, \ldots x_m) = x_i\, p_i(x_1, \ldots, x_n) - h_i(x_i).$$

In fact, it is natural to assume that the cost function h_i of the ith firm depends only on production level of the ith firm itself. It is also important to note that in an oligopolistic structure, the number of firms is not very large. Further, we assume that each firm i has a strategy set $U_i \subset \mathbb{R}$ and we can safely assume it to be convex. This strategy set allows the firm i to set its production level once it has idea of the production level of other firms. This is quite natural since the number of firms is quite less. Thus, a point $\bar{x} = (\bar{x}_1, \ldots, \bar{x}_n) \in U = U_1 \times U_2 \times \cdots \times U_n$ is a *Nash equilibrium* if for each $i = 1, \ldots, n$

$$f_i(\bar{x}_1, \bar{x}_2, \ldots, \bar{x}_{i-1}, y_i, \bar{x}_{i+1}, \ldots \bar{x}_n) \leq f(\bar{x}_1, \bar{x}_2, \ldots, \bar{x}_i, \ldots, \bar{x}_n),$$

for all $y_i \in U_i$. In fact, one can express this as a sequence of minimization problem. Let x^{-i} denote the production levels of all the firms except the ith firm. Thus, we can write

$$x^{-i} = (x_1, x_2, \ldots, x_{i-1}, x_{i+1}, \ldots, x_n)^T.$$

Traditionally in the study of Nash equilibrium, one can write the vector x as $x = (x_i, x^{-i})$. Let us write the loss function for the ith firm as

$$\theta_i(x_i, x^{-i}) = -f_i(x_1, \ldots, x_n) = h_i(x_i) - x_i p_i(x_1, \ldots, x_n).$$

Thus for any given x^{-i}, the object of the ith firm is to choose a strategy which solves the problem $P_i(x^{-i})$ given as

$$\min_{x_i \in U_i} \theta_i(x_i, x^{-i}).$$

Let $S(x^{-i})$ denote the solution set of the problem $P_i(x^{-i})$. A vector $\bar{x}$ is a Nash equilibrium if $\bar{x}_i \in S(\bar{x}^{-i})$ for each $i = 1, \ldots, n$. In order solve the above problem, most economists would like to have at least have that θ_i is convex in x_i. Thus, this means that $h_i(x_i) - x_i p_i(x_1, \ldots, x_n)$ must be convex in x_i. In fact, Muu et al. [3] considers $p(x_1, \ldots, x_n) = \alpha_i - \beta_i(x_1 + \cdots + x_n)$ where, α_i and β_i are constants with $\beta_i \geq 0$. Note that in this case we have

$$x_i p_i(x_1, \ldots x_n) = \alpha_i x_i - \beta_i(x_1 x_i + \cdots + x_i^2 + \cdots + x_n x_i).$$

This is in fact concave in x_i. Further as per the standard assumptions in economic theory we consider that the cost function h_i is strongly convex and this proves that θ_i is convex in x_i . In fact, a careful inspection would show that it is actually jointly convex in all the variables. Through the following proposition our aim would be to show that under the above assumptions the Nash equilibrium can be computed by solving a hemivariational inequality through of the non-monotone type.

Proposition 6.4.1 *Let us assume that $\bar{x}$ is the Nash equilibrium of the oligopolistic market model discussed above. Let us assume that the cost function h_i of each of the ith firm is strongly convex and the unit price p_i quoted by the ith firm is given as*

$$p_i(x_1, \ldots, x_n) = \alpha_i - \beta_i(x_1 + \cdots + x_n),$$

where $\alpha_i \in \mathbb{R}$ and $\beta_i \geq 0$. Then $\bar{x}$ solves the hemivariational inequality $VI(F, +\nabla\varphi, U)$, where $F(x) = \tilde{B}x - \alpha$, $\alpha = (\alpha_1, \ldots, \alpha_n)^T$ with $\tilde{B}$ is a $n \times n$ matrix whose ith row has the entry 0 at the ith column and all other entries are β_i and φ is given as

$$\varphi(x) = \langle x, Bx \rangle + h(x),$$

where B is a diagonal matrix given as $B = diag(\beta_1, \ldots \beta_n)$ and $h(x) = \sum_{i=1}^{n} h_i(x_i)$. Conversely if $\bar{x}$ is a solution to $VI(F + \nabla\varphi, U)$ with F and φ as given above then $\bar{x}$ is indeed a Nash equilibrium for the oligopolistic market model.

Proof: Let us begin by assuming that $\bar{x}$ is the Nash equilibrium of the oligopolistic market model described above. Thus for each $i = 1, \ldots, n$, we have

$$f_i(\bar{x}_1, \bar{x}_2, \ldots, \bar{x}_{i-1}, y_i, \bar{x}_{i+1}, \ldots \bar{x}_n) \leq f(\bar{x}_1, \bar{x}_2, \ldots, \bar{x}_i, \ldots, \bar{x}_n),$$

for all $y_i \in U_i$. Of course, we know that $U = U_1 \times U_2 \times \cdots \times U_n$. From the above expression, a simple manipulation will show that

$$h_i(y_i) - y_i \left(\alpha_i - \beta_i \left(y_i + \sum_{j=1, j\neq i}^{n} \bar{x}_j \right) \right) \geq h_i(\bar{x}_i) - \bar{x}_i p_i(\bar{x}_1, \ldots, \bar{x}_n).$$

Further simplification shows that

$$h_i(y_i) - h_i(\bar{x}_i) + (\beta_i \bar{x}_1 + \cdots + \beta_i \bar{x}_{i-1} + \beta_i x_{i+1} + \cdots + \beta_i \bar{x}_n - \alpha_i)(y_i - \bar{x}_i) + \beta_i y_i^2 - \beta_i \bar{x}_i^2 \geq 0,$$

for all $y_i \in U_i$. Summing over all i from 1 to n we have

$$\sum_{i=1}^{n} h_i(y_i) - \sum_{i=1}^{n} h_i(\bar{x}_i) + \langle \tilde{B}\bar{x} - \alpha, y - \bar{x} \rangle + \langle y, By \rangle - \langle x, Bx \rangle \geq 0 \quad \forall y \in U.$$

This shows that $\bar{x}$ solves $VI(F + \nabla\varphi, U)$.

Conversely let $\bar{x}$ solve $VI(F + \nabla\varphi, U)$ with F and φ as described in the statement of the proposition. Thus, we have

$$\sum_{i=1}^{n} h_i(y_i) - \sum_{i=1}^{n} h_i(\bar{x}_i) + \langle \tilde{B}\bar{x} - \alpha, y - \bar{x} \rangle + \langle y, By \rangle - \langle x, Bx \rangle \geq 0 \quad \forall y \in U. \tag{6.7}$$

Let us choose $y \in U$ as follows:

$$y = (\bar{x}_1, \bar{x}_2, \ldots, \bar{x}_{i-1}, y_i, \bar{x}_{i+1}, \ldots, \bar{x}_n),$$

where y_i is any element from U_i. Plugging this y in (6.7), we get

$$h_i(y_i) - h_i(\bar{x}_i) + (\beta_i \bar{x}_1 + \cdots + \beta_i \bar{x}_{i-1} + \beta_i x_{i+1} + \cdots + \beta_i \bar{x}_n - \alpha_i)(y_i - \bar{x}_i) + \beta_i y_i^2 - \beta_i \bar{x}_i^2 \geq 0,$$

which implies that $f_i(\bar{x}_1, \ldots, \bar{x}_{i-1}, y_i, \bar{x}_{i+1}, \ldots, \bar{x}_n) \leq f_i(\bar{x}_1, \ldots, \bar{x}_n)$. This clearly shows that $\bar{x}$ is the Nash equilibrium of the oligopolistic market model. □.

As mentioned earlier in Muu et al. [3], it was assumed that h_i is an increasing concave function for each i. Then φ becomes a difference convex function, and thus, the $VI(F + \nabla\varphi, U)$ would truly become an equilibrium problem which cannot be solved by solving a VI. However as we had discussed this issue with several economists, they have clearly told us that concavity assumption on the cost function is fundamentally incorrect since in such a case the graph of the cost function of a firm may always remain below the price curve $x_i p_i(x_1, \ldots x_n)$ which leads the possibility of arbitrarily large amount of production in principle. However, no firm can make an arbitrarily large amount of commodities. The assumption of a convex curve limits the amount of commodities produced by the firm i and thus makes U_i a compact and convex set. This will make it much easier to handle the problems $(P_i(x^{-i}))$. Thus as we see that under the strong convexity assumption on the cost function of each firm, we have φ to be strongly convex and thus $VI(F + \nabla\varphi, U)$ is same as $VI(F + 2B + \nabla h, U)$, since the cost functions are assumed to be twice differentiable. Thus, the analysis of the Nash equilibrium of an oligopolistic market under natural assumptions does not lead us to an equilibrium problem different from a VI. To the best of our knowledge, the problem of finding an application which can be modelled as an equilibrium problem that is not equivalent to its associated variational inequality remains to be open.

Thus given assumptions (H1) and (H2), it appears that the most general form of an equilibrium problem is a variational inequality problem. However, a variational inequality problem is more general than an optimization problem. This is what the following example will demonstrate.

Example 6.4.2 Consider the following convex optimization problem (CP):

$$\min f(x) \quad \text{subject to} \quad g_i(x) \leq 0, i = 1, \ldots, m, \quad x \in X,$$

where f and each $g_i, i = 1, \ldots, m$ are finite-valued convex functions on X or $\mathbb{R}^n$ and X is a closed convex subset of $\mathbb{R}^n$. Associated with (CP) is the Lagrangian function $L : X \times \mathbb{R}^m_+ \to \mathbb{R}$ given as

$$L(x, \lambda) = f(x) + \lambda_1 g_1(x) + \cdots + \lambda_m g_m(x).$$

Assume that the Slater's condition holds, i.e. there exists $\hat{x} \in \mathbb{X}$ such that $g_i(\hat{x}) < 0$ for all $i = 1. \ldots, m$. It is a well-known result in convex optimization (see, for example, Dhara and Dutta [20]) that if Slater condition holds then $\bar{x} \in X$ is a minimizer of (CP) if and only if there exists $\bar{\lambda} \in \mathbb{R}^m_+$ such that

$$L(\bar{x}, \lambda) \leq L(\bar{x}, \bar{\lambda}) \leq L(x, \bar{\lambda}), \quad \text{for all} \quad x \in X, \lambda \in \mathbb{R}^m_+. \tag{6.8}$$

The point $(\bar{x}, \bar{\lambda}) \in X \times \mathbb{R}^m_+$ is called a saddle point of the Lagrangian function. From (6.8), it is clear that

$$L(\bar{x}, \bar{\lambda}) = \min_{x \in X} L(x, \bar{\lambda})$$
$$L(\bar{x}, \bar{\lambda}) = \max_{\lambda \in \mathbb{R}^m_+} L(\bar{x}, \lambda).$$

Now using the standard necessary optimality for convex optimization (see Rockafellar [21]), we conclude that

$$-\nabla_x L(\bar{x}, \bar{\lambda}) \in N_X(\bar{x})$$

and

$$\nabla_\lambda L(\bar{x}, \bar{\lambda}) \in N_{\mathbb{R}^m_+}(\bar{\lambda}).$$

Noting that

$$N_X(x) \times N_{\mathbb{R}^m_+}(\lambda) = N_{X \times \mathbb{R}^m_+}(x, \lambda)$$

we conclude that under the Slater condition $(\bar{x}, \bar{\lambda}) \in X \times \mathbb{R}^m_+$ is a saddle point of the Lagrangian function if and only if $(\bar{x}, \bar{\lambda})$ solves the following variational inequality:

$$0 \in F(x, \lambda) + N_{X \times \mathbb{R}^m_+}(x, \lambda)$$

where $F(x, \lambda) = (\nabla_x L(x, \lambda), -\nabla_\lambda L(x, \lambda))$. It is clear that $F(x, \lambda)$ is not the gradient of a convex function and in fact, it is not the gradient of the Lagrangian function jointly in both variable. Thus, we have a variational inequality which is not the optimality condition of convex optimization problem.

For more examples of this type, see Borwein and Dutta [22] and Borwein and Lewis [23].

6.5 Link Between QEP and QVI

A more general version of the variational inequality problem is the quasi-variational inequality (QVI) problem (see [24]) and the quasi-equilibrium problem (QEP) generalizes the standard equilibrium problem with set-valued maps. For more details on quasi-equilibrium problems, see, for example, [25, 26]. In this section, we present the observations about the relation between QEP and QVI, i.e. under which assumptions a QEP is equivalent to a QVI.

Given two set-valued maps $T : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ and $K : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$, the problem $QVI(T, K)$ is defined as:

$$\text{Find } x \in K(x) \text{ such that there exists } x^* \in T(x) \text{ with } \langle x^*, y - x\rangle \geq 0 \text{ for all } y \in K(x).$$

For a bifunction $f : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ and a set-valued map $K : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$, $QEP(f, K)$ is

$$\text{Find } x \in K(x) \text{ such that } f(x, y) \geq 0 \text{ for all } y \in K(x).$$

The following assumptions have been taken for $QEP(f, K)$ in this chapter which also appear in literature:
$\mathscr{A}1 : f(x, .)$ is convex for each $x \in \mathbb{R}^n$.
$\mathscr{A}2 : f(x, x) = 0$ for each $x \in \mathbb{R}^n$.
$\mathscr{A}3 : K(x)$ is non-empty, convex and closed for all $x \in \mathbb{R}^n$.

If x^* is a solution of $QEP(f, K)$, by assumption we get that $x^* \in K(x^*)$ and

$$f(x^*, y) \geq f(x^*, x^*) \quad \forall y \in K(x^*),$$

which is nothing but a solution of the following minimization problem:

$$\min f(x^*, y) \quad \text{subject to } y \in K(x^*).$$

Additionally if we assume that $f(x, .)$ is lsc, proper for any $x \in \mathbb{R}^n$, by the necessary and sufficient optimality condition for this problem we get that there exists $\xi^* \in \partial_2 f(x^*, .)(x^*)$ such that

$$\langle \xi^*, y - x\rangle \geq 0 \quad \forall y \in K(x^*). \tag{6.9}$$

This implies that x^* is a solution of the $QVI(T_f, K)$, where $T_f(x) = \partial_2 f(x, .)(x)$. Further if x^* solves $QVI(T_f, K)$, (6.9) holds with some $\xi^* \in \partial_2 f(x^*, .)(x^*)$. This together with $\mathscr{A}1$ and $\mathscr{A}2$ implies that x^* is a solution of $QEP(f, K)$. Hence

$$QEP(f, K) = QVI(T_f, K) \quad \text{where } T_f(x) = \partial_2 f(x, .)(x).$$

In particular when $f(x, .)$ is differentiable for any $x \in \mathbb{R}^n$, using the gradient for subdifferential we get

$$QEP(f, K) = QVI(T_f, K) \quad \text{where } T_f(x) = \nabla_2 f(x, .)(x).$$

Finally, if $\mathscr{A}2$ and $\mathscr{A}3$ are satisfied and the function $f(x, .)$ is semi-strictly quasi-convex function for each $x \in \mathbb{R}^n$, we still get an equivalence relation between QEP and QVI. If x^* is a solution of $QEP(f, K)$, we have

$$f(x^*, y) \geq 0 \quad \forall y \in K(x^*),$$

which implies that x^* is a solution of the equilibrium problem $EP(f, K(x^*))$. Then by Proposition 2.1, we can say that $EP(f, K(x^*)) = VI(T_f, K(x^*))$, where $T_f(x) = N^a_{f(x,.)}(x)\backslash\{0\}$, which implies there exists $y^* \in T_f(x^*) = N^a_{f(x^*,.)}(x^*)\backslash\{0\}$ such that

$$\langle y^*, y - x^* \rangle \geq 0 \quad \forall y \in K(x^*).$$

Hence, x^* solves $QVI(T_f, K)$. Again if we assume that x^* is a solution of the $QVI(T_f, K)$, following the previous arguments in reverse way we can easily show that x^* also solves $QEP(f, K)$.

The above-stated observations are summarized in the following table with the assumption that $\mathscr{A}2$ and $\mathscr{A}3$ hold:

Initial problem	QEP reformulation	Hypothesis	QVI reformulation	Hypothesis
QVI(T,K)	$QVI(T, K) = QEP(f_T, K)$ with $f_T(x, y) = \sup_{\xi \in T(x)} \langle \xi, y - x \rangle$	$T(x)$ is compact $\forall x \in \mathbb{R}^n$		
QEP(f,K)			$QEP(f, K) = QVI(T_f, K)$ with $T_f(x) = \nabla_2 f(x, .)(x)$	$f(x, .)$ diff. convex $\forall x \in \mathbb{R}^n$
			$QEP(f, K) = QVI(T_f, K)$ with $T_f(x) = \partial_2 f(x, .)(x)$	$f(x, .)$ convex, lsc, proper $\forall x \in \mathbb{R}^n$
			$QEP(f, K) = QVI(T_f, K)$ with $T_f(x) = N^a_{f(x,.)}(x)\backslash\{0\}$	f(x,.) continuous and semi-strictly quasi-convex $\forall x \in \mathbb{R}^n$ $K(x) \cap \operatorname{argmin}_{\mathbb{R}^n} f = \emptyset \quad \forall x \in \mathbb{R}^n$

References

1. Aussel, D.: Adjusted sublevel sets, normal operator and quasiconvex programming. SIAM J. Optim. **16**, 358–367 (2005)
2. Aussel, D., Ye, J.: Quasiconvex minimization on locally finite union of convex sets. J. Optim. Theory Appl. **139**, 1–16 (2008)
3. Le, D., Muu, D., Nguyen, V.H., Quy, N.V.: On Nash Cournot oligopolistic market equilibrium models with concave cost functions. J. Glob. Optim. **41**, 351–364 (2008)
4. Blum, E., Oettli, W.: From optimization and variational inequalities to equilibrium problems. Math. Stud. **63**, 123–145 (1994)
5. Chadli, O., Chbani, Z., Riahi, H.: Equilibrium problems with generalized monotone bifunctions and applications to variational inequalities. J. Optim. Theory Appl. **105**, 299–323 (2000)
6. Iusem, A.N., Sosa, W.: Iterative algorithms for equilibrium problems. Optimization **52**, 301–316 (2003)
7. Nguyen, T.T.V., Strodiot, J.J., Nguyen, V.H.: The interior proximal extragradient method for solving equilibrium problems. J. Glob. Optim. **44**, 175–192 (2009)
8. Iiduka, H., Yamada, I.: A subgradient-type method for the equilibrium problem over the fixed point set and its applications. Optimization **58**, 251–261 (2009)

9. Dinh, N., Strodiot, J.J., Nguyen, V.H.: Duality and optimality conditions for generalized equilibrium problems involving DC functions. J. Glob. Optim. **48**, 183–208 (2010)
10. Iusem, A.N., Sosa, W.: On the proximal point method for equilibrium problems in Hilbert spaces. Optimization **59**, 1259–1274 (2010)
11. Charitha, C.: A note on D-gap functions for equilibrium problems. Optimization **62**, 211–226 (2013)
12. Konnov, I.V.: On penalty methods for non monotone equilibrium problems. J. Glob. Optim. **59**, 131–138 (2014)
13. Anh, P.N., Hai, T.N., Tuan, P.M.: On ergodic algorithms for equilibrium problems. J. Glob. Optim. **64**, 179–195 (2016)
14. Eaves, B.C.: On the basic theorem of complementarity. Math. Program. **1**, 68–75 (1971)
15. Hartman, P., Stampacchia, G.: On some nonlinear elliptic differential functional equations. Acta Math. **115**, 153–188 (1966)
16. Aussel, D.: Quasimonotone quasivariational inequalities: existence results and applications. J. Optim. Theory Appl. **158**, 637–652 (2013)
17. Fan, K.: A minimax inequality and applications. In: Shisha, O. (ed.) Inequality III, pp. 103–113. Academic, New York (1972)
18. Iusem, A.N., Kassay, G., Sosa, W.: On certain conditions for the existence of solutions of equilibrium problems. Math. Program. **116**, 259–273 (2009)
19. Facchinei, F., Pang, J.-S.: Finite-Dimensional Variational Inequalities and Complementarity Problems. Springer Series in Operations Research, vol. I. Springer, New York (2003)
20. Dhara, A., Dutta, J.: Optimality Conditions in Convex Optimization: A finite-Dimensional View. With a Foreword by Stephan Dempe. CRC Press, Boca Raton (2012)
21. Rockafellar, R.T.: Convex Analysis. Princeton Mathematical Series, vol. 28. Princeton University Press, Princeton (1970)
22. Borwein, J.M., Dutta, J.: Maximal monotone inclusions and Fitzpatrick functions. J. Optim. Theory Appl. **171**, 757–784 (2016)
23. Borwein, J.M., Lewis, A.S.: Convex analysis and nonlinear optimization. Theory and examples. CMS Books in Mathematics/Ouvrages de Mathématiques de la SMC, vol. 3, Second edition edn. Springer, New York (2006)
24. Chan, D., Pang, J.S.: The generalized quasivariational inequality problem. Math. Oper. Res. **7**, 211–222 (1982)
25. Bianchi, M., Pini, R.: A note on equilibrium problems with properly quasimonotone bifunctions. J. Glob. Optim. **20**, 67–76 (2001)
26. Nasri, M., Sosa, W.: Equilibrium problems and generalized Nash games. Optimization **60**, 1161–1170 (2011)

Chapter 7
The Shrinking Projection Method and Resolvents on Hadamard Spaces

Yasunori Kimura

7.1 Introduction

Let H be a real Hilbert space and $f : H \to]-\infty, +\infty]$ a proper lower semicontinuous convex function. We consider the following problem called a convex minimization problem: Find an element $z \in H$ such that

$$f(z) = \inf_{y \in H} f(y).$$

The solution to this problem is called a minimizer of f. A large number of researchers have been working on this simple problem because it is related to various types of nonlinear problems such as equilibrium problems, variational inequality problems, saddle point problems, and others. In particular, there is a strong relation between convex minimization problems and fixed point problems for nonexpansive mappings and the concept of resolvent plays an important role to connect these two nonlinear problems.

The definition of the resolvent for f is as follows: For fixed $x \in H$, define $g_x : H \to]-\infty, +\infty]$ by

$$g_x(y) = f(y) + \|y - x\|^2.$$

Then, we know that g_x has a unique minimizer $y_x \in H$. Using this fact, we define an operator $J_f : H \to H$ by $J_f x = y_x$. Namely,

$$J_f x = \mathrm{argmin}_{y \in H}(f(y) + \|y - x\|^2).$$

This work was supported by JSPS KAKENHI Grant Number 15K05007.

Y. Kimura (✉)
Department of Information Science, Toho University, Miyama, Funabashi, Chiba 274-8510, Japan
e-mail: yasunori@is.sci.toho-u.ac.jp

S. K. Neogy et al. (eds.), *Mathematical Programming and Game Theory*, Indian Statistical Institute Series, https://doi.org/10.1007/978-981-13-3059-9_7

Since we know that the resolvent operator is nonexpansive and the set $\mathrm{Fix}\, J_f = \{z \in H : z = J_f z\}$ of its fixed points coincides with the set of minimizers of f, we may apply the notion of this operator to the theory of nonexpansive mappings and their fixed points, and consequently, we can obtain various kinds of useful results related to the convex minimization problem.

In this note, we will discuss approximation techniques to the solution of convex minimization problems by using iterative sequences with resolvent operators. As well as the history of this topic, we will show a new iterative scheme with respect to the common minimization problem for a finite family of convex functions.

7.2 Preliminaries

Let X be a metric space with a metric d. For $x, y \in X$, we say that a mapping $c : [0, l] \to X$ is a geodesic with endpoints x, y if $c(0) = x$, $c(l) = y$, and $d(c(t), c(s)) = |t - s|$ for any $t, s \in [0, l]$. We say that X is a geodesic metric space if a geodesic with endpoints x, y exists for every $x, y \in X$. Moreover, if a geodesic is unique for each $x, y \in X$, then X is said to be uniquely geodesic. In what follows, we assume that X is uniquely geodesic.

The image of a geodesic c with endpoints $x, y \in X$ is called a geodesic segment joining x and y, and is denoted by $[x, y]$. A geodesic triangle with vertices $x, y, z \in X$ is defined by $\triangle(x, y, z) = [y, z] \cup [z, x] \cup [x, y]$. For $\triangle(x, y, z) \subset X$, the comparison triangle $\triangle(\overline{x}, \overline{y}, \overline{z})$ is defined by the triangle in the 2-dimensional Euclidean space $\mathbb{E}^2$ with vertices $\overline{x}, \overline{y}, \overline{z} \in \mathbb{E}^2$ such that

$$d(y, z) = |\overline{y} - \overline{z}|_{\mathbb{E}^2}\,,\ d(z, x) = |\overline{z} - \overline{x}|_{\mathbb{E}^2}\,,\ d(x, y) = |\overline{x} - \overline{y}|_{\mathbb{E}^2}\,,$$

where $|\cdot|_{\mathbb{E}^2}$ is the Euclidean norm on $\mathbb{E}^2$. A point $\overline{p} \in [\overline{x}, \overline{y}]$ is called a comparison point for $p \in [x, y]$ if $d(x, p) = |\overline{x} - \overline{p}|_{\mathbb{E}^2}$. If for any $p, q \in \triangle(x, y, z)$ and their comparison points $\overline{p}, \overline{q} \in \overline{\triangle}(\overline{x}, \overline{y}, \overline{z})$, the inequality

$$d(p, q) \le |\overline{p} - \overline{q}|_{\mathbb{E}^2}$$

holds for all triangles in X, we call X a CAT(0) space. A Hadamard space is defined as a complete CAT(0) space.

For $x, y \in X$ and $t \in [0, 1]$, there exists a unique point $z \in [x, y]$ such that $d(x, z) = (1 - t)d(x, y)$ and $d(z, y) = td(x, y)$. We denote it by $tx \oplus (1 - t)y$. A subset C of X is said to be convex if $tx \oplus (1 - t)y \in C$ for every $x, y \in C$ and $t \in [0, 1]$. In a Hadamard space X, we know that the following inequality holds:

$$d(z, tx \oplus (1 - t)y)^2 \le td(z, x)^2 + (1 - t)d(z, y)^2 - t(1 - t)d(x, y)^2$$

for every $x, y, z \in X$ and $t \in [0, 1]$.

For a nonempty subset C of a Hadamard space X and $x \in X$, we define the distance $d(x, C)$ between x and C by

$$d(x, C) = \inf_{y \in C} d(x, y).$$

Suppose that C is nonempty, closed, and convex. Then, we know that for each $x \in X$, there exists a unique point $y_x \in C$ such that $d(x, y_x) = d(x, C)$. Using this fact, we define a mapping $P_C : X \to C$ by $P_C x = y_x$ for $x \in X$ and we call it the metric projection of X onto C.

The following result shows a relation between a decreasing sequence of closed convex subsets with respect to inclusion and the sequence of corresponding metric projections.

Theorem 7.1 (Kimura [4]) *Let $\{C_n\}$ be a sequence of nonempty closed convex subsets of a Hadamard space X and suppose that $C_{n+1} \subset C_n$ for all $n \in \mathbb{N}$. Let $u \in X$. If $C_0 = \bigcap_{n \in \mathbb{N}} C_n$ is nonempty, then the corresponding sequence $\{P_{C_n} u\}$ of metric projections to $\{C_n\}$ converges to $P_{C_0} u$.*

For more details of Hadamard spaces and their fundamental properties, see [2].

7.3 The Shrinking Projection Method

The shrinking projection method was originally proposed by Takahashi, Takeuchi, and Kubota [9] as an iterative method approximating a common fixed point of nonexpansive mappings defined on a subset of a Hilbert space. For a metric space X, a mapping $T : X \to X$ is said to be nonexpansive if

$$d(Tx, Ty) \leq d(x, y)$$

for all $x, y \in X$. We say that $z \in X$ is a fixed point of T if $z = Tz$, and we denote the set of fixed points of T by $\mathrm{Fix}\, T$.

For a nonexpansive mapping T defined on a closed convex subset of a Hadamard space X, $\mathrm{Fix}\, T$ is always closed and convex. Therefore, under the assumption that $\mathrm{Fix}\, T \neq \emptyset$, the metric projection $P_{\mathrm{Fix}\, T} : H \to \mathrm{Fix}\, T$ is defined. These properties also hold on a Hilbert space since a Hilbert space is an example of Hadamard spaces. The following is a convergence theorem with a simple version of the shrinking projection method.

Theorem 7.2 (Takahashi, Takeuchi, and Kubota [9]) *Let H be a Hilbert space and C a nonempty closed convex subset of H. Let $T : C \to C$ be a nonexpansive mapping such that $\mathrm{Fix}\, T \neq \emptyset$. Let $\{\alpha_n\}$ be a nonnegative real sequence such that $\sup_{n \in \mathbb{N}} \alpha_n < 1$. For an arbitrary point $u \in H$, generate a sequence $\{x_n\}$ by the following iterative scheme: $x_1 \in C$, $C_1 = C$, and*

$$
\begin{aligned}
&y_n = \alpha_n x_n + (1-\alpha_n)Tx_n,\\
&C_{n+1} = \{z \in H : \|y_n - z\| \le \|x_n - z\|\} \cap C_n,\\
&x_{n+1} = P_{C_{n+1}} u
\end{aligned}
$$

for $n \in \mathbb{N}$. Then, $\{x_n\}$ converges strongly to $P_{\mathrm{Fix}\,T} u \in C$.

This result was generalized to the setting of real Hilbert ball, a special case of Hadamard spaces, by Kimura [4]. We remark that the underlying space of this result can be changed to a Hadamard space satisfying that a set $\{x \in X : d(x,u) \le d(x,v)\}$ is always convex for $u, v \in X$ as follows:

Theorem 7.3 (Kimura [4]) *Let X be a Hadamard space such that $\{z \in X : d(u,z) \le d(v,z)\}$ is convex for every $u, v \in X$. Let $T : X \to X$ be a nonexpansive mapping such that $\mathrm{Fix}\,T \neq \emptyset$. Let $\{\alpha_n\}$ be a nonnegative real sequence in $[0,1]$ such that $\liminf_{n\to\infty} \alpha_n < 1$. For $u \in X$, generate an iterative sequence $\{x_n\}$ by $x_1 \in X$, $C_1 = X$, and*

$$
\begin{aligned}
&y_n = \alpha_n x_n \oplus (1-\alpha_n)Tx_n,\\
&C_{n+1} = \{z \in X : d(y_n, z) \le d(x_n, z)\} \cap C_n,\\
&x_{n+1} = P_{C_{n+1}} u
\end{aligned}
$$

for all $n \in \mathbb{N}$. Then $\{x_n\}$ converges to $P_{\mathrm{Fix}\,T} u \in X$.

In this method, we need to calculate a metric projection for a convex subset of X for every iteration, and it may be difficult to obtain its exact value in a practical computation. To overcome this difficulty, we can use the following result.

Theorem 7.4 (Kimura [5]) *Let X be a Hadamard space and suppose that a subset $\{z \in X : d(u,z) \le d(v,z)\}$ is convex for every $u, v \in X$. Let $T : X \to X$ be a nonexpansive mapping such that $\mathrm{Fix}\,T \neq \emptyset$. Let $\{\delta_n\}$ be a sequence of nonnegative numbers and $\delta_0 = \limsup_{n\to\infty} \delta_n$. For a given point $u \in X$, generate a sequence $\{x_n\}$ by $x_1 \in X$, $C_1 = X$, and*

$$
\begin{aligned}
&C_{n+1} = \{z \in X : d(Tx_n, z) \le d(x_n, z)\} \cap C_n,\\
&x_{n+1} \in C_{n+1} \text{ such that } d(u, x_{n+1})^2 \le d(u, C_{n+1})^2 + \delta_{n+1}^2
\end{aligned}
$$

for each $n \in \mathbb{N}$. Then,

$$\limsup_{n\to\infty} d(x_n, Tx_n) \le 2\delta_0.$$

Moreover, if $\delta_0 = 0$, then $\{x_n\}$ converges to $P_{\mathrm{Fix}\,T} u$.

7.4 Common Minimizers for a Family of Convex Functions

As we stated in the introduction, the resolvent operator for a convex function is defined in the setting of Hilbert spaces and more general spaces such as Banach spaces and Hadamard spaces. In this section, we first see the definition and fundamental properties of resolvents defined on Hadamard spaces. Then, we obtain an approximation result for a common minimizing problem for a finite family of convex functions.

Let X be a Hadamard space and $f : X \to]-\infty, +\infty]$. We say that f is proper if $f(x_0) < \infty$ for some $x_0 \in X$. f is said to be lower semicontinuous if

$$f(x_0) \leq \liminf_{n\to\infty} f(x_n)$$

whenever $\{x_n\} \subset X$ converges to $x_0 \in X$. f is said to be convex if for $x, y \in X$ and $\tau \in]0, 1[$,

$$f(\tau x \oplus (1-\tau)y) \leq \tau f(x) + (1-\tau) f(y)$$

holds.

Suppose that f is proper, lower semicontinuous, and convex. For $x \in X$, define $g_x : X \to X$ by $g(y) = f(x) + d(y, x)^2$ for $y \in X$. Then, we know that g_x has a unique minimizer $y_x \in X$. The resolvent operator $J_f : X \to X$ is defined by $J_f x = y_x$ for each $x \in X$, that is,

$$J_f x = \operatorname{argmin}_{y\in X}(f(y) + d(y, x)^2)$$

for $x \in X$. This definition was firstly given by Jost [3]. See also [8].

The resolvent operator has the following useful properties; see [1, 7].

- The set of fixed points of J_f coincides with the set of minimizers of f; $\operatorname{Fix} J_f = \operatorname{argmin}_{y\in X} f(y)$;
- J_f is firmly nonexpansive in the sense that

$$2d(J_f x, J_f y)^2 + d(J_f x, x)^2 + d(J_f y, y)^2 \leq d(J_f x, y)^2 + d(J_f y, x)^2$$

for all $x, y \in X$ and thus it is nonexpansive.

Since the set of fixed points of nonexpansive mappings on Hadamard spaces is closed and convex, so is $\operatorname{argmin}_{y\in X} f(y)$.

Using the notion of the resolvent for convex functions, we consider the problem of finding common minimizers for a finite family of convex functions. To take calculation errors for the metric projections into consideration, we employ the technique used in Theorem 7.4.

Theorem 7.5 *Let X be a Hadamard space and suppose that $\{x \in X : d(x, u) \leq d(x, v)\}$ is convex for every $u, v \in X$. For fixed $k \in \mathbb{N}$, let $\{f_j : X \to]-\infty, +\infty]$,*

$j = 0, 1, \ldots, k-1\}$ be a finite family of proper lower semicontinuous convex functions such that the set $M = \bigcap_{j=0}^{k-1} \operatorname{argmin}_X f_j$ of common minimizers of $\{f_j\}$ is nonempty. For a positive real sequence $\{\rho_n\}$ such that $\rho_0 = \liminf_{n\to\infty} \rho_n > 0$ and for $n \in \mathbb{N}$, let

$$J_n = J_{\rho_n f_{(n \bmod k)}},$$

where $J_{\rho_n f}$ is the resolvent of $\rho_n f$. For given $u, x_1 \in X$ with $d(u, x_1) \le \delta_1$, generate an iterative sequence $\{x_n\}$ as follows: $C_1 = X$,

$$\begin{aligned} C_{n+1} &= \{z \in X : d(J_n x_n, z) \le d(x_n, z)\} \cap C_n, \\ x_{n+1} &\in C_{n+1} \text{ such that } d(u, x_{n+1})^2 \le d(u, C_{n+1})^2 + \delta_{n+1}^2, \end{aligned}$$

where $\{\delta_n\}$ is a nonnegative real sequence. Let $\delta_0 = \limsup_{n\to\infty} \delta_n$. Then,

$$\limsup_{m\to\infty} f_j(J_{km+j} x_{km+j}) - \min_{y\in X} f_j(y) \le \frac{4\delta_0(2d(p,u) + \delta_0)}{\rho_0}$$

for all $j \in \{0, 1, \ldots, k-1\}$.

Moreover, if $\delta_0 = 0$, then $\{x_n\}$ converges to $P_M u \in \bigcap_{j=0}^{k-1} \operatorname{argmin}_{y\in X} f_j(y)$.

Proof We first prove the well-definedness of the sequence $\{x_n\}$ and $M \subset \bigcap_{n\in\mathbb{N}} C_n$ by induction. Note that $x_1 \in X$ is given and it is trivial that $M \subset C_1 = X$. For arbitrarily fixed $n \in \mathbb{N}$, we suppose that $x_n \in X$ is defined and $M \subset C_n$. Then, we have that

$$\begin{aligned} C_{n+1} \supset \operatorname{Fix} J_n \cap M &= \operatorname{argmin}_{y\in X} \rho_n f_{(n \bmod k)}(y) \cap M \\ &= \operatorname{argmin}_{y\in X} f_{(n \bmod k)}(y) \cap M \supset M \neq \emptyset. \end{aligned}$$

Thus there exists $x_{n+1} \in C_{n+1}$ such that

$$d(u, x_{n+1})^2 \le d(u, C_{n+1})^2 + \delta_{n+1}^2.$$

It follows that $\{x_n\}$ is well defined and $M \subset \bigcap_{n\in\mathbb{N}} C_n$.

It is easy to see that that every C_n is closed by the continuity of the metric d. We also know that C_n is convex from the assumption of the space. Hence $\{C_n\}$ is a decreasing sequence of nonempty closed convex subsets of X with respect to inclusion. Let $p_n = P_{C_n} u$ for $n \in \mathbb{N}$ and $p_0 = P_{C_0} u$, where $C_0 = \bigcap_{n\in\mathbb{N}} C_n$. Then, by Theorem 7.1 we have that $\{p_n\}$ converges to p_0. From the definition of the metric projections, we have that

$$d(u, x_n)^2 \le d(u, C_n)^2 + \delta_n^2 = d(u, p_n)^2 + \delta_n^2.$$

For $\tau \in]0, 1[$, it follows that

$$\begin{aligned} d(u, p_n)^2 &\le d(u, \tau p_n \oplus (1-\tau) x_n)^2 \\ &\le \tau d(u, p_n)^2 + (1-\tau) d(u, x_n)^2 - \tau(1-\tau) d(x_n, p_n)^2, \end{aligned}$$

and thus

$$\tau d(x_n, p_n)^2 \leq d(u, x_n)^2 - d(u, p_n)^2.$$

Tending $\tau \uparrow 1$, we have that

$$d(x_n, p_n)^2 \leq d(u, x_n)^2 - d(u, p_n)^2 \leq \delta_n^2,$$

and hence $d(x_n, p_n) \leq \delta_n$. Since $p_{n+1} \in C_{n+1}$, we have that

$$\begin{aligned} d(J_n x_n, x_n) &\leq d(J_n x_n, p_{n+1}) + d(p_{n+1}, x_n) \\ &\leq 2d(p_{n+1}, x_n) \\ &\leq 2(d(p_{n+1}, p_n) + d(p_n, x_n)) \\ &\leq 2(d(p_{n+1}, p_n) + \delta_n) \end{aligned}$$

for all $n \in \mathbb{N}$. Therefore, we obtain that

$$\limsup_{n\to\infty} d(J_n x_n, x_n) \leq 2\delta_0.$$

Let $p = P_M u$. Fix $j \in \{0, 1, \ldots, k-1\}$ arbitrarily. Then, for $m \in \mathbb{N}$, we have that

$$J_n = J_{km+j} = J_{\rho_n f_j},$$

where $n = km + j$. For $\tau \in]0, 1[$, we have that

$$\begin{aligned} &\rho_n f_j(J_n x_n) + d(J_n x_n, x_n)^2 \\ &\leq \rho_n f_j(\tau J_n x_n + (1-\tau)p) + d(\tau J_n x_n \oplus (1-\tau)p, x_n)^2 \\ &\leq \tau\rho_n f_j(J_n x_n) + (1-\tau)\rho_n f_j(p) \\ &\quad + \tau d(J_n x_n, x_n)^2 + (1-\tau)d(p, x_n)^2 - \tau(1-\tau)d(J_n x_n, p)^2. \end{aligned}$$

It follows that

$$\begin{aligned} &(1-\tau)\rho_n f_j(J_n x_n) - (1-\tau)\rho_n f_j(p) \\ &\qquad \leq (1-\tau)d(p, x_n)^2 - (1-\tau)d(J_n x_n, x_n)^2 - \tau(1-\tau)d(J_n x_n, p)^2. \end{aligned}$$

Dividing by $1-\tau$ and tending $\tau \uparrow 1$, we have that

$$\rho_n f_j(J_n x_n) - \rho_n f_j(p) \leq d(x_n, p)^2 - d(J_n x_n, x_n)^2 - d(J_n x_n, p)^2.$$

On the other hand, we have that

$$
\begin{aligned}
&d(x_n, p)^2 - d(J_n x_n, x_n)^2 - d(J_n x_n, p)^2 \\
&= (d(x_n, p) - d(J_n x_n, p))(d(x_n, p) + d(J_n x_n, p)) - d(J_n x_n, x_n)^2 \\
&\le d(J_n x_n, x_n)(d(x_n, p) + d(J_n x_n, p)) - d(J_n x_n, x_n)^2 \\
&= d(J_n x_n, x_n)(d(x_n, p) + d(J_n x_n, p) - d(J_n x_n, x_n)) \\
&\le 2d(J_n x_n, x_n) d(x_n, p) \\
&\le 4(d(p_{n+1}, p_n) + \delta_n)(d(p, u) + d(u, p_n) + d(p_n, x_n)) \\
&\le 4(d(p_{n+1}, p_n) + \delta_n)(2d(p, u) + \delta_n).
\end{aligned}
$$

Since $n = km + j$, we have that

$$
\begin{aligned}
&f_j(J_{km+j} x_{km+j}) - f_j(p) \\
&= f_j(J_n x_n) - f_j(p) \\
&\le \frac{4(d(p_{n+1}, p_n) + \delta_n)(2d(p, u) + \delta_n)}{\rho_n} \\
&\le \frac{4(d(p_{km+j+1}, p_{km+j}) + \delta_{km+j})(2d(p, u) + \delta_{km+j})}{\rho_{km+j}}.
\end{aligned}
$$

Since $f_j(p) = \min_{y \in X} f_j(y)$ and $\{p_n\}$ converges strongly to p_0, tending $m \to \infty$, we have that

$$
\begin{aligned}
&\limsup_{m \to \infty} f_j(J_{km+j} x_{km+j}) - \min_{y \in X} f_j(y) \\
&= \limsup_{m \to \infty} f_j(J_{km+j} x_{km+j}) - f_j(p) \\
&\le \limsup_{m \to \infty} \frac{4(d(p_{km+j+1}, p_{km+j}) + \delta_{km+j})(2d(p, u) + \delta_{km+j})}{\rho_{km+j}} \\
&\le \frac{4\delta_0(2d(p, u) + \delta_0)}{\rho_0}
\end{aligned}
$$

for any $j \in \{0, 1, \ldots, k-1\}$. Hence we obtain the desired result.

For the latter part of the theorem, suppose that $\delta_0 = 0$. Then we have that

$$
\lim_{n \to \infty} d(x_n, p_n) \le \lim_{n \to \infty} \delta_n = \delta_0 = 0.
$$

Since $\{p_n\}$ converges to p_0, so does $\{x_n\}$. We also have that

$$
\lim_{n \to \infty} d(J_n x_n, x_n) = \lim_{n \to \infty} 2\delta_n = 2\delta_0 = 0,
$$

$\{J_{\rho_n} x_n\}$ also converges to p_0. Since each f_j is lower semicontinuous, we have that

$$\begin{aligned} f_j(p_0) - \min_{y\in X} f(y) &\le \liminf_{m\to\infty} f(J_{km+j}x_{km+j}) - \min_{y\in X} f(y) \\ &\le \limsup_{m\to\infty} f(J_{km+j}x_{km+j}) - \min_{y\in X} f(y) \\ &= \frac{4\delta_0(2d(u,\,p)+\delta_0)}{\rho_0} \\ &= 0. \end{aligned}$$

Therefore, $p_0 \in \operatorname{argmin}_X f_j$ for all $j \in \{0, 1, \ldots, k-1\}$ and it follows that $p_0 \in M$. Since $p_0 = P_{C_0}u$ and $M \subset C_0$, we have that

$$p_0 = P_M u \in M = \bigcap_{j=0}^{k-1} \operatorname{argmin}_{y\in X} f_j(y),$$

which completes the proof.

In the end of this chapter, we remark some recent development for this theory. The notion of resolvent for convex functions has been generalized to that defined on a complete CAT(1) space [6]. It is also obtained that this new resolvent operator has useful properties called firm spherical nonspreadingness, which is an analogy to firm nonexpansiveness of the resolvent defined on Hadamard spaces. By using this operator, we may obtain various kinds of approximation schemes for the convex minimization problem on complete CAT(1) spaces.

References

1. Bačák, M.: Convex Analysis and Optimization in Hadamard Spaces. De Gruyter Series in Nonlinear Analysis and Applications, vol. 22. De Gruyter, Berlin (2014)
2. Bridson, M.R., Haefliger, A.: Metric Spaces of Non-positive Curvature. Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences], vol. 319. Springer, Berlin (1999)
3. Jost, J.: Convex functionals and generalized harmonic maps into spaces of nonpositive curvature. Comment. Math. Helv. **70**, 659–673 (1995)
4. Kimura, Y.: Convergence of a sequence of sets in a Hadamard space and the shrinking projection method for a real Hilbert ball. Abstr. Appl. Anal. Art. ID **582475**, 11 (2010)
5. Kimura, Y.: A shrinking projection method for nonexpansive mappings with nonsummable errors in a Hadamard space. Ann. Oper. Res. **243**, 89–94 (2016)
6. Kimura, Y., Kohsaka, F.: Spherical nonspreadingness of resolvents of convex functions in geodesic spaces. J. Fixed Point Theory Appl. **18**, 93–115 (2016)
7. Kimura, Y., Kohsaka, F.: Two modified proximal point algorithms for convex functions in Hadamard spaces. Linear Nonlinear Anal. **2**, 69–86 (2016)
8. Mayer, U.F.: Gradient flows on nonpositively curved metric spaces and harmonic maps. Comm. Anal. Geom. **6**, 199–253 (1998)
9. Takahashi, W., Takeuchi, Y., Kubota, R.: Strong convergence theorems by hybrid methods for families of nonexpansive mappings in Hilbert spaces. J. Math. Anal. Appl. **341**, 276–286 (2008)

Chapter 8
Some Hard Stable Marriage Problems: A Survey on Multivariate Analysis

Sushmita Gupta, Sanjukta Roy, Saket Saurabh and Meirav Zehavi

8.1 Introduction

Matching under preferences is a rich topic central to both economics and computer science, which has been consistently and intensively studied for over several decades. One of the main reasons for interest in this topic stems from the observation that it is extremely relevant to a wide variety of practical applications modeling situations where the objective is to *match* agents to other agents (or to resources). In the most general setting, a matching is defined as an allocation (or assignment) of agents to resources that satisfies some predefined criterion of compatibility/acceptability. Here, the (arguably) best-known model is the *two-sided model*, where the agents on one side are referred to as *men*, and the agents on the other side are referred to as *women*. A few illustrative examples of real-life situations where this model is employed in practice include matching hospitals to residents, students to colleges, kidney patients to donors and users to servers in a distributed Internet service. At the heart of all of these applications lies the fundamental STABLE MARRIAGE problem. In particular, the Nobel Prize in Economics was awarded to Shapley and Roth in 2012 "for the theory of stable allocations and the practice of market design." Moreover, several books have been dedicated to the study of STABLE MARRIAGE as well as optimization variants of this classical problem such as the EGALITARIAN STABLE MARRIAGE, SEX-EQUAL STA-

S. Gupta · M. Zehavi
University of Bergen, Bergen, Norway
e-mail: sushmita.gupta@uib.no

M. Zehavi
e-mail: meirav.zehavi@uib.no

S. Roy · S. Saurabh (✉)
The Institute of Mathematical Sciences, HBNI, Chennai, India
e-mail: saket@imsc.res.in

S. Roy
e-mail: sanjukta@imsc.res.in

S. K. Neogy et al. (eds.), *Mathematical Programming and Game Theory*,
Indian Statistical Institute Series, https://doi.org/10.1007/978-981-13-3059-9_8

BLE MARRIAGE, BALANCED STABLE MARRIAGE, MAXIMUM (MINIMUM) STABLE MATCHING WITH TIES and STABLE MATCHING MANIPULATION problems [1–3]. A solution to STABLE MARRIAGE problem can be computed in polynomial time [4]. However, these variants of STABLE MARRIAGE problem except EGALITARIAN STABLE MARRIAGE are NP-hard. Consequently, different ways to cope with the hardness has been looked at for these problems.

In this article, we survey works on NP-hard variants of problems related to STABLE MARRIAGE in the area of exact exponential time algorithms. In particular, we look at these problems through the lens of Parameterized Complexity, a finer notion of complexity for NP-hard problems. We will introduce the fundamentals of Parameterized Complexity in the next section. Prior to that, we define the problems we study in this article.

8.1.1 Stable Matching and Its Variants

In algorithmic game theory, it is common to model games in terms of graph theory terminology. Primarily inspired by the seminal work of Myerson [5] on cooperative game theory, this has become a standard practice, especially but not limited to the study of *assignment problems* of which STABLE MARRIAGE, STABLE ROOMMATES are special cases. In this model vertices of a graph are used to represent players, and edges represent some notion of compatibility or acceptability among pairs of players.

Stable Marriage. The input of the STABLE MARRIAGE (SM) problem consists of a set of men, M, and a set of women, W, each person ranking a subset of people of the opposite gender, modeled as a bipartite graph $G = (M, W, E)$. That is, each person a has a set of *acceptable partners*, $\mathcal{A}(a)$, whom this person subjectively ranks. Consequently, each person a has a so-called *preference list*, where $p_a(b)$ denotes the position of $b \in \mathcal{A}(a)$ in a's preference list. Without loss of generality, it is assumed that if a person a ranks a person b, then the person b ranks the person a as well. The sets of preference lists of the men and the women are denoted by $\mathcal{L}^M$ and $\mathcal{L}^W$, respectively. In this context, we say that a pair of a man and a woman, (m, w), is an *acceptable pair* if both $m \in \mathcal{A}(w)$ and $w \in \mathcal{A}(m)$ (equivalently, $(m, w) \in E$). Accordingly, the notion of a *marriage* refers to a matching between men and women, where two people that are matched to one another form an acceptable pair. Roughly speaking, the goal of the STABLE MARRIAGE problem is to *find* a matching that is *stable* in the following sense: there should not exist two people who prefer being matched to each other over their current "status". More precisely, a matching μ is said to be stable if it does not have a *blocking pair*, which is an acceptable pair (m, w) such that **(i)** either m is unmatched by μ or $p_m(w) < p_m(\mu(m))$, and **(ii)** either w is unmatched by μ or $p_w(m) < p_w(\mu(w))$. Here, the notation $\mu(a)$ represents the person to whom μ matches the person a. Note that a person always prefers being matched to an acceptable partner to being unmatched. Keeping in line with the graph

terminology, we will refer to a blocking pair as a *blocking edge*. We denote the set of all stable matchings by SM.

Stable Roommate. When the underlying graph $G = (V, E)$ is not necessarily a bipartite graph, then the STABLE MARRIAGE problem is known as the STABLE ROOMMATE problem. For each vertex $v \in V$, $\mathcal{L}(v)$ is a strict ranking over the set of the neighbors of v in G, denoted by $N(v)$. Quite clearly, the notions of acceptable pair, and stability are well defined even for the *roommate* setting.

8.1.1.1 Stability in Presence of Ties

When the preference lists are not strict orderings but can contain ties, then in the case of a bipartite graph, we have STABLE MARRIAGE WITH TIES and STABLE ROOMMATE WITH TIES otherwise. Moreover, the preference lists may not be complete, i.e., the underlying graph is not complete, then the problems are called STABLE MARRIAGE WITH TIES AND INCOMPLETE LISTS (SMTI) and STABLE ROOMMATE WITH TIES AND INCOMPLETE LISTS (SRTI), respectively.

In the presence *ties*, there are multiple notions of stability: *weak*, *super*, and *strong* [6, 7]. A matching μ is said to be *(weakly) stable* if there do not exist blocking edges. The above models depict many real-life situations where solutions have to satisfy certain predefined criterion of suitability and compatibility. Every instance of SMTI has a stable matching, and such a matching can be found in polynomial time [8]: Simply break ties arbitrarily and run Gale–Shapley algorithm to find a stable matching in the new instance. The resulting matching is weakly stable in the original instance. However, note that, there can be an *exponential* number of stable matchings in a given instance [1, Theorem 1.3.3, pg 24]. Furthermore, the manner in which the ties are broken can affect the size of the resulting stable matching, up to a factor of 2. Depending on the application at hand, some of these (exponentially many) matchings might be better suited than others. The two (arguably) most natural objectives are to maximize or minimize the size of the matching as it might be desirable to maintain stability while either maximizing or minimizing the use of available "resources". These objectives define the well-known NP-hard variants of SM problem, namely the MAX- SMTI and MIN- SMTI problems [8].

8.1.2 Stability and Equality

Gale and Shapley in the 1960s [4], while analyzing a heuristic that was in use for over a decade to match medical residents to teaching hospitals in the Boston area under the National Resident Matching Program (NRMP), showed that every instance of the STABLE MARRIAGE problem admits a stable matching. The heuristic has since come to be known as the famous Gale–Shapley algorithm works in polynomial time and can be used to find a stable matching. In other words, given any set of preference lists

of men and women there exists at least one stable matching and as many as exponential number of stable matchings, and they should be viewed as a *spectrum* where the two extremes are known as the *man-optimal stable matching* and the *woman-optimal stable matching*. The man-optimal stable matching, denoted by μ_M, is a stable matching such that every stable matching μ satisfies the following condition: every man m is either unmatched by both μ_M and μ or $p_m(\mu_M(m)) \leq p_m(\mu(m))$. The woman-optimal stable matching, denoted by μ_W, is defined analogously. These two extremes, which give the best possible solution for one party at the expense of the other party, always exist and can be computed in polynomial time [4].

Gale–Shapley Algorithm. This algorithm exists in two versions: men-proposing and women-proposing. The version that yields the man (woman)-optimal stable matching is the men-proposing (women-proposing) Gale–Shapley algorithm. It has been customary to use the men-proposing version of the algorithm, and our discussion in this survey will stick to that convention. In the next paragraph, we will only describe the men-proposing version of the algorithm; the other one can be described analogously.

A man who is currently unmatched to any woman, *proposes* to the woman who is at the top of his current list, which is obtained by removing from his original preference list, all the women who have rejected him at an earlier step. On the woman's side, when a woman w receives a proposal from a man m, she accepts the proposal if it is her first proposal, or if she prefers m to her current partner. If w prefers her current partner to m, then w *rejects* m. If m is rejected by w, then m removes w from his list. This process continues until every man is either matched or his preference list is empty. The output of this algorithm is the men-optimal stable matching. For more details, see [1] Gusfeld and Irving's authoritative treatise on stable matching. Let $(\mathcal{L}^M, \mathcal{L}^W)$ denote the set of preference lists of men and women, and the men-optimal stable matching with respect to these lists are denoted by $\text{GS}(\mathcal{L}^M, \mathcal{L}^W)$. Henceforth, unless explicitly stated otherwise, any mention of a stable matching should be interpreted by the reader as the man-optimal stable matching.

Naturally, it is desirable to analyze stable matchings that lie somewhere in the middle of the two extremes, being *globally desirable*, *fair towards both sides* or *desirable by both sides*. Each of these notions yields a desirable stable matching that leads to a natural, *different* optimization problem. The determination of which notion best describes an appropriate outcome depends on the specific situation at hand. Here, the value $p_a(\mu(a))$ is viewed as the "satisfaction" of a in a matching μ, where a smaller value signifies a greater amount of satisfaction. Under this interpretation, the *egalitarian stable matching* attempts to be *globally desirable* by minimizing $e(\mu) = \sum_{(m,w)\in\mu}(p_m(\mu(m)) + p_w(\mu(w)))$ over the set of all stable matchings (recall that we denote it by SM). The problem of finding an egalitarian stable matching, called EGALITARIAN STABLE MARRIAGE, is known to be solvable in polynomial time due to Irving et al. [9]. Roughly speaking, this problem does not distinguish between men and women, and therefore, it does not fit scenarios where it is necessary to differentiate between the individual satisfaction of each party. In such scenarios,

the SEX-EQUAL STABLE MARRIAGE and BALANCED STABLE MARRIAGE problems come into play.

In the SEX-EQUAL STABLE MARRIAGE problem, the objective is to find a stable matching that minimizes the absolute value of $\delta(\mu)$ over SM, where $\delta(\mu) = \sum_{(m,w)\in\mu} p_m(\mu(m)) - \sum_{(m,w)\in\mu} p_w(\mu(w))$. It is thus clear that SEX-EQUAL STABLE MARRIAGE seeks a stable matching that is fair toward both sides by minimizing the difference between their individual amounts of satisfaction. Unlike the EGALITARIAN STABLE MARRIAGE, the SEX-EQUAL STABLE MARRIAGE problem is known to be NP-hard [10]. On the other hand, in BALANCED STABLE MARRIAGE, the objective is to find a stable matching that minimizes $\mathsf{balance}(\mu) = \max\{\sum_{(m,w)\in\mu} p_m(w), \sum_{(m,w)\in\mu} p_w(m)\}$ over SM. At first sight, this measure might seem conceptually similar to the previous one, but in fact, the two measures are quite different. Indeed, BALANCED STABLE MARRIAGE does not attempt to find a stable matching that is fair, but one that is desirable by both sides. In other words, BALANCED STABLE MARRIAGE examines the amount of dissatisfaction of each party *individually*, and attempts to minimize the worse one among the two. This problem fits the common scenario in economics where each party is selfish in the sense that it desires a matching where its own dissatisfaction is minimized, irrespective of the dissatisfaction of the other party, and our goal is to find a matching desirable by both parties by ensuring that each individual amount of dissatisfaction does not exceed some threshold. In some situations, the minimization of $\mathsf{balance}(\mu)$ may indirectly also minimize $\delta(\mu)$, but in other situations, this may not be the case. Indeed, McDermid [11] constructed a family of instances where there does *not* exist any matching that is both a sex-equal stable matching and a balanced stable matching (the construction is also available in the book [3]).

We study BALANCED STABLE MARRIAGE problem in the realm of fast exact exponential time algorithms as defined by the field of Parameterized Complexity (see Sect. 8.2). Recall that SM is the set of all stable matchings. In this context, we would like to remark that McDermid and Irving [12] showed that SEX-EQUAL STABLE MARRIAGE is NP-hard even if it is only necessary to decide whether the target $\Delta = \min_{\mu\in\mathrm{SM}} |\delta(\mu)|$ is 0 or not [12]. In particular, this means that SEX-EQUAL STABLE MARRIAGE is not only W[1]-hard with respect to Δ, but it is even paraNP-hard with respect to this parameter.[1] In the case of BALANCED STABLE MARRIAGE, however, fixed-parameter tractability with respect to the target $\mathsf{Bal} = \min_{\mu\in\mathrm{SM}} \mathsf{balance}(\mu)$ trivially follows from the fact that this value is lower bounded by $\max\{|M|, |W|\}$.[2]

[1] If a parameterized problem cannot be solved in polynomial time even when the value of the parameter is a fixed constant (that is, independent of the input), then the problem is said to be paraNP-hard.

[2] In the analysis of the BALANCED STABLE MARRIAGE, it is assumed that any stable matching is perfect.

8.1.3 Optimization Variants

Cseh and Manlove [13] studied NP-hard variants of the STABLE MARRIAGE and STABLE ROOMMATE problems[3]: the input consists of preference lists of every agent, and two subsets of (not necessarily pairwise disjoint) pairs of agents, representing the set of *forbidden pairs* and *forced pairs*. The objective is to find a matching that does not contain any of the forbidden pairs, and contains each of the forced pairs, while simultaneously minimizing the number of blocking pairs in the matching. Mnich and Schlotter [14] studied a variant of this problem where a subset of women and as a subset of men are termed *distinguished*, and the objective is to find a matching with fewest number of blocking pairs that matches all of the distinguished men and women. They consider three parameters to determine the computational tractability of this problem: the maximum length of the preference lists for men and women, the number of distinguished men and women, and the number of blocking pairs allowed in a given instance. A complete trichotomy of computational complexity of the problem is exhibited with respect to these three parameters: polynomial-time solvable, NP-hard and fixed-parameter tractable, and W[1]-hard, respectively.

8.1.4 Manipulation

Strategic manipulation of matching algorithms is a rich area of research on matchings. Working on this topic, specifically with regards to stable matching algorithms, goes back several decades and is anchored on the fact that there are no stable matching algorithms that are *strategyproof*. Informally stated, it means that for any stable matching algorithm, there are instances in which at least some players have an incentive to misrepresent their true preferences to obtain a strictly better outcome for themselves. In the case of STABLE MARRIAGE problems, the misrepresentation takes the form of stating a smaller list of acceptable partners, and/or permuting one's true preference list.

Kobayashi and Matsui [15] studied manipulation of the Gale–Shapley algorithm, where a coalition of agents manipulate with the goal of attaining specific matching partners. Formally speaking, an input consists of the usual preference lists for men and women $\mathcal{L}^M$ and $\mathcal{L}^W$, and a matching; this matching can either be *perfect* (if it contains $n = |M| = |W|$ pairs) or *partial* (possibly, fewer than n pairs). Furthermore, for a couple of problems, we are given a set of preference lists for a subset of women, $\mathcal{L}^{W'}$, where $W' \subseteq W$. The goal is to decide if there exists a set of preference lists for all the women, $\mathcal{L}^W$ that contains $\mathcal{L}^{W'}$, such that when used in conjunction with $\mathcal{L}^M$, the Gale–Shapley man-proposing algorithm yields a matching that contains all the

[3] In STABLE ROOMMATE, the matching market consists of agents of the same type, as opposed to the market modeled the stable marriage problem that consists of agents of two types, men and women. Roommate assignments in college housing facilities is a real-world application of the stable roommate problem.

pairs in the stated matching. Next, we consider two of these problems, and compare and contrast their computational complexity.

ATTAINABLE STABLE MATCHING (ASM)
Input: A set of preference lists $\mathcal{L}^M$ of men over women W, and a perfect matching μ on (M, W).
Question: Does there exist a set of preference lists of women, denoted by $\mathcal{L}^W$, such that $\text{GS}(\mathcal{L}^M, \mathcal{L}^W) = \mu$?

Kobayashi and Matsui in [15] showed that ASM is polynomial-time solvable, and exhibited an $\mathcal{O}(n^2)$ algorithm that computes the set $\mathcal{L}^W$, if one exists. Or else, reports "none exists". Note that the following problem, SEOPM, is identical to ASM, except in one key aspect: *the target matching, denoted by μ, need not be perfect*. The authors show that SEOPM is NP-complete.

STABLE EXTENSION OF PARTIAL MATCHING (SEOPM)
Input: A set of preference lists $\mathcal{L}^M$ of men M over women W, and a partial matching μ on (M, W).
Question: Does there exist preferences of women, denoted by $\mathcal{L}^W$, such that $\mu \subseteq \text{GS}(\mathcal{L}^M, \mathcal{L}^W)$?

These two problems and their differing computational complexities represent a dichotomy with respect to the size of the target matching. Kobayashi and Matsui solve ASM by designing a novel combinatorial structure called the *suitor graph*, which encodes enough information about the men's preferences and the matching pairs in μ, that it allows an efficient search of the possible preference lists of women, which are $n \cdot n!$ in number. The same approach falls short when the target matching is partial.

8.2 Parameterized Complexity

A *parameterization* of a problem P is the association of an integer k with each input instance of P resulting in a *parameterized problem* $\Pi = (P, k)$. Intuitively, the parameter bounds any secondary information known about the problem or the input excluding the size of the input instance. The goal of parameterization is to investigate the complexity of the problem in terms of the input size as well as the parameter. For the purpose of this article, we use three basic concepts of Parameterized Complexity: *kernelization*, *fixed parameter tractability*, and W-*hardness*.

8.2.1 Kernelization

A *kernelization algorithm* for a parameterized problem $\Pi = (P, k)$ translates any input instance (I, k) of Π into an "equivalent instance" (I', k')[4] of Π such that the size of I' is bounded by $f(k)$ and $k' = g(k)$ for some computable functions f and g that *only* depend on k. A parameterized problem Π is said admit a *kernel* of size $f(k)$ if there exists a polynomial-time kernelization algorithm. In case the function f is polynomial in k, Π is said to admit a *polynomial kernel*. Thus, kernelization is seen as a mathematical concept that aims to analyze the power of preprocessing procedures in a formal and rigorous manner.

8.2.2 Fixed-Parameter Tractability

A parameterized problem $\Pi = (I, k)$ is said to be *fixed parameter tractable* (FPT) if there is an algorithm that solves it in time $f(k) \cdot n^{\mathcal{O}(1)}$, where n is the size of the input and f is a function that depends only on k. Such an algorithm is called a *parameterized algorithm*. In other words, the notion of FPT signifies that there is an algorithm that limits the combinatorial explosion in the running time to the parameter k and only allows a polynomial dependence on the input size n.

It is known that if a parameterized problem is FPT, then it admits a kernel, and vice versa. Thus, kernelization can be another way of defining fixed-parameter tractability.

8.2.3 W-Hardness

Parameterized Complexity also provides tools to refute the existence of polynomial kernels and FPT algorithms for certain problems (under plausible complexity-theoretic assumptions). In this context, the W-hierarchy of Parameterized Complexity is analogous to the polynomial hierarchy of classical Complexity Theory. It is widely believed that a problem that is W[1]-hard is unlikely to be FPT, and we refer the reader to the books [16, 17] for more information on this notion in particular, and on Parameterized Complexity in general.

[4]Two instances $\mathscr{I}$ and $\mathscr{J}$ are said to be equivalent if $\mathscr{I}$ is a YES-instance if and only if $\mathscr{J}$ is a YES-instance.

8.2.4 Exact Exponential Algorithms

An algorithm whose running time is expressible entirely in terms of the size of input instance such as $f(n)$ is called an *exact algorithm*. Specifically, if f is an exponential function in n (such as $f(n) = c^n$ for some constant c), then the algorithm is known as an *exact exponential algorithm*.

The notation $\mathcal{O}^*$ is used to hide factors polynomial in the input size.

8.3 Three Problems

Since BALANCED STABLE MARRIAGE, MAXIMUM (MINIMUM) STABLE MARRIAGE WITH TIES, and STABLE MATCHING MANIPULATION problems have been shown to be NP-complete [8, 15, 18], it is natural to study these problems in computational paradigms that are meant to cope with NP-hardness.

8.3.1 STABLE MATCHING MANIPULATION

Manipulation and strategic issues in voting have been well studied in the field of Exact Algorithms and Parameterized Complexity; survey [19] provides an overview. But one cannot say the same regarding the strategic issues in the stable matching model. These problems hold a lot of promise and remain hitherto unexplored in the light of exact algorithms and parameterized complexity, with exceptions that are few and far between [20, 21].

There is a long history of research on manipulation of the Gale–Shapley algorithm by one or more agents working individually or in a coalition. The objective is to misstate the true preference lists (either by truncating, or by permuting the list), to obtain a better partner (in terms of the true preferences) than would be otherwise possible under the Gale–Shapley algorithm.

The SEOPM problem (defined in Sect. 8.1.4) can be viewed as a manipulation game in which a coalition of agents—the subset of women who are matched under the partial matching μ'—called *manipulating agents* have fixed their partners. These agents are colluding, with cooperation from the other women who are not matched in the partial matching, to produce a matching that matches every agent (called a *perfect matching*) while matching each of the manipulating agents to their target partners. There exists a strategy to attain this objective if and only if there exists a set of preference lists of women that yields a perfect matching using Gale–Shapley algorithm that contains the partial matching.

Recall that we have $n = |M| = |W|$. The most basic algorithm for SEOPM would be to generate the preference list of a woman by enumerating *all* possible permutations of men, $n!$ of them. Thus, a total of $n \cdot n!$ possible choices for the set of

preference lists for n women, denoted by $\mathcal{L}^W$; and then check whether the partial matching μ' is contained in the matching $\text{GS}(\mathcal{L}^M, \mathcal{L}^W)$, obtained by applying the Gale–Shapley algorithm to $(\mathcal{L}^M, \mathcal{L}^W)$.

However, this algorithm will have a time complexity of $(n!)^n n^2 = 2^{\mathcal{O}(n^2 \log n)}$. One can improve over this naïve algorithm by using the polynomial-time algorithm by Kobayashi and Matsui for ASM [15]. That is, using the algorithm for ASM, in which given a matching μ can check in polynomial time whether there exists a set of preference lists of women $\mathcal{L}^W$ such that $\mu = \text{GS}(\mathcal{L}^M, \mathcal{L}^W)$. The faster algorithm for SEOPM, using the algorithm for ASM as a subroutine, tries all possible extensions μ of the partial matching μ' and checks in polynomial time whether there exists set of preference lists for women, $\mathcal{L}^W$, such that $\mu = \text{GS}(\mathcal{L}^M, \mathcal{L}^W)$. Thus, if the size of the partial matching is k, then this algorithm would have to try $(n - k)!$ possibilities. In the worst case this can take time $(n!)n^{\mathcal{O}(1)} = 2^{\mathcal{O}(n \log n)}$.

Gupta and Roy [22, 23] give an exact-exponential time algorithm of running time $2^{\mathcal{O}(n)}$ for SEOPM. Clearly, this improves the time complexity established by the naïve algorithm. It relates SEOPM to the problem of Colored Subgraph Isomorphism, where we are given two graphs G and H and a coloring $\chi : V(G) \to \{1, 2, \ldots, |V(H)|\}$, and the objective is to test whether H is isomorphic to some subgraph of G whose vertices have distinct colors. The connection between SEOPM and Colored Subgraph Isomorphism is established by introducing a combinatorial tool, the *universal suitor graph* that extends the notion of the rooted *suitor graph* devised by Kobayashi and Matsui in [15], to solve ASM. It is shown in [15] that an input instance $(\mathcal{L}^M, \mu)$ of ASM is a Yes-instance if and only if the corresponding rooted suitor graph has an *out-branching*: a spanning subgraph in which every vertex has at most one incoming arc, and is reachable from the root. The universal suitor graph satisfies the property that an instance of SEOPM $(\mathcal{L}^M, \mu')$ is a Yes-instance if and only if the corresponding universal suitor graph contains a subgraph that is isomorphic to the out-branching corresponding to $(\mathcal{L}^M, \mu)$ where μ is the perfect matching that "extends" μ'. In this manner, the universal suitor graph succinctly encodes all "possible suitor graphs" and is only polynomially larger than the size of a suitor graph. That is, the size of universal suitor graph is $\mathcal{O}(n^2)$.

Using ideas from the world of exact exponential time algorithms and Parameterized Complexity Gupta and Roy [22, 23] search for a subgraph in the universal suitor graph that is isomorphic to an out-branching corresponding to an extension of μ'. In particular, their algorithm uses a subroutine that enumerates all non-isomorphic out-branchings in a (given) rooted directed graph [24, 25], and a parameterized algorithm for Colored Subgraph Isomorphism [26, 27]. Moveover, it is shown that unless the Exponential Time Hypothesis (ETH) fails [28], their algorithm is asymptotically optimal. That is, unless ETH fails, there is no algorithm for SEOPM with running time $2^{o(n)}$.

8.3.2 Maximum (Minimum) Stable Marriage with Ties

Irving, Iwama et al. [8] showed that Max- SMTI is NP-hard even if inputs are restricted to having ties only in the preference lists of men, preference lists of bounded length, and symmetry in preference lists. Thus, it is natural to study Max- SMTI from the perspectives of Parameterized Complexity.

Marx and Schlotter [20] study Max- SMTI using the local search approach. They consider the following parameters: (i) the maximum number of ties in an instance ($\kappa_1(\mathrm{i})$); (ii) the maximum length of ties in an instance ($\kappa_2(\mathrm{i})$); (iii) the total length of the ties in an instance ($\kappa_3(\mathrm{i})$). Furthermore, it is shown that Max- SMTI is W-hard parameterized by $\kappa_1(\mathrm{i})$, and FPT when parameterized by $\kappa_3(\mathrm{i})$. Since it is known that Max- SMTI is NP-hard when the length of each tie is at most 2 [8], there cannot exist an algorithm with running time $f(\kappa_2(\mathrm{i}))n^{g(\kappa_2(\mathrm{i}))}$, for any functions f and g that depend only on k unless ¶= NP. This motivates us to study this problem with larger parameter such as the solution size of the problem.

Adil, Gupta et al. [29] study the parameterized complexity of NP-hard optimization versions of Stable Matching and Stable Roommates in the presence of ties and incomplete lists. Specifically, the following problems are studied.

Max(resp. Min)- SMTI
Input: A bipartite graph $G = (M \cup W, E)$, and two families of preference lists, $\mathcal{L}^M$ and $\mathcal{L}^W$ and a non-negative integer k.
Question: Find (if there) exists a weakly stable matching of size at least k (resp. at most k).
Parameter: k

Max- SRTI is defined as follows.

Max- SRTI
Input: A graph $G = (V, E)$, the family of preference lists $\mathcal{L}^V$, the size of a maximum matching ℓ, and a positive integer k.
Question: Find (if there) exists a weakly stable matching of size at least k.
Parameter: ℓ

The parameter ℓ. The reason k is not an appropriate parameter for Max- SRTI which follows from the fact that the decision version of the SRTI problem, that is, whether there exists a stable matching is NP-hard [30]. Similar to Max- SMTI, an approach to solve SRTI is by breaking the ties of the instance of SRTI arbitrarily. We know that if a matching is stable in the new instance, then it is stable in the original instance. However, some ordering of the ties may create an instance with no stable matching while some other ordering of the ties may produce an instance which has a stable matching. It is computationally hard to decide how to break the ties in order to test the existence of a stable matching. Thus, there does not exist (unless ¶= NP) any algorithm for Max- SRTI which runs in time of the form $f(k) \cdot |V|^{\mathcal{O}(1)}$ (or even $|V|^{f(k)}$) where function f depends only on k. Indeed, we could set $k = 1$, employ such an algorithm to test whether there *exist* a stable matching of size at

least 1 in polynomial time, thereby contradicting the result that the decision version of the SRTI problem is NP-hard. Consequently, we need to look for an alternate parameter. Toward this, we observe that a stable matching, if one exists, is a *maximal matching* in the underlying graph. Furthermore, the size of any maximal matching is at least $\ell/2$, where ℓ is the size of a maximum matching. Thus, if a stable matching exists, then the size of such a matching differs from the value of ℓ by a factor of at most 2. This leads directly to the parameterization of MAX- SRTI by ℓ instead of the solution size.

The main result of Adil, Gupta et al. [29] is that the above hard variants of STABLE MATCHING and STABLE ROOMMATES, that is, MAX(RESP. MIN)- SMTI and SRTI admit polynomial sized kernels. It implies that MAX- SMTI (MIN- SMTI) is FPT with respect to solution size, and MAX- SRTI is FPT with respect to a structural parameter. Additionally, they show that MAX- SMTI, MIN- SMTI and MAX- SRTI, when parameterized by the treewidth **tw** of the input graph, admit algorithms with running time $n^{\mathcal{O}(\mathbf{tw})}$.

Adil, Gupta et al. [29] design FPT algorithms using the small kernel as follows. First, obtain an equivalent instance by applying the kernelization algorithm, where the output graph $G' = (M' \cup W', E')$ has $\mathcal{O}(k^2)$ edges. This implies that the sum of the sizes of preference lists of any agent (men (M') or women (W')) is $\mathcal{O}(k^2)$. Then, they enumerate all subsets $E'' \subseteq E'$ of edges of size q in G', where $k \leq q \leq 2k$ and test if E'' is a solution for MAX- SMTI. Since $\sum_{q=k}^{2k} \binom{|E'|}{q} \leq \sum_{q=k}^{2k} \left(\frac{e|E'|}{q}\right)^q = |E'|^{\mathcal{O}(k)} = 2^{\mathcal{O}(k)\log|E'|} = 2^{\mathcal{O}(k\log k)}$, the running time of the algorithm is $2^{\mathcal{O}(k\log k)}$. To solve MIN- SMTI, they enumerate all subsets of edges of size at most k in G'; again, the running time is $\sum_{q=1}^{k} \binom{|E'|}{k} = 2^{\mathcal{O}(k\log k)}$. In both cases, for every subset of edges, the test whether it is a stable matching can be conducted in polynomial time. Overall, it is shown that both MAX- SMTI and MIN- SMTI admit a kernel of size $\mathcal{O}(k^2)$, and exhibit an algorithm with running time $2^{\mathcal{O}(k\log k)} + n^{\mathcal{O}(1)}$. In addition, MAX- SRTI admits a kernel of size $\mathcal{O}(\ell^2)$, and exhibit an algorithm with running time $2^{\mathcal{O}(\ell\log\ell)} + n^{\mathcal{O}(1)}$.

Many combinatorial problems that are computationally hard for general graphs, are known to be easier on planar graphs. Moreover, planar graphs are extensively studied in real-life applications. However, since it is known that MIN MAXIMAL MATCHING is NP-hard on planar cubic graphs [31], the reduction by Irving, Manlove et al. in [32, Section 4, Theorem 6] directly implies that MAX- SMTI, MIN- SMTI and MAX- SRTI are NP-hard on planar graphs. This leads us to question: whether or not, MAX- SMTI, MIN- SMTI and MAX- SRTI admit smaller kernels on planar graphs than those known for general graphs. In a similar spirit of research, Peters [33] has recently explicitly asked to study graphical hedonic games (which subsume matching problems such as SM and STABLE ROOMMATE) on bipartite, planar and H-minor free graph topologies. Adill, Gupta et al., [29] showed that for this restricted class of input MAX- SMTI (MIN- SMTI) admits a kernel of size $\mathcal{O}(k)$ and an algorithm running in time $2^{\mathcal{O}(\sqrt{k}\log k)} + n^{\mathcal{O}(1)}$. They also proved that MAX- SRTI on planar graphs admits a kernel of size $\mathcal{O}(\ell)$ and an algorithm running in time $2^{\mathcal{O}(\sqrt{\ell}\log\ell)} + n^{\mathcal{O}(1)}$.

Empirical algorithms for MAX- SMTI has been studied as well. Munera et al. [34] gave an algorithm based on local search. Gent and Prosser [35] formulated MAX-SMTI as a constrained optimization problem. They give an algorithm using constrained programming for both decision and optimization version of the problem.

8.3.3 BALANCED STABLE MARRIAGE

The BALANCED STABLE MARRIAGE problem was introduced in the influential work of Feder [18] on stable matchings. Feder [18] proved that this problem is NP-hard and that it admits a 2-approximation algorithm. Later, it was shown that this problem also admits a $(2 - 1/\ell)$-approximation algorithm where ℓ is the maximum size of a set of acceptable partners [3]. O'Malley [36] phrased the BALANCED STABLE MARRIAGE problem in terms of constraint programming. Recently, McDermid and Irving [12] expressed interest in the design of fast exact exponential time algorithms for BALANCED STABLE MARRIAGE. For EGALITARIAN STABLE ROOMMATES, Feder [18] showed that the problem is NP-complete even if the preferences are complete and have no ties. and gave a 2-approximation algorithm for this case. Recently, Chen et al. [37] showed that EGALITARIAN STABLE ROOMMATE is FPT parameterized by the egalitarian cost.

Gupta et al. [38] consider two parameterizations of BALANCED STABLE MARRIAGE. Specifically, they introduce two "above-guarantee parameterizations" of BALANCED STABLE MARRIAGE. Let us consider the minimum value O_M of the total dissatisfaction of men that can be realized by a stable matching, and the minimum value O_W of the total dissatisfaction of women that can be realized by a stable matching. Formally, $O_M = \sum_{(m,w)\in\mu_M} p_m(w)$, and $O_W = \sum_{(m,w)\in\mu_W} p_w(m)$, where μ_M and μ_W are the man-optimal and woman-optimal stable matchings, respectively. An input integer k would indicate that the objective is to decide whether $\mathsf{Bal} \leq k$. The first parameter they consider is $k - \min\{O_M, O_W\}$, and the second one, is $k - \max\{O_M, O_W\}$. In other words, they ask the following questions (recall that $\mathsf{Bal} = \min_{\mu\in\mathrm{SM}} \mathsf{balance}(\mu)$).

ABOVE- MIN BALANCED STABLE MARRIAGE (ABOVE- MIN BSM)
Input: An instance $(M, W, \mathcal{L}^M, \mathcal{L}^W)$ of BALANCED STABLE MARRIAGE, and a non-negative integer k.
Question: Is $\mathsf{Bal} \leq k$?
Parameter: $t = k - \min\{O_M, O_W\}$

ABOVE- MAX BALANCED STABLE MARRIAGE (ABOVE- MAX BSM)
Input: An instance $(M, W, \mathcal{L}^M, \mathcal{L}^W)$ of BALANCED STABLE MARRIAGE, and a non-negative integer k.
Question: Is $\mathsf{Bal} \leq k$?
Parameter: $t = k - \max\{O_M, O_W\}$

The parameters. Let us consider the choice of these parameters. Note that the best satisfaction the party of men can hope for is O_M, and the best satisfaction the party of women can hope for is O_W.

First, consider the parameter $t = k - \min\{O_M, O_W\}$. Whenever we have a solution such that the amounts of satisfaction of *both* parties are *close enough* to the best they can hope for, this parameter is small. Indeed, the closer the satisfaction of both parties to the best they can hope for (which is exactly the case where both parties would find the solution desirable), the smaller the parameter is, and the smaller the parameter is, the faster a parameterized algorithm is. In other words, if there exists a solution that is desirable by both parties, this parameter is small.

However, in this parameterization above, as the *min* of $\{O_M, O_W\}$ is taken, it is necessary that the satisfaction of *both* parties to be close to optimal in order to have a small parameter. They show that BALANCED STABLE MARRIAGE is FPT with respect to this parameter. Consequently, the next natural parameter to examine is $t = k - \max\{O_M, O_W\}$. In this case, the parameter is smaller even when *at most one* party is closer to the best satisfaction it can achive. So, the demand from a solution in order to have a small parameter is *weaker*. In the vocabulary of Parameterized Complexity, it is said that the parameterization by $t = k - \max\{O_M, O_W\}$ is "above a higher guarantee" than the parameterization by $t = k - \min\{O_M, O_W\}$, since it is *always* the case that $\max\{O_M, O_W\} \geq \min\{O_M, O_W\}$. Unfortunately, they show, the parameterization by $k - \max\{O_M, O_W\}$ results in a problem that is W[1]-hard. Hence, the complexities of the two parameterizations behave very differently. We remark that in Parameterized Complexity, it is *not at all* the rule that when one takes an "above a higher guarantee" parameterization, the problem would suddenly become W[1]-hard, as can be evidenced by the most classical above-guarantee parameterizations in this field, which are of the VERTEX COVER problem. For that problem, three above-guarantee parameterizations were considered in [39–42], each above a higher guarantee than the previous one that was studied, and each led to a problem that is FPT. In that context, unlike this case, it is still not clear whether the bar can be raised higher. Overall, the results accurately draw the line between tractability and intractability with respect to the target value in the context of two very natural, useful parameterizations.

Finally, to be more precise, Gupta et al. [38] prove three main theorems:

- First, it is proved that ABOVE- MIN BSM admits a kernel where *the number of people is linear in* t. For this purpose, the authors introduce notions that might be of independent interest in the context of a "functional" variant of ABOVE- MIN BSM. Their kernelization algorithm consists of several phases, each simplifying a different aspect of ABOVE- MIN BSM, and shedding light on structural properties of the YES-instances of this problem. Note that this result already implies that ABOVE- MIN BSM is FPT.
- Second, it is proved that ABOVE- MIN BSM admits a parameterized algorithm whose running time is *single exponential in the parameter* t. This algorithm first builds upon the kernel described, and then incorporates the method of bounded search trees.

- Third, it is proved that ABOVE- MAX BSM is W[1]-hard. This reduction is quite technical, and its importance lies in the fact that it rules out (under plausible complexity-theoretic assumptions) the existence of a parameterized algorithm for ABOVE- MAX BSM. Thus, they show that although ABOVE- MAX BSM seems quite similar to ABOVE- MIN BSM, in the realm of Parameterized Complexity, these two problems are completely different.

8.4 Conclusion

In this survey, we gave the current status of various stable marriage problems that have been studied in the framework of Parameterized Complexity. This is an emerging area with lots of open problems. There are two ways of defining new problems in this area. A study the problems that have been considered before with respect to other set of parameters. For example, Gupta et al. [43] studied stable marriage problems parameterized by the *treewidth* of the primal graph as well as of the rotation digraph (wherever it makes sense). Other important parameters that can be used to study these problems include *feedback vertex set*, some width parameter associated with the preference profile, the number of people, and different topologies of the input graphs. The next avenue of defining a new problem is to study another variant of hard stable marriage problems and study them using appropriate parameterizations of solution size. Indeed, matching is just one subarea of algorithmic game theory—a lot more is yet to explored on other topics such as auction, manipulation, and computing equilibriums using a multivariate lens.

References

1. Dan Gusfield, D., Irving, R.W.: The Stable Marriage Problem-Structure and Algorithm. MIT Press, Cambridge (1989)
2. Knuth, D. E.: Stable marriage and its relation to other combinatorial problems: an introduction to the mathematical analysis of algorithms. In: CRM Proceedings & Lecture Notes. American Mathematical Society, Providence, R.I. (1997)
3. Manlove, D.F.: Algorithmics of Matching Under Preferences. Series on Theoretical Computer Science, vol. 2. World Scientific, Singapore (2013)
4. David Gale, D., Shapley, L.S.: College admissions and the stability of marriage. Am. Math. Mon. **69**, 9–15 (1962)
5. Myerson, R.B.: Graphs and cooperation games. Math. Op. Res. **2**, 225–229 (1977)
6. Irving, R.: Stable marriage and indifference. Discret. Appl. Math. **48**, 261–272 (1994)
7. Manlove, David F., D.F.: The structure of stable marriage with indifference. Discret. Appl. Math. **122**, 167–181 (2002)
8. Manlove, D.F., Irving, R.W., Iwama, K., Miyazaki, S., Morita, Y.: Hard variants of stable marriage. Theor. Comput. Sci. **276**, 261–279 (2002)
9. Irving, R.W., Leather, P., Gusfield, D.: An efficient algorithm for the "optimal" stable marriage. J. ACM **34**, 532–543 (1987)

10. Kato, A.: Complexity of the sex-equal stable marriage problem. Jpn. J. Ind. Appl. Math. **10**, 1 (1993)
11. McDermid, E.: In Personal communications between Eric McDermid and David F. Manlove (2010)
12. McDermid, E., Irving, R.: Sex-equal stable matchings: complexity and exact algorithms. Algorithmica **68**, 545–570 (2014)
13. Cseh, A., Manlove, D.F.: Stable marriage and roommates problems with restricted edges: complexity and approximability. Discret. Optim. **20**, 62–89 (2016)
14. Mnich, M., Schlotter, I.: Stable marriage with covering constraints: a complete computational trichotomy (2016). CoRR, arXiv:1602.08230
15. Kobayashi, H., Matsui, T.: Cheating strategies for the gale-shapley algorithm with complete preference lists. Algorithmica **58**, 151–169 (2010)
16. Cygan, M., Fomin, F.V., Kowalik, L., Lokshtanov, D., Marx, D., Pilipczuk, M., Pilipczuk, M., Saurabh, S.: Parameterized Algorithms. Springer, Berlin (2015)
17. Downey R.G., Fellows, M.R.: Fundamentals of Parameterized Complexity. Springer, Berlin (2013)
18. Feder, T.: Stable networks and product graphs. Ph.D. thesis, Stanford University (1990)
19. Bredereck, R., Chen, J., Faliszewski, P., Guo, J., Niedermeier, R., Woeginger, G.J.: Parameterized algorithmics for computational social choice: nine research challenges (2014). CoRR, arXiv:1407.2143
20. Marx, D., Schlotter, I.: Parameterized complexity and local search approaches for the stable marriage problem with ties. Algorithmica **58**, 170–187 (2010)
21. Marx, D., Schlotter, I.: Stable assignment with couples: parameterized complexity and local search. Discret. Optim. **8**, 25–40 (2011)
22. Gupta S., Roy, S.: Stable matching games: manipulation via subgraph isomorphism. In: Proceedings of the 36th IARCS Annual Conference on Foundations of Software Technology and Theoretical Computer Science (FSTTCS) volume 65 of LIPIcs, pp. 29:1–29:14 (2016)
23. Gupta, S., Roy, S.: Stable matching games: manipulation via subgraph isomorphism. Algorithmica **10**, 1–23 (2017)
24. Beyer, T., Hedetniemi, S.M.: Constant time generation of rooted trees. SIAM J. Comput. **9**, 706–712 (1980)
25. Otter, Richard: The number of trees. Ann. Math. **49**, 583–599 (1948)
26. Fomin, F. V., Lokshtanov, D., Panolan, F., Saurabh, S.: Representative sets of product families. J. ACM Trans. Algorithms, **13** (2017)
27. Fomin, F.V., Lokshtanov, D., Raman, V., Saurabh, S., Rao, B.V.R.: Faster algorithms for finding and counting subgraphs. J. Comput. Syst. Sci. **78**, 698–706 (2012)
28. Impagliazzo, R., Paturi, R.: The Complexity of k-SAT. In: The Proceedings of 14th IEEE Conference on Computational Complexity, pp. 237–240 (1999)
29. Adil, D., Gupta, S., Roy, S., Saurabh, S., Zehavi, M.: Parameterized algorithms for stable matching with ties and incomplete lists. Manuscript (2017)
30. Ronn, E.: NP-complete stable matching problem. J. Algorithms **11**, 285–304 (1990)
31. Horton, J.D., Kilakos, K.: Minimum edge dominating sets. SIAM J. Discret. Math. **6**, 375–387 (1993)
32. Irving, R.W., Manlove, D.F., O'Malley, G.: Stable marriage with ties and bounded length preference lists. J. Discret. Algorithms **7**, 213–219 (2009)
33. Peters, D.: Graphical hedonic games of bounded treewidth. In: Proceedings of the 30th AAAI Conference on Artificial Intelligence, pp. 586–593 (2016)
34. Munera, D., Diaz, D., Abreu, S., Rossi, F., Saraswat, V., Codognet, P.: Solving hard stable matching problems via local search and cooperative parallelization. In: Proceedings of 29th AAAI Conference on Artificial Intelligence, pp. 1212–1218 (2015)
35. Gent I. P., Prosser, P: An empirical study of the stable marriage problem with ties and incomplete lists. In: Proceedings of the 15th European Conference on Artificial Intelligence, pp. 141–145. IOS Press (2002)

36. O'Malley, G.: Algorithmic aspects of stable matching problems. Ph.D. thesis, University of Glasgow (2007)
37. Chen, J., Hermelin, D., Sorge, M., Yedidsion, H.: How hard is it to satisfy (almost) all roommates? (2017). CoRR, arXiv:1707.04316
38. Gupta, S., Roy, S., Saurabh, S., Zehavi, M., Balanced stable marriage: how close is close enough? (2017). CoRR, arXiv:1707.09545v1
39. Cygan, M., Pilipczuk, M., Pilipczuk, M., Wojtaszczyk, J.O.: On multiway cut parameterized above lower bounds. TOCT **5**, 3:1–3:11 (2013)
40. Garg, S., Philip, G.: Raising the bar for vertex cover: fixed-parameter tractability above A higher guarantee. In: Proceedings of the 27th Annual ACM-SIAM Symposium on Discrete Algorithms (SODA), pp. 1152–1166 (2016)
41. Lokshtanov, D., Narayanaswamy, N.S., Raman, V., Ramanujan, M.S., Saurabh, S.: Faster parameterized algorithms using linear programming. ACM Trans. Algorithms **11**, 15:1–15:31 (2014)
42. Raman, V., Ramanujan, M.S., Saurabh, S.: Paths, flowers and vertex cover. In: Proceedings of 19th Annual European Symposium of Algorothms (ESA), pp. 382–393 (2011)
43. Gupta, S., Saurabh, S., Zehavi, M.: On treewidth and stable marriage (2017). CoRR, arXiv:1707.05404

Chapter 9
Approximate Quasi-linearity for Large Incomes

Mamoru Kaneko

9.1 Introduction

Quasi-linear utility functions are widely used in economics and game theory. This assumption greatly simplifies the development of theories; for example, in the theory of cooperative games with side payments, Pareto optimality for a given coalition of agents can be expressed by a one-dimensional value of the maximum total surplus, while in the theory without the assumption, Pareto optimality should be described by a set of feasible utility vectors for the coalition. In the cost–benefit analysis, similarly, the total surplus (minus the total cost) from a policy is used as the criterion to recommend it or not.

Quasi-linearity ignores income effects on individual evaluations of alternative choices. It is captured by a condition of no-income effects on such evaluations; a simple axiomatization of quasi-linearity is found in Aumann [1] and Kaneko [6] (see also Kaneko-Wooders [9], Mas-Collel et al. [15], Section 3.C). However, income effects are observed when expenditures for the economic activities in question are non-negligible relative to total incomes; typical examples are individual behaviors in the purchase of a house, automobile, and so forth. Hence, it is desirable to study quasi-linearity from the domain that allows for income effects. As far as the present author knows, only Miyake [11] studied quasi-linearity explicitly from this point of view. In this chapter, we study how much the case of income effects and the case of no-income effects are reconciled; indeed, we give an axiomatic approach to this problem and study its implications.

Miyake [11] studies the above problem in the classical economics context with two commodities. Under the normality condition on income effects and quasi-concavity,

The author thanks two referees for many helpful comments. He is supported by Grant-in-Aids for Scientific Research No. 26245026, and No.17H02258, Ministry of Education, Science and Culture.

M. Kaneko (✉)
Faculty of Political Science and Economics, Waseda University, Tokyo 169-8050, Japan
e-mail: mkanekoepi@waseda.jp

S. K. Neogy et al. (eds.), *Mathematical Programming and Game Theory*,
Indian Statistical Institute Series, https://doi.org/10.1007/978-981-13-3059-9_9

he gave various conditions to guarantee the result that the utility function U is approximated by a quasi-linear function for large incomes.[1] In Miyake [12], he studied the behavior of the demand function for large incomes under similar conditions, but we will discuss his studies in Sect. 9.3.[2]

Our treatment is more direct to approximate quasi-linearity than in [11]. We start with the characterization of quasi-linearity. Let $\succsim$ be a given preference relation over $X \times R_+$, where X is an arbitrary set of the alternatives in question and R_+ is the set of nonnegative real numbers, interpreted as a consumption level measured by a composite commodity (Marshall's money, see Hicks [4], Chap.III, and [5], Chap.5). In addition to certain basic conditions on $\succsim$, when we add a condition – $C4^{PI}$ (*parallel indifference curves*) in Sect. 9.2, we have a quasi-linear utility function $u^* : X \to R$ so that for all $(x, c), (x', c') \in X \times R_+$,

$$(x, c) \succsim (x', c') \Longleftrightarrow u^*(x) + c \geq u^*(x') + c'. \tag{9.1}$$

Our main theorem (Theorem 9.3.1) replaces condition $C4^{PI}$ with a weaker condition, C4 – a Cauchy property, given in Sect. 9.3, and states that a utility function $U^* : X \times R_+ \to R$ representing $\succsim$ is approximated by a quasi-linear utility function $u^* : X \to R$ in the sense that for any $x \in X$ and any $\varepsilon > 0$, there is a c_0 such that

$$\left|U^*(x, c) - (u^*(x) + c)\right| < \varepsilon \quad \text{for all } c \geq c_0. \tag{9.2}$$

Both functions U^* and u^* are derived from $\succsim$ with the basic conditions; in the following, the asterisk * is used to signify that it is derived. Condition (9.2) itself was first mentioned in Miyake [11]. We will show that under other basic conditions, our C4 is equivalent to (9.2).

Condition (9.2) means that $u^*(x) + c$ approximates $U^*(x, c)$ for a large c. The essential part of (9.2) is that $u^*(x) < \infty$ is independent of c. This is justified by assuming that x is tradable in society, but we exclude some familiar mathematical functions from the candidates of approximate quasi-linearity. We will discuss these implications in the end of Sect. 9.3.1.

To study quasi-linearity and the implications mentioned above more clearly, we give another set of sufficient conditions for (9.2) in terms of normality, which is a weakening of Condition $C4^{PI}$. This will be given in Sect. 9.4.

Since economic theory and/or game theory with quasi-linear utility functions are well investigated, it is convenient to connect these cases to the large finite cases. Specifically, we ask the question of how we can convert results obtained in the case with quasi-linearity to the case with large finite incomes. We apply our theorem to

[1] He used the term "asymptotic quasi-linearity." We use "approximate quasi-linearity" to emphasize approximation of a utility function including income effects by a quasi-linear utility function.

[2] It is indirectly related but relevant to mention Vives [22]; he showed that in economies that possibly have many commodities, the income effects on demand of each commodity become negligible relative to the number of commodities as the number increases to infinity.

the theory of cooperative games with side payments in Sect. 9.3.2, and to the theory of Lindahl-ratio equilibrium in a public goods economy in Sect. 9.4.2.

Diagram 9.1 gives a schematic explanation of these applications. We start with a base model E_B and its quasi-linear approximation E_Q, which is the double arrow $\rightleftharpoons$. Some results are obtained in E_Q, and then they are converted to E_B and hold approximately in E_B. The other way to start with E_Q and to find an approximating E_B will be briefly discussed in the end of Sect. 9.3.1.

$$E_B \underset{\text{approximation}}{\rightleftharpoons} E_Q \underset{\text{analysis}}{\Longrightarrow} \text{results} \underset{\text{conversion}}{\rightsquigarrow} \text{approximate result in } E_B,$$

Diagram 9.1

This chapter is written as follows: Sect. 9.2 reviews the characterization of a quasi-linear utility function in terms of a preference relation by Kaneko [6]. In Sect. 9.3, we give a characterization for a preference relation to be approximately represented by a quasi-linear utility function, and we consider its application to the theory of cooperative games side payments. Section 9.4 gives another axiomatization in terms of normality, and an application to the theory of Lindahl-ratio equilibrium. Section 9.5 extends the result in Sect. 9.3 to expected utility theory. Section 9.6 gives a summary of the chapter and states two remaining issues.

9.2 Quasi-linear Utility Function

A *preference relation* $\succsim$ is a binary relation over $X \times R_+$. An expression $(x, c) \succsim (x', c')$ means that (x, c) is weakly preferred to (x', c'). First, we assume Condition C0:
C0 (Complete preordering): $\succsim$ is complete and transitive over $X \times R_+$.

Under C0, we define the *strict part* $\succ$ and the *indifference part* $\sim$ as follows: $(x, c) \succ (x', c') \Longleftrightarrow$ not $(x', c') \succsim (x, c)$; and $(x, c) \sim (x', c') \Longleftrightarrow (x, c) \succsim (x', c')$ and $(x', c') \succsim (x, c)$.

We assume the following three basic conditions.
C1 (Monotonicity): For any $x \in X$, if $c > c'$, then $(x, c) \succ (x, c')$.
C2 (Monetary substitutability): If $(x, c) \succ (x', c')$, then there is an $\alpha > 0$ such that $(x, c) \sim (x', c' + \alpha)$.
C3 (Fixed reference): There is an $x_o \in X$ such that $(x, 0) \succsim (x_o, 0)$ for all $x \in X$. Condition C1 is coherent with the interpretation of R_+ in terms of the composite commodity. Condition C2 means that the economic activities behind the composite commodity R_+ are rich enough to substitute for a transition from x' to x. Condition C3 means that x_o is the worst alternative in X with zero consumption. This is guaranteed by C0 when X is a finite set.[3]

Quasi-linearity can be captured by adding Condition $C4^{PI}$:
$C4^{PI}$ (Parallel indifferences): If $(x, c) \sim (x', c')$ and $\alpha \geq 0$, then $(x, c + \alpha) \sim (x', c' + \alpha)$.

This was given in the case of the domain $X \times R$, instead of $X \times R_+$, in Kaneko [6] (cf., also Kaneko-Wooders [9]), where ξ of $(x, \xi) \in X \times R$ means the increment or decrement from

[3] When X is an infinite set with some topology, under C0, a sufficient condition for C3 is: for any $y \in Y$, $\{(x, 0) \in X \times R : (y, 0) \succsim (x, 0)\}$ is a compact set in $X \times R$. This is proved by using the finite intersection property.

the normalized initial consumption level zero. In the domain $X \times R_+$, $c \in R_+$ is an absolute consumption level, and we can impose an explicit income constraint. As noted in Sect. 9.1, this can be regarded as the no-income-effect condition on evaluations of alternatives $x \in X$. If utility maximization is included here, a change in income may still affect the concluded behavior.[4]

Proposition 9.2.1 (Quasi-linearity) *A preference relation $\succsim$ on $X \times R_+$ satisfies Conditions C0 to C3 and C4^{PI} if and only if there is a function $u^* : X \to R$ such that $u^*(x) \geq u^*(x_o)$ for all $x \in X$ and (9.1) holds for all $(x, c), (x', c') \in X \times R_+$.*

Proof The *only-if* part is essential. Then, let $(x, c) \in X \times R_+$. Since $(x, 0) \succsim (x_o, 0)$ by C3, we have a unique $\alpha_x \geq 0$ by C1 and C2 so that $(x, 0) \sim (x_o, \alpha_x)$. Then, by C4^{PI}, $(x, c) \sim (x_o, \alpha_x + c)$. Define $u^* : X \to R$ by $u^*(x) = \alpha_x$ for all $x \in X$. Now, let $(x, c), (x', c') \in X \times R_+$. Then, by the above definition of α_x, and also by C0 and C1, it holds that $(x, c) \succsim (x', c') \iff (x_o, \alpha_x + c) \sim (x, c) \succsim (x', c') \sim (x_o, \alpha_{x'} + c') \iff \alpha_x + c \geq \alpha_{x'} + c' \iff u^*(x) + c \geq u^*(x') + c'$. □

9.3 Approximate Quasi-linearity

We give a condition for a preference relation $\succsim$ to be approximated by a quasi-linear utility function as an idealization. This approximate representation theorem is given as Theorem 9.3.1. We also give an application to the theory of cooperative games with side payments.

9.3.1 Condition for Approximate Quasi-linearity

Consider the problem of when condition C4^{PI} holds approximately for large incomes. This is answered by relaxing C4^{PI} in the following way:

C4 (Approximate monetary substitutes): Let $x, x' \in X$. For any $\varepsilon > 0$, there is a $c_0 \geq 0$ such that for any $c, c' \geq c_0$ and $\alpha, \alpha' \geq 0$, if $(x, c) \sim (x', c + \alpha)$ and $(x, c') \sim (x', c' + \alpha')$, then $\left|\alpha - \alpha'\right| < \varepsilon$.

The additional α, α' are compensations for the transitions from x to x' with consumptions c, c', and C4 requires these to be close for large c and c'. This is a kind of Cauchy property of a sequence $\{a_v\}$ (cf., Royden-Fitzpatrick [18], Section 1.5). Condition C4 is an weakening of C4^{PI} under C1 (i.e., C4^{PI} implies that the conclusion of C4 becomes $\left|\alpha - \alpha'\right| = 0$).

The following lemma is basic for the development of our theory.

Lemma 9.3.1 (Measurement along the consumption axis with x_o) *Suppose that $\succsim$ satisfies C0 to C3. Then there is a real-valued function $\delta^* : X \times R_+ \to R$ such that for any $(x, c), (x', c') \in X \times R_+$,*

[4] This is pointed out by a referee. Let a utility function representing $\succsim$ be given as $u(x) + c$ over R_+^2. When $u(x)$ is a strictly concave function, utility maximization gives a choice of x, independent of an income if it is large. However, if $u(x) = x^2$, then utility maximization gives a corner solution; a change in income affects this corner solution.

$$(x, c) \sim (x_o, \delta^*(x, c) + c); \tag{9.3}$$

$$(x, c) \succsim (x', c') \Longleftrightarrow \delta^*(x, c) + c \geq \delta^*(x', c') + c'. \tag{9.4}$$

It holds that

$$\delta^*(x_o, c) = 0 \text{ for all } c \in R_+. \tag{9.5}$$

Proof Consider any $(x, c) \in X \times R_+$. Then, $(x, c) \succsim (x_o, 0)$ by C0, C1, and C3. Thus, there is a unique value $\delta^*(x, c) + c$ by C1 and C2 such that $(x, c) \sim (x_o, \delta^*(x, c) + c)$. Thus, we have the function $\delta^*(\cdot, \cdot) : X \times R_+ \to R$ satisfying (9.3). We show (9.4). Let $(x, c), (x', c') \in X \times R_+$. Then, by C0 and C1, $(x_o, \delta^*(x, c) + c) \sim (x, c) \succsim (x', c') \sim (x_o, \delta^*(x', c') + c')$ $\Longleftrightarrow \delta^*(x, c) + c \geq \delta^*(x', c') + c'$. Letting $(x, c) = (x_o, c)$, by (9.3), we have $\delta^*(x_o, c) = 0$, i.e., (9.5). □

Define $U^* : X \times R_+ \to R$ by

$$U^*(x, c) = \delta^*(x, c) + c \text{ for all } (x, c) \in X \times R_+. \tag{9.6}$$

Equation (9.4) states that this U^* represents the preference relation $\succsim$. These particular functions, $\delta^*(x, c)$ and $U^*(x, c)$, play crucial roles in the following development. The third statement (9.5) means that $U^*(x, c) = \delta^*(x, c) + c$ is measured by the scale of the consumption axis at x_o.

Approximate quasi-linearity (9.2) is then written as: there is some real-valued function u^* over X such that

$$\left|U^*(x, c) - (u^*(x) + c)\right| \to 0 \text{ as } c \to +\infty. \tag{9.7}$$

Our main theorem states that C4 is exactly the condition for the existence of such a function $u^*(x)$.

Theorem 9.3.1 (Approximate quasi-linearity) *Let $\succsim$ be a preference relation on $X \times R_+$ satisfying C0 to C3, and δ^* the function given by Lemma 9.3.1. Then, $\succsim$ satisfies C4 if and only if for each $x \in X$, there is a $u^*(x) \in R$ such that*

$$\lim_{c \to +\infty} \delta^*(x, c) = u^*(x). \tag{9.8}$$

Proof *If*:[5] Let $x, x' \in X$ and $\varepsilon > 0$. By (9.8), there is a $c_o \geq 0$ such that for all $d \geq c_o$,

$$\left|\delta^*(x, d) - u^*(x)\right| < \tfrac{\varepsilon}{4} \text{ and } \left|\delta^*(x', d) - u^*(x')\right| < \tfrac{\varepsilon}{4}. \tag{9.9}$$

Now, let $c, c' \geq c_o$ and $\alpha, \alpha' \geq 0$. Suppose that $(x, c) \sim (x', c + \alpha)$ and $(x, c') \sim (x', c' + \alpha')$. Then, applying (9.3) of Lemma 9.3.1, we have

$$\delta^*(x, c) = \delta^*(x', c + \alpha) + \alpha \text{ and } \delta^*(x, c') = \delta^*(x', c' + \alpha') + \alpha'. \tag{9.10}$$

[5] This proof is given by a referee and is clearer than the original proof by the author.

Letting $d = c$ in (9.9), we obtain

$$\left|\delta^*(x,c) - u^*(x)\right| < \tfrac{\varepsilon}{4}. \tag{9.11}$$

Letting $d = c + \alpha$ in the second inequality of (9.9), we have $\left|\delta^*(x', c+\alpha) - u^*(x')\right| < \frac{\varepsilon}{4}$. Since $\delta^*(x', c+\alpha) = \delta^*(x,c) - \alpha$ by (9.10), we have

$$\begin{aligned}\left|u^*(x') + \alpha - \delta^*(x,c)\right| &= \left|(\delta^*(x,c) - \alpha) - u^*(x')\right| \\ &= \left|\delta^*(x', c+\alpha) - u^*(x')\right| < \tfrac{\varepsilon}{4}.\end{aligned} \tag{9.12}$$

In a parallel manner to the derivations of (9.11) and (9.12), we have

$$\left|\delta^*(x', c'+\alpha') - u^*(x')\right| < \tfrac{\varepsilon}{4} \text{ and } \left|u^*(x) + \alpha' - \delta^*(x', c'+\alpha')\right| < \tfrac{\varepsilon}{4}. \tag{9.13}$$

Using the triangle inequality and summing up (9.11)–(9.13), we have

$$\left|\alpha - \alpha'\right| \le \left|\delta^*(x,c) - u^*(x)\right| + \left|u^*(x') + \alpha - \delta^*(x,c)\right|$$

$$+ \left|\delta^*(x', c'+\alpha') - u^*(x')\right| + \left|u^*(x) + \alpha' - \delta^*(x', c'+\alpha')\right| < 4 \times \frac{\varepsilon}{4} = \varepsilon.$$

Only-if: Let $\delta^*(x,c)$ be the function given by (9.3). We show that for each fixed $x \in X$, there is a $u^*(x) \in R$ satisfying (9.8). Consider the sequence $\{\delta^*(x,\nu)\} = \{\delta^*(x,\nu) : \nu = 1, ...\}$. C4 states that for any $\varepsilon > 0$, there is a ν_0 such that for any $\nu, \nu' \ge \nu_0$, $\left|\delta^*(x,\nu) - \delta^*(x,\nu')\right| < \varepsilon$. This means that $\{\delta^*(x,\nu)\}$ is a Cauchy sequence. Hence, it converges to some real number, which is denoted by $u^*(x)$. Now, each $\delta^*(x,\nu)$ in $\{\delta^*(x,\nu)\}$ is defined for a natural number $\nu \ge 1$. However, we prove $\lim_{c\to+\infty} \delta^*(x,c) = u^*(x)$. Let ε be an arbitrary positive number. Then there is a ν_0 such that for any $\nu \ge \nu_0$, $\left|\delta^*(x,\nu) - u^*(x)\right| < \varepsilon/2$. By C4, there is a c_0 such that for any $\nu \ge c_0$ and $c \ge c_0$, $\left|\delta^*(x,\nu) - \delta^*(x,c)\right| < \varepsilon/2$. Now, let $\nu_1 = \max(\nu_0, c_0)$. Then, for any $c \ge \nu_1$, we have $\left|\delta^*(x,c) - u^*(x)\right| \le \left|\delta^*(x,c) - \delta^*(x,\nu_1)\right| + \left|\delta^*(x,\nu_1) - u^*(x)\right| < \varepsilon$. Thus, $\lim_{c\to+\infty} \delta^*(x,c) = u^*(x)$. This is (9.8). □

Theorem 9.3.1 states that the central condition for approximate quasi-linearity should be C4 or equivalently (9.8). Now, we use either condition to think about various implications.

As stated above, the functions $\delta^*(x,c)$ and $U^*(x,c)$ are defined particularly by (9.3) and (9.6). Let $U(x,c)$ be any utility function representing a given preference relation $\succsim$, and we define $\delta(x,c) = U(x,c) - c$ for $(x,c) \in X \times R_+$. These may look like candidates for $\delta^*(x,c)$ and $U^*(x,c)$. In general, however, they may not satisfy (9.3). When we talk about examples of utility functions $U(x,c)$, we should not forget that $\delta^*(x,c)$ is defined by (9.3), rather than $U(x,c) - c$.

To see this, consider the following necessary condition of (9.8): for each $x \in X$,

$$\{\delta^*(x,c) : c \in R_+\} \text{ is bounded.} \tag{9.14}$$

Thus, the compensation for x from x_o is bounded even if c is very large.

Consider $U_0(x,c) = u(x) + \sqrt{c}$ for $(x,c) \in R_+ \times R_+$ with $u(x_o) < u(x)$ for all $x \neq x_o = 0$, which satisfies the law of diminishing marginal utility for c and differentiability at any $c > 0$. The function $\delta^*(x,c)$ derived from this $U_0(x,c)$, however, violates (9.14).

Choose an x with $u(x) > u(x_o)$ and let $h = u(x) - u(x_o)$. Then, $u(x) + \sqrt{c} = u(x_o) + \sqrt{c + \delta^*(x,c)}$, i.e., $h = \sqrt{c + \delta^*(x,c)} - \sqrt{c}$; so $\delta^*(x,c) = (h + \sqrt{c})^2 - c = h^2 + 2h\sqrt{c}$. Hence, $\delta^*(x,c) \to +\infty$ as $c \to +\infty$; (9.14) is violated.

A positive example is: $U_1(x,c) = (1 - \frac{1}{1+c})u(x) + c$. In this case, it is possible to directly verify (9.14). To facilitate such applications, we provide further conditions on approximate quasi-linearity on $\succsim$. One is related to boundedness (9.14), which is studied now, and the other is a normality condition, which is studied in Sect. 9.4. We represent boundedness in terms of the preference relation $\succsim$.

C5 (Boundedness for compensations): For any $x \in X$, there is an $m > 0$ such that $(x_o, c + m) \succsim (x, c)$ for any $c \in R_+$.

That is, there is a compensation m for x_o from x independent of consumption level c. Under C0 to C3, this is equivalent to boundedness of $\delta^*(x, \cdot)$ for each $x \in X$. The example $U_1(x,c) = (1 - \frac{1}{1+c})u(x) + c$ satisfies this condition, but $U_0(x,c) = u(x) + \sqrt{c}$ does not.

Lemma 9.3.2 *Suppose that $\succsim$ satisfies C0 to C3. Then, $\succsim$ satisfies C5 if and only if (9.14) holds for $\delta^*(x, \cdot)$ for each $x \in X$.*

Proof If: By (9.14), there is an $m \in R_+$ such that $m > \delta^*(x,c)$ for all $c \in R_+$. By (9.3), we have $(x_o, c + \delta^*(x,c)) \sim (x,c)$ for any c. By C1, we have $(x_o, c+m) \succsim (x,c)$ for any c.

Only-if: By (9.3), $(x_o, c + \delta^*(x,c)) \sim (x,c)$ for any $c \in R_+$. By C5, $(x_o, c+m) \succsim (x_o, c + \delta^*(x,c)) \sim (x,c)$ for any $c \in R_+$. By C1, we have $m \geq \delta^*(x,c)$ for any $c \in R_+$. □

As mentioned in Sect. 9.1, approximate quasi-linearity was first studied in Miyake [11], who aimed to study the Marshallian demand theory; he starts with the domain $X \times R_+ = R_+ \times R_+$ and assumes that a utility function U of C^2 (twice continuously differentiable in the interior of $R_+ \times R_+$) is given. U is assumed to be quasi-concave and satisfies normality (formulated in terms of first and second partial derivatives) in $R_+ \times R_+$. He then gives some other conditions to guarantee approximate quasi-linearity in the sense of (9.2), and various results on the limit demand function.

Miyake [12] continued his study of the Marshallian demand theory, where he gave a criterion for a demand function, called "asymptotically well-behaved demand;" this appears to be related to our approximate quasi-linearity. He gave three examples:

(a) $U_a(x,c) = \log(x+1) + \log(c+1) + c$;

(b) $U_b(x,c) = \frac{(x+1)(c+1)}{x+c+2} + c \ (= (x+1) - \frac{(x+1)^2}{x+c+2} + c)$;

(c) $U_c(x,c) = 2\sqrt{x} + \sqrt{c} + c$.

He showed that (a) and (c) satisfy his criterion but (b) does not. These three, however, satisfy condition C5. For example, consider (a); since $\log(x+1) + \log(c+1) + c > \log(c+1) + c$, it suffices to take an $m > \log(x+1)$. Indeed,

$$\begin{aligned} U_a(x,c) &= \log(x+1) + \log(c+1) + c < m + \log(c+1) + c \\ &< \log(c+1+m) + (c+m) = U_a(0, c+m). \end{aligned}$$

The verification of (c) is similar. For (b), since $(x+1)+c > (x+1) - \frac{(x+1)^2}{x+c+2} + c > 1 - \frac{1}{c+2} + c \geq \frac{1}{2} + c$, it suffices to take an $m > x - \frac{1}{2}$. Moreover, we will see, using another characterization of approximate quasi-linearity in Sect. 9.4.1, that these three examples satisfy approximate quasi-linearity. Thus, Miyake's "asymptotically well-behaved demand" conceptually differs from approximate quasi-linearity.[6]

The exclusion of utility function such as $U_0(x, c) = u(x) + \sqrt{c}$ from C5 may give rise to some inconvenience. In fact, we can avoid it by changing utility functions slightly. For example, the above utility function is changed into

$$V_{c_o}(x, c) = \begin{cases} u(x)+\sqrt{c} & \text{if } c \leq c_o \\ u(x) + \beta(c - c_o)+\sqrt{c_o} & if\, c > c_o, \end{cases} \tag{9.15}$$

where $c_o >$ is a given parameter and $\beta = \frac{1}{2\sqrt{c_o}}$. That is, $V_{c_o}(x, c)$ is obtained from $U_0(x, c)$ by linearizing $\sqrt{c}$ after c_o. This $V_{c_o}(x, c)$ satisfies C5 and the normality condition C5^{NM} to be given in Sect. 9.4, from which we will see that the preference relation derived by $V_{c_o}(x, c)$ satisfies C0-C4.[7]

This example is related to a typical justification of quasi-linearity: when a utility function $U(x, c)$ is partially differentiable with respect to c, it is regarded as locally approximated by a linear function of c. When the expenditures for possible choices are small relative to incomes, we could have a quasi-linear approximation of U. However, our theory reveals that this interpretation is incorrect since our theory requires all consumption levels after some c_o. It is an open question of whether our definition can be modified to capture this interpretation.

A schematic representation of the above argument was given as Diagram 9.1. Here, we start with a given preference relation $\succsim$ and give the conditions, C0–C4, for $\succsim$ to be approximately represented by a quasi-linear utility function $u^*(x) + c$. Another approach is to ask whether for a given $u^*(x) + c$, we find a preference relation $\succsim$ to be represented approximately by $u^*(x) + c$ in a nontrivial sense (i.e., $\succsim$ differs from the relation represented by $u^*(x) + c$). This direction is depicted in Diagram 9.2. Here, we give only a simple answer to this question. A full study remains open.

$$E_Q \underset{\text{approximation}}{\Longrightarrow} E_B$$

Diagram 9.2

Suppose that $u : X \to R$ is given with $0 = u(x_o) \leq u(x)$ for any $x \in X$, and we specifically define $\delta : X \times R_+ \to R$ to be

$$\delta(x, c) = u(x) - \frac{u(x)}{(c+1)^\alpha}, \tag{9.16}$$

[6]Miyake [13] provided a result (Theorem 2 in p.561) related to this approach. He studied the behavior of "willingness-to-pay" and willingness-to-accept," and he provided many results on the behavior of these concepts.

[7]Kaneko-Ito [10] conducted an equilibrium-econometric analysis to study how utility functions have "significant income effects," adopting utility functions of the form $U(x, c) = u(x) + c^\alpha$ $(0 < \alpha < 1)$. It was shown that this α is bounded away from 1 using rental housing market data in Tokyo. Since incomes of households are distributed over some interval, we do not need the above modification of a utility function.

where $\alpha > 0$ is a parameter. The utility function $U_2(x, c) = \delta(x, c) + c$ for all $(x, c) \in X \times R_+$ derives the preference relation $\succsim$ satisfying C0-C4, and $\delta^*(x, c)$ derived by (9.3) is $\delta(x, c)$ itself, and $u^*(x)$ derived by (9.8) is $u(x)$, too. Thus, E_B is obtained from E_Q. Incidentally, the parameter α represents the convergence speed of $\delta(x, c) = \delta^*(x, c)$ to $u(x) = u^*(x)$ (i.e., when α is large, the convergence speed is fast, but when α is close to 0, it is slow).

Finally, we raise the question of whether approximate quasi-linearity is an appropriate concept from the viewpoint of economics. Our theory formulates "large income" simply as "c tends to $+\infty$." Mathematically, there are two possibilities: (A) $\delta^*(x, c)$ is in a bounded region, and (B) it goes to $+\infty$. There is a subtlety in the interpretation of "large incomes." To have a meaningful interpretation, we should consider how much richness is hidden behind the compound commodity c and/or the richness of X, which was mentioned to justify condition C2. The two mathematical possibilities are examined from the socioeconomic point of view.

When income gets larger for a person, his/her scope of consumption (economic behavior in general) gets larger. Suppose that there is an alternative y, hidden behind the composite commodity c or in X, similar to x in the sense that the person can switch from x to y. When this is applied to any person in a similar economic situation, a value of each of x or y is more or less determined. In this interpretation, $\delta^*(x, c)$ is not very different from the social/market value. Here, possibility (A) is justified, and approximate quasi-linearity is applied.

In possibility (B), alternative x is unique and has no substitution for the person either behind the composite commodity or in X; x may be indispensable for him/her and its value may be unbounded when $c \to +\infty$. In this case, approximate quasi-linearity does not hold, and even condition C2 is not justified. Nevertheless, this is only a logically possible world.

9.3.2 An Application to Cooperative Game Theory

Here, we consider an application of Theorem 9.3.1 to the theory of cooperative games with side payments (cf., Osborne-Rubinstein [17], Chap.13, Maschler et al. [14], Chap.16). This is one example for Diagram 9.1.

We denote the set of agents by $N = \{1, ..., n\}$. For each nonempty subset $S \subseteq N$, X_S is given as a *finite* nonempty set of social alternatives to be controlled by S, and $C_S : X_S \to R$ is a cost function. It can be assumed that $X_S \cap X_{S'} = \emptyset$ if $S \neq S'$. The value $C_S(x)$ for each $x \in X_S$ is allocated among the members in S. Let $X^i = \cup_{i \in S \subseteq N} X_S$. Each agent $i \in N$ has a preference relation $\succsim_i$ over the set $X^i \times R_+$ and an initial income $I_i \geq 0$. Here, $(x, c_i) \in X_S \times R_+$ means that an alternative x for S is chosen, and agent i's consumption is c_i after paying his/her cost assignment. The *base model* is expressed as $E_B = (\{C_S\}_{S \subseteq N}, \{\succsim_i\}_{i \in N}, \{I_i\}_{i \in N})$. Here, $(\{C_S\}_{S \subseteq N}, \{\succsim_i\}_{i \in N})$ are fixed, but only $\{I_i\}_{i \in N}$ are variable parameters. In this sense, E_B may be written as $E_B(\{I_i\}_{i \in N})$. The above formulation includes market games[8] (cf., Shapley–Shubik [19]), voting games (cf., Kaneko–Wooders [9]).

Under C0–C4 for the preference relations $\succsim_i$ for each $i \in N$, we have two functions $u_i^* : X^i \to R$ and $U_i^* : X^i \times R_+ \to R$ satisfying (9.6) and (9.8). In a parallel manner as above, the *quasi-linear approximation* is given as $E_Q = E_Q(\{I_i\}_{i \in N}) = (\{C_S\}_{S \subseteq N}, \{u_i^*\}_{i \in N}, \{I_i\}_{i \in N})$. In E_Q, we define the characteristic function v by, for all $S \subseteq N$,

[8] When the set of commodity bundles is infinite, we need some modifications.

$$v(S) = \max_{x \in X_S} \left(\sum_{i \in S} u_i^*(x) - C_S(x) \right). \tag{9.17}$$

The value $v(S)$ is the maximum total surplus obtained by S. When $\sum_{i \in S} I_i \geq C_S(x)$ for all $x \in X_S$, this maximization meets the budget constraint. The pair (N, v) is a *game with side payments*.

We ask the question of how (N, v) is related to the base model E_B. The aim of (N, v) is to consider a distribution of the total surplus for each S expressed by v. Such a distribution is described by an imputation: A vector $\alpha_S = \{\alpha_i\}_{i \in S}$ is called an S- *imputation* iff $\sum_{i \in S} \alpha_i = v(S)$ and $\alpha_i \geq v(\{i\})$ for all $i \in S$. We denote the set of all S-imputations in (N, v) by $I_S(N, v)$.[9] Then, the question is what the set $I_S(N, v)$ is in the base model E_B.

Let $\alpha_S = \{\alpha_i\}_{i \in S} \in I_S(N, v)$ and let x_S^* be a solution for (9.17). We consider the corresponding allocation in the base model E_B. The cost assignment for agent $i \in S$ is given as $\gamma_i(\alpha_i) := u_i^*(x_S^*) - \alpha_i$. Indeed, $\alpha_i = u_i^*(x_S^*) - \gamma_i(\alpha_i)$ is the net surplus for agent i. When the budget constraint $I_i \geq \gamma_i(\alpha_i)$ holds for each $i \in S$, we can construct an S-allocation in the base model E_B :

$$\psi(\alpha_S) = (x_S^*, \{I_i - \gamma_i(\alpha_i)\}_{i \in S}). \tag{9.18}$$

In E_B, the utility level for agent i is given as $U_i^*(x_S^*, I_i - \gamma_i(\alpha_i))$, and in $E_Q = E_Q(\{I_i\}_{i \in N})$, the utility level for agent i is given as

$$u_i^*(x_S^*) + (I_i - \gamma_i(\alpha_i)) = I_i + \alpha_i, \tag{9.19}$$

because $\gamma_i(\alpha_i) = u_i^*(x_S^*) - \alpha_i$. That is, the surplus α_i is the increment of utility from the initial I_i. If the initial state is normalized as 0, the utility level is exactly α_i.

The question is now how the cost allocation $\{\gamma_i(\alpha_i)\}_{i \in S}$ is interpreted in E_B. Here, we assume C0 to C4 for the preference relations $\succsim_i$ for each $i \in S$. Recall that the functions $u_i^* : X^i \to R$ and $U_i^* : X^i \times R_+ \to R$ are defined by (9.8) and (9.6).

Theorem 9.3.2 (Approximation by a game with side payments) *For any $\varepsilon > 0$, there is an $I^* \geq 0$ such that for any $I_i \geq I^*$ for all $i \in S$, and for all $\alpha_S = \{\alpha_i\}_{i \in S} \in I_S(N, v)$,*

$$I_i \geq \gamma_i(\alpha_i) \text{ for all } i \in S; \tag{9.20}$$

$$\left| U_i^*(x_S^*, I_i - \gamma_i(\alpha_i)) - (u_i(x_S^*) + (I_i - \gamma_i(\alpha_i)) \right| < \varepsilon \text{ for all } i \in S. \tag{9.21}$$

Proof First, we fix an agent $i \in S$. The set $\{\gamma_i(\alpha_i) : \alpha_S \in I_S(N, v)\}$ is bounded. Let I_i^0 be an income level greater than the maximum of this set. Hence, for all $I_i \geq I_i^0$, we have (9.20) for i.

Consider (9.21) for i. Applying Theorem 9.3.1 to i, we have some c_i^* such that for any $c_i \geq c_i^*$, $\left| U_i^*(x_S^*, c_i) - (u_i^*(x_S^*) + c_i) \right| < \varepsilon$. Since $\gamma_i^*(\alpha_i) = u_i^*(x_S^*) - \alpha_i$ and $\alpha_i \geq v(\{i\})$ for all $\alpha_S \in I_S(N, v)$, we can take an I_i^1 so that $I_i^1 - (u_i^*(x_S^*) - \alpha_i) \geq c_i^*$ for all $\alpha_S \in I_S(N, v)$. Then, we have, for all $I_i \geq I_i^1$,

$$\begin{aligned} &\left| U_i^*(x_S^*, I_i - \gamma_i(\alpha_i)) - (u_i^*(x_S^*) + (I_i - \gamma_i(\alpha_i)) \right| \\ = &\left| U_i^*(x_S^*, I_i - (u_i^*(x_S^*) - \alpha_i)) - (u_i^*(x_S^*) + I_i - (u_i(x_S^*) - \alpha_i)) \right| < \varepsilon \end{aligned}$$

[9] The set $I_S(N, v)$ is nonempty under some additional condition (e.g., $v(S) \geq \sum_{i \in S} v(\{i\})$).

for all $\alpha_S \in I_S(N, v)$. We take $I^* = \max\{I_i^0, I_i^1 : i \in S\}$. Then, for this I^*, (9.20) and (9.21) hold for all $i \in S$. □

In Theorem 9.3.2, we focus on a particular coalition S. The theorem can be extended to the existence of I^* uniformly for all $S \subseteq N$. Once this is obtained, we can apply it to a solution theory for (N, v). For example, the core of (N, v) can be translated into the approximate core in the base model $E_B(\{I_i\}_{i\in N})$. Thus, the theory of cooperative games with side payments is viewed as an ideal approximation of the theory without quasi-linearity.

9.4 Characterization by Normality

Under C0 to C3, condition C4 is equivalent to approximate quasi-linearity. Some sufficient conditions are useful for applications in economics and game theory. Here, we weaken condition $C4^{PI}$ in a different manner from C4; it is normality, which together with C5 (boundedness) implies C4. We will apply this result to the theory of Lindahl-ratio equilibrium in a public good economy, which is another example of conversion suggested in Diagram 9.1.

9.4.1 Normality and Approximate Quasi-linearity

Boundedness C5 is a necessary condition for approximate quasi-linearity. When C5 is assumed in addition to C0–C3, the monotonicity (weakly increasing) of $\delta^*(x, c)$ with c is enough to have (9.8). In fact, this monotonicity is guaranteed by a normality condition. First, we look at a weak form of normality, which is equivalent to the monotonicity of $\delta^*(x, c)$.

$C4^{NM_o}$ (Normality$_o$): Let $(x, c) \in X \times R_+$, $c' \in R_+$, and $\alpha \geq 0$. If $(x, c) \sim (x_o, c')$ and $c \leq c'$, then $(x, c + \alpha) \succsim (x_o, c' + \alpha)$.

An additional α to (x, c) gives more (or equal) satisfaction than to (x_o, c').

Lemma 9.4.1 (Monotonicity) *Suppose C0 to C3 for $\succsim$. Let $x \in X$.*
*(**1**): Suppose $C4^{NM_o}$. Then, $(x, c) \succsim (x_o, c)$ and $\delta^*(x, c) \geq 0$ for all $c \geq 0$.*
*(**2**): $C4^{NM_o}$ holds if and only if $\delta^*(x, c)$ is weakly increasing with respect to c.*

Proof **(1)**: Since $(x, 0) \succsim (x_o, 0)$ by C3, we have $(x, 0) \sim (x_o, \alpha)$ for some $\alpha \geq 0$ by C2. Hence, we have $(x, 0 + c) \succsim (x_o, \alpha + c)$ by $C4^{NM_o}$. By C0 and C1, we have $(x, c) \succsim (x_o, c)$. By (9.3), $(x, c) \sim (x_o, \delta^*(x, c) + c)$. Since $(x, c) \succsim (x_o, c)$, by C0 and C1, we have $\delta^*(x, c) \geq 0$.
(2): *Only-if*: Now, let $\alpha \geq 0$. Then, since $(x, c) \sim (x_o, \delta^*(x, c) + c)$ by (9.3) and $\delta^*(x, c) \geq 0$ by (1), we have, by $C4^{NM_o}$, $(x, c + \alpha) \succsim (x_o, \delta^*(x, c) + c + \alpha)$. Since $(x, c + \alpha) \sim (x_o, \delta^*(x, c + \alpha) + c + \alpha)$, we have $(x_o, \delta^*(x, c + \alpha) + c + \alpha) \succsim (x_o, \delta^*(x, c) + c + \alpha)$ by C0. This and C1 imply $\delta^*(x, c + \alpha) \geq \delta^*(x, c)$.

If: Suppose $(x, c) \sim (x_o, c')$ and $c \leq c'$. By (9.3), $(x_o, \delta^*(x, c) + c) \sim (x, c) \sim (x_o, c')$. By C0 and C1, we have $\delta^*(x, c) + c = c'$. Since $\delta^*(x, c)$ is increasing with c, we have $\delta^*(x, c + \alpha) \geq \delta^*(x, c)$. Since $\delta^*(x, c) + c = c'$, we have $\delta^*(x, c) + c + \alpha = c' + \alpha$. Thus, $\delta^*(x, c +$

$\alpha) + c + \alpha \geq c' + \alpha$. By (9.3) and C1, we have $(x, c + \alpha) \sim (x_o, \delta^*(x, c + \alpha) + c + \alpha) \succsim (x_o, c' + \alpha)$. This is the conclusion of $C4^{NM_o}$. □

Under C0 to C3, $C4^{NM_o}$ and C5, the function $\delta^*(x, c)$ is increasing (Lemma 9.4.1.(2)) and bounded (Lemma 9.4.2) with c for each $x \in X$. Hence, $\delta^*(x, c)$ converges to $u^*(x)$ as $c \to +\infty$. This is (9.8) of Theorem 9.3.1, and thus C4 is derived.

Theorem 9.4.1 (Characterization by normality) *Suppose C0 to C3, $C4^{NM_o}$, and C5 for $\succsim$. Then, (9.8) holds for $\succsim$.*

The examples in Sect. 9.3 satisfy condition $C4^{NM_o}$. For example, $U_b(x, c) = \frac{(x+1)(c+1)}{x+c+2} + c$ $(= (x + 1) - \frac{(x+1)^2}{x+c+2} + c)$ is a concave function of c, which implies $C4^{NM_o}$. The other example $U_2(x, c) = u(x) - \frac{u(x)}{(c+1)^\alpha} + c$ in (9.16) provides that $\delta^*(x, c) = u(x) - \frac{u(x)}{(c+1)^\alpha}$ is increasing with c; the derived preference relation $\succsim$ satisfies $C4^{NM_o}$ by Lemma 9.4.1.(2).

The above form of normality $C4^{NM_o}$ is enough for (9.8) but it requires nothing direct about the relationship between different alternatives x and x'. It may be more convenient to mention the following stronger form:[10]

$C4^{NM}$ (Normality[11]**)**: Let $(x, c), (x', c') \in X \times R_+$ and $\alpha \geq 0$. If $(x, c) \sim (x', c')$ and $c \leq c'$, then $(x, c + \alpha) \succsim (x', c' + \alpha)$.

We have the full monotonicities for $\succsim$ and $\delta^*(x, c)$ over $x \in X$ and $c \in R_+$.

Lemma 9.4.2 (Monotonicities over X *and* R_+) Suppose C0 to C3 and $C4^{NM}$ for $\succsim$. Let $x, x' \in X$. If $(x, 0) \succsim (x', 0)$, then $(x, c) \succsim (x', c)$ and $\delta^*(x, c) \geq \delta^*(x', c)$ for all $c \geq 0$.

Proof Let $(x, 0) \succsim (x', 0)$. The first conclusion is obtained from the proof of Lemma 9.4.1.(1) by replacing x_o by x'. Hence, by (9.3), $(x_o, \delta^*(x, c) + c) \sim (x, c) \succsim (x', c) \sim (x_o, \delta^*(x', c) + c)$. By C0 and C1, we have $\delta^*(x, c) \geq \delta^*(x', c)$. □

For applications in Sect. 9.4.2, we provide certain specific properties on the derived function $u^* : X \to R$. Suppose that $X = Z_o = \{0, ..., z_o\}$ and the worst x_o in C3 is fixed to be 0.[12] The set $X \times R_+ = Z_o \times R_+$ is not convex in the standard sense. However, we can modify the definition of convexity slightly, which enables us to discuss convexity almost in the same way as the standard.

We say that a subset S of $Z_o \times R_+$ is *convex* iff for any $(x, c), (x', c') \in Z_o \times R_+$ and any $\lambda \in [0, 1]$ with $\lambda x + (1 - \lambda)x' \in Z_o$, it holds that $\lambda x + (1 - \lambda)x' \in S$. Using this notion, we have the following definition of convexity of $\succsim$: the preference relation $\succsim$ is said to be *convex* iff $\{(x', c') \in Z_o \times R_+ : (x', c') \succsim (x, c)\}$ is a convex set for any $(x, c) \in Z_o \times R_+$. Similarly, we say that a function $f : Z_o \to R$ is *concave (convex)* iff for any $x, x' \in Z_o$ and $\lambda \in (0, 1)$

[10] This strict version is used in Kaneko [8].

[11] This term "normality" is motivated by the following observation. Suppose that $\succsim$ is weakly increasing with respect to $x \in X = R_+$. Then, the demand function, assumed to exist here, for the commodity in $X = R_+$ is weakly monotonic with an income. Indeed, let $p > 0$. Let $(x, I - px) \succsim (x', I - px')$ and $x > x'$. By C1 and C2, $(x, I - px) \sim (x', I - px' + \alpha)$ for some $\alpha \geq 0$. Let $I' > I$. Then, since $I - px < I - px' + \alpha$, we have $(x, I' - px) \succsim (x', I' - px' + \alpha) \succsim (x', I' - px')$ by $C4^{NM}$ and C1. This means that the quantity demanded weakly increases when an income increases.

[12] The finiteness of Z_o is assumed to have the uniform convergence result in Theorem 9.4.2. Otherwise, we could take the set of all nonnegative integers Z_+.

with $\lambda x + (1-\lambda)x' \in Z_o$, it holds that $f(\lambda x + (1-\lambda)x') \geq (\leq)\ \lambda f(x) + (1-\lambda) f(x')$. This implies $f(x) - f(x-1) \geq (\leq)\ f(x+1) - f(x)$ for all $x \in Z_o$ with $0 < x < z_o$.

We have the following result for the function u^* derived from $\succsim$ with C0 to C4 in Theorem 9.3.1.

Lemma 9.4.3 (Concavity) *If $\succsim$ is convex, then $u^*(x)$ is a concave function over Z_o.*

Proof Let $x, x' \in Z_o$ and $c \in R_+$. Suppose $(x, c) \succsim (x', c)$. Then, by C1, C2, we have a unique $c' \geq c$ such that $(x, c) \sim (x', c')$. This implies $\delta^*(x', c') + c' = \delta^*(x, c) + c$. We denote $c' = c'(c)$.

Let $\lambda \in (0, 1)$ with $\lambda x + (1-\lambda)x' \in Z_o$. Then, by convexity for $\succsim$, we have $(\lambda x + (1-\lambda)x', \lambda c + (1-\lambda)c') \succsim (x, c) \sim (x', c')$. Thus, $\delta^*(\lambda x + (1-\lambda)x', \lambda c + (1-\lambda)c') + (\lambda c + (1-\lambda)c') \geq \delta^*(x, c) + c = \delta^*(x', c') + c'$, and also $\delta^*(x, c) + c = \lambda[\delta^*(x, c) + c] + (1-\lambda)[\delta^*(x', c') + c']$. Then, it holds that

$$\delta^*(\lambda x + (1-\lambda)x', \lambda c + (1-\lambda)c') \geq \lambda\delta^*(x, c) + (1-\lambda)\delta^*(x', c'). \tag{9.22}$$

This holds for any c with $c' = c'(c)$. When $c \to \infty$, $c'(c) \to \infty$. Since $\lim_{c \to +\infty} \delta^*(x, c) = u^*(x)$ and $\lim_{c' \to +\infty} \delta^*(x', c') = u^*(x')$, we have, by (9.22), $u^*(\lambda x + (1-\lambda)x') \geq \lambda u^*(x) + (1-\lambda)u^*(x')$. □

The monotonicity of $u^*(x)$ follows from Lemma 9.4.2 that if $(x, 0) \succsim (x', 0)$, then $u^*(x) = \lim_{c \to \infty} \delta^*(x, c) \geq \lim_{c \to \infty} \delta^*(x', c) = u^*(x')$. Sometimes, we need strict monotonicity of $u^*(x)$, which is obtained by the following condition for $\succsim$. We say that $\succsim$ over $Z_o \times R_+$ is *strict increasing* with $x \in Z_o$ iff for any $x, x' \in Z_o$ with $x > x'$, there is an $\varepsilon > 0$ such that $(x, c) \succsim (x', c + \varepsilon)$ for any $c \in R_+$. This guarantees the strict monotonicity of u^* over $X = Z_o$ derived in Theorem 9.3.1.

Lemma 9.4.4 *Suppose $\succsim$ over $Z_o \times R_+$ is* strict increasing *with $x \in Z_o$. Then, $u^* : Z_o \to R$ is strictly increasing.*

Proof Let $x > x'$. By (9.3) and strict increasingness for $\succsim$, we have $(x_o, \delta^*(x, c) + c) \sim (x, c) \succsim (x', c + \varepsilon) \sim (x_o, \delta^*(x', c + \varepsilon) + c + \varepsilon)$. Hence, by C0 and C1, $\delta^*(x, c) + c \geq \delta^*(x', c + \varepsilon) + c + \varepsilon$. When $c \to +\infty$, this inequality implies $u^*(x) \geq u^*(x') + \varepsilon$. □

Finally, we give a comment on the converse of Lemma 9.4.3. It was shown in Kaneko [6] that u^* derived C0 to C3 and $C4^{PI}$ in Proposition 9.2.1 is concave if and only if the preference relation $\succsim$ is convex, where X is assumed to have a convex structure. A question is whether Lemma 9.4.3 holds in the form of "if and only if". This is answered negatively, since if u^* is linear, it is concave as well as convex; it is possibly derived from a non-convex $\succsim$. A counterexample is given below. Of course, it holds that if u^* is concave, there is a convex $\succsim$ such that u^* is derived from $\succsim$.

Let $X \times R_+ = Z_o \times R_+$ with $Z_o = \{0, ..., 4\}$ (Z_o can be R_+). Consider the utility function U defined by

$$U(x, c) = x - \frac{\sqrt{x}}{c+1} + 2c.$$

Then δ^* derived by (9.3) is $\delta^*(x, c) = (x - \frac{\sqrt{x}}{c+1})/2$, and the derived $U^*(x, c)$ is given as $\delta^*(x, c) + c = U(x, c)/2$. In this case, $u^*(x) = \lim_{c \to +\infty} \delta^*(x, c) = x/2$, which is concave

in Z_o. However, $U(x, c)$ is not quasi-concave (equivalently, $\succsim$ is not convex). Indeed, consider $(4, 0)$ and $(0, 1)$. Then, $U(4, 0) = 2 = U(0, 1)$. The middle point is $\frac{1}{2}(4, 0) + \frac{1}{2}(0, 1) = (2, \frac{1}{2})$, and $U(2, \frac{1}{2}) = 2 - \frac{\sqrt{2}}{\frac{3}{2}} + \frac{1}{2} = 2 - \frac{2}{3}\sqrt{2} + \frac{1}{2} < 2$.

9.4.2 *Lindahl-Ratio Equilibrium for a Public Goods Economy*

Let us apply the results in Sect. 9.4.1 to the theory of Lindahl-ratio equilibrium in a public goods economy (cf., Kaneko [7], van den Nouweland et al. [20], and van den Nouweland [21]).

Let $X = Z_o$. A cost function $C : Z_o \to R_+$ is given as a convex and strictly increasing function over X with $C(0) = 0$. Each agent $i \in N$ has a preference relation $\succsim_i$ over $Z_o \times R_+$ and an income $I_i \geq 0$. We call $E_B = (C; \{\succsim_i\}_{i\in N}, \{I_i\}_{i\in N})$ the *base (public good) economy*. We assume that each $\succsim_i$ satisfies C0-C3, $C4^{NM}$, C5, and that $\succsim_i$ is convex over $Z_o \times R_+$ and strictly increasing with $x \in Z_o$.

We say that $r = (r_1, ..., r_n)$ is a *ratio vector* iff $\sum_{i\in N} r_i = 1$ and $r_i > 0$ for all $i \in N$. A pair $(x^*, r) = (x^*, (r_1, ..., r_n))$ of an $x^* \in Z_o$ and a ratio vector $(r_1, ..., r_n)$ is called a (Lindahl-) *ratio equilibrium* in the base economy E_B iff for all $i \in N$,

$$r_i C(x^*) \leq I_i; \tag{9.23}$$

$$(x^*, I_i - r_i C(x^*)) \succsim_i (x, I_i - r_i C(x)) \text{ for all } x \in Z_o \text{ with } r_i C(x) \leq I_i. \tag{9.24}$$

That is, with an appropriate choice of a ratio vector for cost-sharing, every agent agrees on the same choice x^*.

Kaneko [7] formulated this concept taking $X = R_+$, and proved the existence of a ratio equilibrium, using the standard fixed-point argument. His result cannot directly be obtained when $X = Z_o$, since Z_o is a discrete set. Here, we first study a ratio equilibrium in an economy with quasi-linearity and then convert the result to E_B.

Now, for each $i \in N$, we have $u_i^* : Z_o \to R$ with $\lim_{c\to+\infty} \delta_i^*(x, c) = u_i^*(x)$ for each $x \in Z_o$. The quasi-linear approximation is given as $E_Q = (C; \{u_i^*\}_{i\in N}, \{I_i\}_{i\in N})$. In E_Q, a pair $(x^*, r) = (x^*, (r_1, ..., r_n))$ is called a *ratio equilibrium in* E_Q iff (9.23) and (9.25) hold:

$$u_i^*(x^*) + I_i - r_i C(x^*) \geq u_i^*(x) + I_i - r_i C(x) \text{ for all } x \in Z_o \text{ with } I_i \geq r_i C(x). \tag{9.25}$$

When I_i is large enough, we can ignore I_i in (9.25).

The analysis of ratio equilibrium is much simpler in the economy E_Q than in the base economy E_B. We consider the maximization of the total surplus in E_Q:

$$\max_{x\in Z_o} \left(\sum_{i\in N} u_i^*(x) - C(x) \right). \tag{9.26}$$

Then, we have the existence of an optimal solution $x^* \in Z_o$.

We have the following lemma. Recall that each $\succsim_i$ satisfies C0 to C3, C4^{NM}, C5, and that $\succsim_i$ is convex over $Z_o \times R_+$ and strictly increasing with $x \in Z_o$.

Lemma 9.4.5 *Let x^* be a solution for (9.26). Then, there is a ratio vector $r = (r_1, ..., r_n)$ such that (r, x^*) is a ratio equilibrium in the economy E_Q.*[13]

Proof When $z_o = 0$, this lemma holds with any ratio vector r. We assume $z_o > 0$. For a function $f : Z_o \to R$, we denote the left and right differentials $f^-(x) = f(x) - f(x-1)$ and $f^+(x) = f(x+1) - f(x)$ at $x \in Z_o$, where $f^-(0)$ or $f^-(z_o)$ are not defined. Let $g(x) = \sum_{i \in N} u_i^*(x) - C(x)$, which is a concave function. We consider the three cases: $x^* = 0$, $0 < x^* < z_o$, and $x^* = z_o$.

Suppose $0 < x^* < z_o$. Then, it holds that

$$g^+(x^*) = \sum_{i \in N} u_i^{*+}(x^*) - C^+(x^*) \leq 0 \leq g^-(x^*) = \sum_{i \in N} u_i^{*-}(x^*) - C^-(x^*). \tag{9.27}$$

For $\theta \in [0, 1]$, let $\alpha_i(\theta) = \theta u_i^{*+}(x^*) + (1-\theta) u_i^{*-}(x^*)$ for all $i \in N$. Then, $\sum_{i \in N} \alpha_i(\theta^*) = \theta^* \sum_{i \in N} u_i^{*+}(x^*) + (1-\theta^*) \sum_{i \in N} u_i^{*-}(x^*)$, and since u_i^* is strictly increasing by Lemma 9.4.4, we have $\alpha_i(\theta) > 0$.

In fact, there is a $\theta^* \in [0, 1]$ such that

$$C^-(x^*) \leq \sum_{i \in N} \alpha_i(\theta^*) \leq C^+(x^*). \tag{9.28}$$

Let us see this. By (9.27),

$$\sum_{i \in N} u_i^{*+}(x^*) \leq C^+(x^*) \text{ and } C^-(x^*) \leq \sum_{i \in N} u_i^{*-}(x^*). \tag{9.29}$$

Suppose $C^-(x^*) \leq \sum_{i \in N} u_i^{*+}(x^*)$. By (9.29), we also have $\sum_{i \in N} u_i^{*+}(x^*) \leq C^+(x^*)$. In this case, we can put $\theta^* = 1$; Eq. (9.28) holds. In the case $\sum_{i \in N} u_i^{*-}(x^*) \leq C^+(x^*)$, we have a parallel argument; we can put $\theta^* = 0$. Finally, consider the case $\sum_{i \in N} u_i^{*+}(x^*) < C^-(x^*)$ and $C^+(x^*) < \sum_{i \in N} u_i^{*-}(x^*)$. Since $C^-(x^*) \leq C^+(x^*)$ by the convexity of C, there is some θ^* satisfying (9.28). In the three cases, we have (9.28).

Let $r_i = \alpha_i(\theta^*) / \sum_{j \in N} \alpha_j(\theta^*)$ for all $i \in N$. Then, since $u_i^{*+}(x^*) \leq u_i^{*-}(x^*)$, it holds that

$$\begin{aligned} u_i^{*+}(x^*) - (\theta^* u_i^{*+}(x^*) + (1-\theta^*) u_i^{*-}(x^*)) \leq 0 \\ \leq u_i^{*-}(x^*) - (\theta^* u_i^{*+}(x^*) + (1-\theta^*) u_i^{*-}(x^*)). \end{aligned} \tag{9.30}$$

Since $r_i = \alpha_i(\theta^*) / \sum_{j \in N} \alpha_j(\theta^*)$, we have, by (9.28),

$$u_i^{*+}(x^*) - r_i C^+(x^*) \leq u_i^{*+}(x^*) - (\theta^* u_i^{*+}(x^*) + (1-\theta^*) u_i^{*-}(x^*)).$$

[13] This is a variant of the method of obtaining the existence of a competitive equilibrium from the maximization of the total social surplus, which was first given by Negishi [16].

Using the first inequality of (9.30), we have $u_i^{*+}(x^*) - r_i C^+(x^*) \leq 0$. Similarly, it holds that $0 \leq u_i^{*-}(x^*) - r_i C^-(x^*)$. Thus, since $u_i^*(x) + I_i - r_i C(x)$ is concave, the solution x^* maximizes $u_i^*(x) + I_i - r_i C(x)$ for each $i \in N$.

Suppose $x^* = 0$. Then, $\sum_{i \in N} u_i^{*+}(0) \leq C^+(0)$. Let $\alpha_i = u_i^{*+}(0) > 0$ and $r_i = u_i^{*+}(0) / \sum_{j \in N} u_j^{*+}(0)$. Now, we have $u_i^{*+}(0) - r_i C^+(0) \leq 0$ for all $i \in N$. This means that $x^* = 0$ maximizes $u_i^*(x) + I_i - r_i C(x)$ for each $i \in N$. In the case where $x^* = z_o$, we have a parallel argument. □

Now, we have the conversion theorem under C0-C4 for $\succsim_i$ over $Z_o \times R_+$. This theorem needs neither the convexity nor strict increasingness for $\succsim_i$, since the existence of a ratio equilibrium is assumed.

Theorem 9.4.2 (Conversion of a ratio equilibrium from E_Q to E_B) Let $(x^*, r) = (x^*, (r_1, ..., r_n))$ be a ratio equilibrium in E_Q. Then, for any $\varepsilon > 0$, there is an I^* such that for any $i \in N$ and $I_i \geq I^*$,

$$I_i \geq r_i C(x^*); \tag{9.31}$$

$$U_i^*(x^*, I_i - r_i C(x^*)) + \varepsilon > U_i^*(x, I_i - r_i C(x)) \text{ for any } x \in Z_o \text{ with } I_i \geq r_i C(x). \tag{9.32}$$

Proof We choose I_i^0 so that $I_i^0 \geq r_i C(x^*)$. Now, let $x \in Z_o$. If $I_i^0 < r_i C(x)$, (9.32) holds in the trivial sense. In the following, consider the case $I_i^0 \geq r_i C(x)$. Then, by (9.25), we have

$$u_i^*(x^*) + I_i - r_i C(x^*) \geq u_i^*(x) + I_i - r_i C(x). \tag{9.33}$$

Take $\varepsilon > 0$. Then, by Theorem 9.3.1, we can choose an $I_i^1 \geq I_i^0$ so that for any $I_i \geq I_i^1$,

$$\begin{aligned} \left| U_i^*(x^*, I_i - r_i C(x_i^*)) - (u_i^*(x^*) + I_i - r_i C(x^*)) \right| &< \varepsilon/2 \\ \left| U_i^*(x, I_i - r_i C(x)) - (u_i^*(x) + I_i - r_i C(x)) \right| &< \varepsilon/2. \end{aligned}$$

Using these inequalities and (9.33), we have (9.32)

$$\begin{aligned} U_i^*(x^*, I_i - r_i C(x^*)) &> u_i^*(x^*) + I_i - r_i C(x^*) - \varepsilon/2 \\ &\geq u_i^*(x) + I_i - r_i C(x) - \varepsilon/2 \\ &> U_i^*(x, I_i - r_i C(x_i)) - \varepsilon/2 - \varepsilon/2 \\ &= U_i^*(x, I_i - r_i C(x_i)) - \varepsilon. \end{aligned}$$

The above choice of $I_i^1 = I_i^1(x)$ depends upon agent $i \in N$ and $x \in Z_o$. However, because N and Z_o are finite, it suffices to take $I^* = \max\{I_x^1 : i \in N \text{ and } x \in Z_o\}$. □

9.5 Extension to Expected Utility Theory

Quasi-linear utility functions are also used in the environment with risks. In this case, the characterization of quasi-linearity should be connected to expected utility theory, or *vice*

versa. This was discussed in Kaneko–Wooders [9]. Here, we will discuss the extension of Theorem 9.3.1.

Let $m_F(X \times R_+) := \{f : X \times R_+ \to [0, 1] : \sum_{(x,c)\in S} f(x, c) = 1$ for some finite subset S of $X \times R_+\}$ (i.e., the set of all probability distributions with finite supports over $X \times R_+$). Regarding $m_F(X \times R_+)$ as a subset of the linear space of all real-valued functions endowed with the standard sum and scalar (real) multiplication, $m_F(X \times R_+)$ is a convex set (i.e., if $f, g \in m_F(X \times R_+)$ and $\lambda \in [0, 1]$, the convex combination (mixture) $\lambda f * (1 - \lambda)g$ belongs to $m_F(X \times R_+)$). Let $\succsim^e$ be a binary relation over $m_F(X \times R_+)$.

We assume the following:

Condition E0 (Complete preordering): $\succsim^e$ is a complete and transitive relation on $m_F(X \times R_+)$;

Condition E1 (Intermediate value): If $f \succ^e g \succ^e h$, then $\lambda f * (1 - \lambda)h \sim^e g$ for some $\lambda \in [0, 1]$;

Condition E2 (Independence): For any $f, g, h \in m_F(X \times R_+)$ and $\lambda \in (0, 1)$,

(1): $f \succ^e g$ implies $\lambda f * (1 - \lambda)h \succ^e \lambda g * (1 - \lambda)h$;

(2): $f \sim^e g$ implies $\lambda f * (1 - \lambda)h \sim^e \lambda g * (1 - \lambda)h$.

It is known (cf., Herstein–Milnor [3], Fishburn [2], Kaneko–Wooders [9]) that these three conditions are enough to derive a utility function $U^e : m_F(X \times R_+) \to R$ representing $\succsim^e$ and satisfying $U^e(\lambda f * (1 - \lambda)g) = \lambda U^e(f) + (1 - \lambda)U^*(g)$ for all $f, g \in m_F(X \times R_+)$ and $\lambda \in [0, 1]$.

We can regard $X \times R_+$ as a subset of $m_F(X \times R_+)$ by the identity mapping. Restricting the preference relation $\succsim^e$ to $X \times R_+$, we have the preference relation over $\succsim$ on $X \times R_+$, which satisfies Condition C0. Conditions E1-E2 require nothing about $\succsim$ over the base set $X \times R_+$. We can assume C1-C4 on $\succsim$. We denote the restriction of U^e to the base set $X \times R_+$ also by U^*.

Theorem 9.5.1 (Expected utility theory version) *Suppose that a preference relation $\succsim^e$ over $m_F(X \times R_+)$ satisfies E0-E2, and that the derived preference $\succsim$ on $X \times R_+$ satisfies C1-C4.*
(1)*: There is a utility function $U^e : m_F(X \times R_+) \to R$ such that*

$$U^e(f) = \sum_{(x,c)\in T_f} f(x, c)U^e(x, c) \text{ for each } f \in m_F(X \times R_+), \tag{9.34}$$

where T_f is a finite support of $f \in m_F(X \times R_+)$.
(2)*: There is a (strictly) monotone $f : R \to R$ such that*

$$U^e(x, c) = f(\delta^*(x, c) + c) \text{ for all } (x, c) \in X \times R_+. \tag{9.35}$$

(3)*: There is a function $u^* : X \to R$ such that (9.8) holds for each $x \in X$.*

Proof (1) is known from expected utility theory.
(2): It is shown in Lemma 9.3.1 that over the domain $X \times R_+$, the relation $\succsim$ is represented by the function $\delta^*(x, c) + c$. This implies that if $\delta^*(x, c) + c = \delta^*(x', c') + c'$, then $U^e(x, c) = U^e(x', c')$. Hence, we can define a function $f : \{\delta^*(x, c) + c : (x, c) \in X \times R_+\} \to R$ by $f(\delta^*(x, c) + c) = U^e(x, c)$ for all $(x, c) \in X \times R_+$. This f is monotone, and can be extended to R.
(3): This is simply Theorem 9.3.1. □

We have still the difference that Theorem 9.5.1.(2) is stated in terms of $U^e = f(\delta^*(x, c) + c)$ rather than $\delta^*(x, c) + c$. Expected utility theory is cardinal, while the theory in Sect. 9.3 is ordinal. Hence, it may be informative to connect (3) with (2) directly. This connection is made to assume risk neutrality:

E3: (Risk Neutrality): $\frac{1}{2}(x_o, c) * \frac{1}{2}(x_o, c') \sim^e (x_o, \frac{1}{2}c + \frac{1}{2}c')$ for $c, c' \in R_+$.

The preference relation $\succsim^e$ is risk neutral with respect to the axis of composite commodity at the worst x_o. This is a connection between our theory and expected utility theory. Then, we have the following lemma on the function f given by Theorem 9.5.1.(2):

Lemma 9.5.1 *There are $\alpha > 0$ and β such that $f(c) = \alpha c + \beta$ for $c \in R_+$.*

Proof Recall (9.5) of Lemma 9.3.1: $\delta^*(x_o, c) = 0$ for all $c \in R_+$. Thus, E3 is expressed as

$$\frac{1}{2}f(c) + \frac{1}{2}f(c') = f(\frac{1}{2}c + \frac{1}{2}c') \text{ for all } c, c' \in R_+.$$

This implies that for some $\alpha > 0$ and β, $f(c) = \alpha c + \beta$ for $c \in R_+$. □

Under the above assumptions on $\succsim^e$, the function f is linear, and in particular, we can assume

$$U^e(x, c) = \delta^*(x, c) + c \text{ for all } (x, c) \in X \times R_+. \tag{9.36}$$

In sum, we obtain the approximately quasi-linear function by adding E3 in the extended theory. Of course, if we assume risk aversion (lover), f is a concave (convex) function.

9.6 Summary and Remaining Issues

We gave characterizations of a preference relation $\succsim$ to be approximately represented by a quasi-linear utility function for large incomes. The main condition is C4, which is a weakening of the parallel indifferences condition $C4^{PI}$. It guarantees the limit function $u^*(x) = \lim_{c \to +\infty} \delta^*(x, c)$, which is a representation of the monetary equivalence of the transition from the origin x_o to alternative x.

We provided another approach in terms of the normality condition $C4^{NM}$. Under C0 to C3, condition $C4^{NM}$ and boundedness C5 imply C4. These are easier to check whether a given relation satisfies approximate quasi-linearity. We also made an explicit connection between our approximate quasi-linearity and expected utility theory.

We gave two applications of our results to the theories of cooperative games with side payments and of Lindahl-ratio equilibrium for a public goods economy with quasi-linearity. We discussed the conversions the results in these theories to the base models. We started our considerations with the base models and went to the limit cases; the conversions went back to the base model. In the end of Sect. 9.3.1, we gave a brief discussion on the other direction directly from the limit E_Q to a base model E_B.

Mathematically speaking, condition C4 excludes some familiar utility functions given in closed forms. In the end of Sect. 9.3.1, we gave how to avoid this difficulty and also argued that the existence of the limit function $u^*(x)$ is justified in the case where the composite commodity behind c is rich enough or the alternatives in X are rich enough.

Nevertheless, there remain various issues. Here, only two issues are mentioned. The first one is how to formulate the richness behind the composite commodity c or the richness of

alternatives in X. Perhaps, this is an important but difficult problem. Another issue is to evaluate the standard interpretation of no-income effect in terms of local approximation, mentioned in the paragraph after (9.15). This may involve double approximations "large incomes" and "small expenditures". Although this may turn to be an inappropriate interpretation, it would be helpful to understand the nature of quasi-linearity and/or no-income effect.

References

1. Aumann, R.J.: Linearity of unrestricted transferable utilities. Naval Res. Logist. Q. **7**, 281–284 (1960)
2. Fishburn, P.: The Foundations of Expected Utility. Springer-Science-Bussiness Media, Dordrecht (1982)
3. Herstein, I.N., Milnor, J.: An axiomatic approach to measurable utility. Econometrica **21**, 291–297 (1953)
4. Hicks, J.R.: A Value and Capital. Oxford University Press, Oxford (1939)
5. Hicks, J.R.: A Revision of Demand Theory. Clarendon Press, Oxford (1956)
6. Kaneko, M.: Note on transferable utility. Int. J. Game Theory **6**, 183–185 (1976)
7. Kaneko, M.: The ratio equilibrium and a voting game in a public goods economy. J. Economic Theory **16**, 123–136 (1977)
8. Kaneko, M.: Housing market with indivisibilities. J. Urban Econ. **13**, 22–50 (1983)
9. Kaneko, M., Wooders, M.H.: Utility Theories in Cooperative Games. Handbook of Utility Theory, vol. 2, pp. 1065–1098. Kluwer Academic Press, Dordrecht (2004)
10. Kaneko, M., Ito, T.: An equilibrium-econometric analysis of rental housing markets with indivisibilities. In: Lina Mallozzi, L., Pardalos, P. (eds.) Spatial Interaction Models: Facility Location using Game Theory, pp. 193–223. Springer (2017)
11. Miyake, M.:Asymptotically quasi-linear utility function. TERGN Working Paper No. 154, Tohoku University (2000)
12. Miyake, M.: On the applicability of Marshallian partial-equilibrium analysis. Math. Soc. Sci. **52**, 176–196 (2006)
13. Miyake, M.: Convergence theorems of willingness-to-pay and willingness-to-accept for non-maket goods. Soc. Choice Welf. **34**, 549–570 (2010)
14. Maschler, M., Solan, E., Zamir, S.: Game Theory. Cambridge University Press, Cambridge (2013)
15. Mas-Collel, A., Whinston, M., Green, J.: Microeconomic Theory. Oxford University Press, Oxford (1995)
16. Negishi, T.: Welfare economics and existence of an equilibrium for a competitive economy. Metroeconomica **12**, 92–97 (1960)
17. Osborne, M.J., Rubinstein, A.: A Course in Game Theory. The MIT Press, London (1994)
18. Royden, H.L., Fitzpatrick, P.M.: Real Analysis, Prentice Hall, Upper Saddle River (2010)
19. Shapley, L.S., Shubik, M.: Competitive outcomes in the cores of market games. Int. J. Game Theory **4**, 229–237 (1975)
20. van den Nouweland, A., Tijs, S., Wooders, M.H.: Axiomatization of ratio equilibria in public good economies. Soc. Choice Welf. **19**, 627–6363 (2002)
21. van den Nouweland, A.: Lindahl and equilibrium. In: Binder, C. et al. (ed.) Individual and Collective Choice and Social Welfare, pp. 335–362. Springer (2015)
22. Vives, X.: Small income effects: a Marshallian theory of consumer surplus and downward sloping demand. Rev. Econ. Stud. **54**, 87–103 (1987)

Chapter 10
Cooperative Games in Networks Under Uncertainty on the Costs

L. Mallozzi and A. Sacco

10.1 Introduction

In many situations arising from Engineering or Economics, as in transportations and logistics, an important aspect is to find efficient and optimal plans to design collaborative service networks when two or more agents are involved. For example, efficiency can be measured in lower cost or more flexibility. An important aspect of the collaboration is to decide on how to share the profits, the cost, or some resources. In literature several sharing mechanisms or cost allocations can be found, and some of them are founded in game theory (see e.g., [17–19, 21, 25]). Of many problems related to collaborating in transportation, some of them regard transportation planning, traveling salesman, vehicle routing, or minimal cost spanning tree (see e.g., [4, 7, 10, 15]).

In this chapter, we approach a cooperative game model that describe a multicommodity network flow problem: the objective in this problem is to share the revenue generated by simultaneously shipping different commodities. Since different possibilities may appear in terms of paths, a maximum revenue (or a minimum cost) network problem can be considered too and solved by using some game theory tools.

Our first assumption is that the network is given and does not have any cycle, so that each agent that has to ship his commodity from an origin to a destination point has just one route for the shipment. In this case a revenue sharing problem arises and a cooperative game problem can be set between agents: the core of the corresponding

L. Mallozzi (✉)
Department of Mathematics and Applications, University Federico II, V. Claudio 21, 80125 Naples, Italy
e-mail: mallozzi@unina.it

A. Sacco
Department of Methods and Models for Economics, Territory and Finance, Sapienza University, V. C. Laurenziano 9, 00161 Rome, Italy
e-mail: armando.sacco@uniroma1.it

S. K. Neogy et al. (eds.), *Mathematical Programming and Game Theory*,
Indian Statistical Institute Series, https://doi.org/10.1007/978-981-13-3059-9_10

cooperative game is not empty under some concavity conditions on the costs (see e.g., [22, 24]).

As in reality, some uncertainty may be in the data of the problem. For example, in [8] the net return that each agent has for the shipment of the commodity has been considered as a real interval, not a real number. Then an interval cooperative approach has been presented in order to provide interval core solutions. The first example of the use of cooperation under interval uncertainty was [5], where it is applied to bankruptcy situations, and later further extensively studied (see for example [1–3] and the references section of [6] for more).

The literature is completed by a stream of non-classical models of cooperative games incorporating some kind of uncertainty such as games with random payoffs [17, 19, 21], games with fuzzy uncertainty [14] or the so-called cooperative fuzzy interval games, a combination of fuzzy and interval games [13].

Main contribution of this chapter is to present the model under cost uncertainty by considering a probability distribution on the set (that is an interval) of the possible values of the costs. In this case we study the stochastic cooperative resulting game and give conditions in order to have a non-empty core. The situation of an expansion cost effect is also discussed, i.e., we study the case where the upper bound of the cost is proportional to the cost according to an expansion factor.

In the chapter, same mathematical preliminaries are recalled in Sect. 10.2, the network and the model are presented in Sect. 10.3 together with some existence results and some examples. Some new research suggestions are discussed in the concluding section.

10.2 Preliminaries

A cooperative game is an ordered pair $< N, v >$ where $N = \{1, \ldots, n\}$ $(n \in \mathbb{N})$ is the set of the players and $v : 2^N \to \mathbb{R}$ is the characteristic function from the set 2^N of all possible coalitions of players N to a set of payments that satisfies $v(\emptyset) = 0$. The function describes how much collective payoff a set of players can gain by forming a coalition, and the game is sometimes called a *value game* or a *profit game*. The players are assumed to choose which coalitions to form, according to their estimate of the way the payment will be divided among coalition members; N is called the grand coalition.

A cooperative game can also be defined with a characteristic cost function $c : 2^N \to \mathbb{R}$ satisfying $c(\emptyset) = 0$. In this setting, the characteristic function c represents the cost of a set of players accomplishing the task together. A game of this kind is known as a cost game. Although most cooperative game theory deals with profit games, the duality of the two approaches made them equivalent (the games v and $-c$ are strategically equivalent).

Several solution concepts have been introduced in the literature. A natural and well-known solution concept for this cooperative game is the core [17, 20, 23]. The core $\mathscr{C}(v)$ of the cooperative game $< N, v >$ gives a share of the worth of the grand

coalition satisfying the so-called coalitional efficiency and is defined by

$$\mathscr{C}(v) = \{(x_1, \ldots, x_n) \in \mathbb{R}^n : \sum_{i \in N} x_i = v(N),\ \sum_{i \in S} x_i \geq v(S), \forall\ S \subseteq N\}.$$

Recall that a cooperative game $< N, v >$ is convex if

$$v(S \cup T) + v(S \cap T) \geq v(S) + v(T),\ \forall S, T \in 2^N$$

and if the game is convex, the core is non-empty [20, 23]. Other solution concepts are the Shapley value, the nucleolus, and many others.

The choice of the core as solution concept is linked with the assumption of acyclic network: the core of a game with cycle may be empty.

Example 1 Let us consider a network design situation (N, G, h, OD, r, IC) where $N = \{1, 2, 3\}$ is the set of players, $V = \{1, 2, 3, 4, 5, 6, 7\}$ is the set of vertexes and $E = \{1, 2, 3, 4, 5, 6, 7, 8, 9\}$ is the set of directed edges (with $G = (V, E)$), $h = (1, 1, 1)$ represents the units of commodity shipped, $OD = \big((1, 7), (3, 7), (5, 7)\big)$ is the set of ordered pair of origin/destination, $r = (3, 2, 4)$ is the vector of revenues and $c_j(y) = \sqrt{y}$, $j \in E$ is the cost function. In this case (see Fig. 10.1)

$$P_1 = \{18, 67, 123457, 1239\},\ P_2 = \{2167, 28, 3457, 39\},\ P_3 = \{57, 49, 432167, 4328\}$$

$$Q_1 = \{18, 123457, 1239, 2167, 432167\},\ Q_2 = \{123457, 1239, 2167, 28, 432167, 4328\}$$

$$Q_3 = \{123457, 1239, 3457, 39, 432167, 4328\},\ Q_4 = \{123457, 3457, 49, 432167, 4328\}$$

$$Q_5 = \{123457, 3457, 57\},\ Q_6 = \{67, 2167, 432167\},$$

$$Q_7 = \{67, 123457, 2167, 3457, 57, 432167\},\ Q_8 = \{18, 28, 4328\},\ Q_9 = \{1239, 39, 49\}$$

and it is easy to compute

$$c(\{1\}) = c(\{2\}) = c(\{3\}) = 2$$

$$c(\{1, 2\}) = c(\{2, 3\}) = c(\{1, 3\}) = 2 + \sqrt{2}$$

$$c(\{1, 2, 3\}) = 4 + \sqrt{2}.$$

The characteristic function is

$$v(\{1\}) = 1, v(\{2\}) = 0, v(\{3\}) = 2$$

$$v(\{1, 2\}) = 3 - \sqrt{2}, v(\{2, 3\}) = 4 - \sqrt{2}, v(\{1, 3\}) = 5 - \sqrt{2}$$

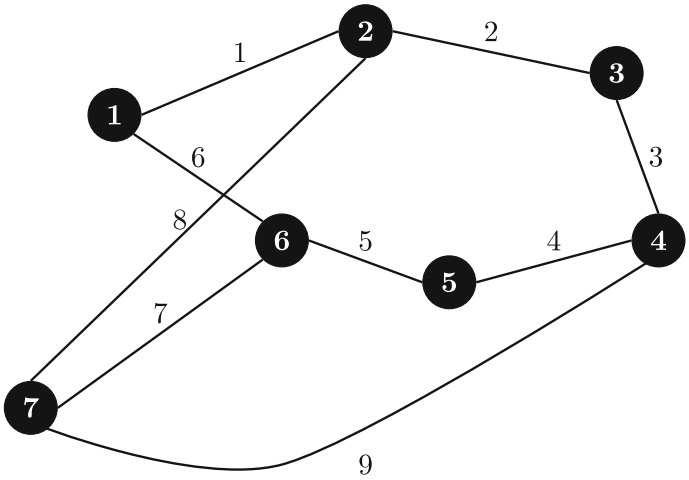


Fig. 10.1 Scheme for the network described into Example 1

$$v(\{1, 2, 3\}) = 5 - \sqrt{2}$$

and the core of this game is empty [8, 24].

Sometimes we deal in reality with uncertainty: we do not know exactly the worth of a coalition S, but we can have an estimate of a lower bound and an upper bound of it. A way to deal with uncertain characteristic function is to consider interval cooperative games that allow to study the case where the value of a coalition is a real interval by using interval analysis tools. Since the work of Branzei, Dimitrov and Tijs in 2003 to study bankruptcy situations [5] many examples of interval games were studied in the literature (see [1–3, 9, 11, 13] and the references section of [6] for more).

A cooperative interval game is an ordered pair $< N, w >$ where N is the set of the players and $w : 2^N \to \text{IR}$ is the characteristic function such that $w(\emptyset) = [0, 0]$. Here IR be the set of real intervals $\text{IR} = \{[\underline{I}, \bar{I}] \subset \mathbb{R}, \underline{I}, \bar{I} \in \mathbb{R}, \underline{I} \leq \bar{I}\}$. We denote by $w(S) = [\underline{w}(S), \bar{w}(S)]$ the worth of the coalition S. By considering the partial order $I \succeq J$ iff $\underline{I} \geq \underline{J}$ and $\bar{I} \geq \bar{J}$, it is possible to introduce the core solution concept for the interval game, namely the interval core that is defined by

$$\mathscr{C}(w) = \{(I_1, \ldots, I_n) \in \text{IR}^n : \sum_{i \in N} I_i = w(N), \ \sum_{i \in S} I_i \succeq w(S), \forall\, S \subseteq N\}.$$

We point out that this approach can leave some ambiguity on the preferences since interval core solutions offer many possibilities of profit sharing scheme. In order to refine the model, we introduce some additional probabilistic information and present a stochastic version of the network design situation.

10.3 The Network and the Model

The game is defined by a set of players $N = \{1, \ldots, n\}$ and by a graph $G = (V, E)$, where $V = \{1, \ldots, k\}$ is the finite set of k vertexes or nodes and $E = \{1, \ldots, m\}$ the set of m directed edges (k and m are natural numbers). The couple (o_i, d_i), with $o_i, d_i \in V$, is an ordered pair of nodes, identifying origin and destination, between which each player $i \in N$ has to ship $h_i > 0$ units of a commodity. We denote $h = (h_1, \ldots, h_n)$ and $OD = \big((o_1, d_1), \ldots, (o_n, d_n)\big)$ the vectors of $\mathbb{R}^n$ and $\mathbb{R}^{2n}$ respectively.

Moreover the shipment produces for each player i a return r_i. The setting of the network provides that at the initial status the capacity of each edge of E for accommodating shipments of the players' commodities is zero, and there is an investment cost $c_j(y)$ for installing y units of capacity on edge $j \in E$. Considering admissible network, that is the network is able to satisfy the requirements of any player that participates to the construction, then any coalition $S \subseteq N$ of players could construct capacities on the edges of E. The assumption of the model is that the coalition S chooses the admissible network of minimum cost.

Two other relevant sets of the network are the set $P_i = \{\text{path connecting } o_i \text{ and } d_i\}$ for any player $i \in N$ and the set $Q_j = \{\text{path of edges from E including } j\}$ for any edge $j \in E$. A path is the union of consecutive edges (ijk is the path given by edge i, then edge j, then edge k). We consider in this chapter acyclic networks, so that each P_i consists of a single path denoted by p_i. Here it is implicitly assumed that players have to ship h, even if it gives them a negative payoff.

The sum of the costs of each edge j by considering all the players of coalition S that are using that edge j when they use the path $p_i, \forall i \in S$, is given by the quantity

$$c(S) = \sum_{j \in E} c_j \Bigg(\sum_{\substack{i : i \in S \\ p_i \in Q_j}} h_i \Bigg).$$

The vector $r = (r_1, \ldots, r_n)$ denotes the revenue profile vector ($r_i > 0$), while $IC = \{c_1, \ldots, c_m\}$ denotes the installing cost functions, where $c_j : [0, +\infty) \to [0, +\infty)$, $c_j(0) = 0$ and c_j is an increasing function over the entire domain. We call the tuple (N, G, h, OD, r, IC) a network design situation.

Definition 1 Given a network design situation (N, G, h, OD, r, IC), we define the network design cooperative game $< N, v >$ where N is the set of the players and $v : 2^N \to \mathbb{R}$ is the characteristic function such that $v(\emptyset) = 0$ and for each coalition $S \subseteq N$ the worth of the coalition is given by

$$v(S) = \sum_{i \in S} r_i - c(S).$$

This game has been studied in [24] and the extension to the case of interval uncertainty in rewards in [8]. By assuming concave cost functions c_j, $j \in E$, the

cooperative game is a convex game and there exist core solutions and interval core solutions.

Proposition 1 *Let (N, G, h, OD, r, IC) be a network design situation where c_j, $j \in E$, are concave cost functions. Then the core of the cooperative game $< N, v >$ is not empty.*

Proof The proof follows by the supermodularity of the game [24]. Here we give a direct proof. Let us prove that the game is convex, i.e., $\forall S, T \in 2^N$, we have that

$$v(S \cup T) + v(S \cap T) \geq v(S) + v(T).$$

Since

$$\sum_{i \in S} r_i + \sum_{i \in T} r_i = \sum_{i \in S \cup T} r_i + \sum_{i \in S \cap T} r_i$$

and for each j the function $-c_j(t + \delta) + c_j(t)$ is increasing in t for any $\delta > 0$, we have:

$$v(S \cup T) + v(S \cap T) - v(S) - v(T) =$$

$$\sum_{j \in E} \Big[-c_j\Big(\sum_{\substack{i : i \in S \cup T \\ p_i \in Q_j}} h_i \Big) - c_j\Big(\sum_{\substack{i : i \in S \cap T \\ p_i \in Q_j}} h_i \Big) + c_j\Big(\sum_{\substack{i : i \in S \\ p_i \in Q_j}} h_i \Big) + c_j\Big(\sum_{\substack{i : i \in T \\ p_i \in Q_j}} h_i \Big) \Big] =$$

$$\sum_{j \in E} \big[-c_j(y'' + \delta) - c_j(y') + c_j(y'') + c_j(y' + \delta) \big] \geq 0$$

where

$$y' = \sum_{i \in S \cap T} h_i \leq y'' = \sum_{i \in S} h_i \quad \text{and} \quad \delta = \sum_{i \in (S \cup T) \setminus S} h_i = \sum_{i \in T \setminus (S \cap T)} h_i \geq 0.$$

Remark 1 Let us observe that the network design situation, given the installing cost functions IC and without revenue, is nothing but the congestion situation of [12, 16, 19], studied from a non-cooperative point of view: there exists for such games a pure Nash equilibrium, because they are potential games.

An analogous result holds for the extension to the case of interval uncertainty in rewards. We suppose that the reward of player i is an unknown value between a lower and an upper bound. We denote by $R = (R_1, \ldots, R_n) \in \mathrm{IR}^n$ the revenue profile vector and consider the network design situation (N, G, h, OD, R, IC).

Definition 2 (*Uncertainty on returns*) We define the network design cooperative game $< N, w >$ where N is the set of the players and $w : 2^N \to \mathrm{IR}$ is the characteristic function such that $w(\emptyset) = 0$ and for each coalition $S \subseteq N$ the worth of the coalition is given by

$$w(S) = \sum_{i \in S} R_i - c(S).$$

This game has been studied in [8], and by assuming concave cost functions c_j, $j \in E$, the cooperative game is a convex game and there exist interval core solutions. This kind of solutions give an indication of possible outcomes with a vagueness degree since a core solution is a set of values in between a lower bound and an upper bound.

Now, we consider uncertainty on installing costs. As it happens in concrete situations, we suppose that the cost for installing y units of capacity on each edge is not known, but players have a lower bound and an upper bound of it. More precisely, for any edge $i \in E$ there are two increasing functions $\underline{c}_j$ and $\overline{c}_j$ ($\underline{c}_j, \overline{c}_j : [0, +\infty) \to [0, +\infty)$, $\underline{c}_j(0) = \overline{c}_j(0) = 0$) with $\underline{c}_j(y) \leq \overline{c}_j(y)$ for all $y > 0$ such that the installing cost for y units can be any value in the real interval $[\underline{c}_j(y), \overline{c}_j(y)]$. Moreover, we assume that the uncertainty does not depend on the amount y of shipped commodity, but it is edge-specific.

Here we want to better describe the uncertainty by using a stochastic approach. We suppose that the cost of installing the edge j is a random variable t with probability density $\varphi_j(t)$.

One way to approach the problem of the uncertainty is considering the expected installing cost, that is given by

$$C_j(y) = \int_{\underline{c}_j(y)}^{\overline{c}_j(y)} t\varphi_j(t)dt,$$

for any $y \geq 0$. Then, given $\phi = \{\varphi_1, \ldots, \varphi_m\}$, we consider the network design situation $(N, G, h, OD, r, EIC)^\phi$ with cost distribution ϕ, where $EIC = \{C_1, \ldots, C_m\}$ is the vector of expected installing costs.

Definition 3 (*Uncertainty on costs*) Given $(N, G, h, OD, r, EIC)^\phi$, we define the network design cooperative game $< N, v >$ where N is the set of the players and $v : 2^N \to \mathbb{R}$ is the characteristic function such that $v(\emptyset) = 0$ and for each coalition $S \subseteq N$ the worth of the coalition is given by

$$v(S) = \sum_{i \in S} r_i - C(S)$$

being $C(S)$ the cost of the coalition S defined as

$$C(S) = \sum_{j \in E} C_j\Big(\sum_{i : i \in S, p_i \in Q_j} h_i\Big)$$

10.3.1 Extremal Situations

In this section, we start to consider the extremal situations in a network design model with cost uncertainty. This is the case where agents have additional information that allows to choose the best (resp. the worst) possible cost in the interval $[\underline{c}_j(y), \overline{c}_j(y)]$. Players, in an optimistic view, consider the best worth they receive under uncertainty, i.e., one can consider the extremal situations, namely in an optimistic view the worth can be

$$v^{opt}(S) = \sum_{i \in S} r_i - \sum_{j \in E} \underline{c}_j \Big(\sum_{\substack{i : i \in S \\ p_i \in Q_j}} h_i \Big)$$

and in a pessimistic view the worst possible case, i.e.,

$$v^{pes}(S) = \sum_{i \in S} r_i - \sum_{j \in E} \overline{c}_j \Big(\sum_{\substack{i : i \in S \\ p_i \in Q_j}} h_i \Big).$$

In these two cases the uncertainty is solved considering the lower and the upper bound for installing cost. In that way, we can study the minimum and the maximum worth for each possible coalition. The following example show a simple case with two players and concave cost functions.

Example 2 Let us consider the situation $(N, G, h, OD, r, EIC)^{\phi}$ where $N = \{1, 2\}$, $V = \{1, 2, 3, 4\}$ and $E = \{1, 2, 3\}$, $h = (1, 1)$, $OD = \big((1, 3), (2, 4)\big)$, $r = (5, 4)$, $\underline{c}_j(y) = \sqrt{y}, \overline{c}_j(y) = 2\sqrt{y}$, $j \in E$. The characteristic functions in the two extremal cases are:

$$v^{opt}(\{1\}) = 3, v^{opt}(\{2\}) = 2, v^{opt}(\{1, 2\}) = 7 - \sqrt{2},$$

$$v^{pes}(S)(\{1\}) = 1, v^{pes}(S)(\{2\}) = 0, v^{pes}(S)(\{1, 2\}) = 5 - 2\sqrt{2}.$$

Any vector $(x_1, x_2) : 3 \leq x_1 \leq 5 - \sqrt{2},\ x_2 = -x_1 + 7 - \sqrt{2}$ is in the core $\mathscr{C}(v^{opt})$ and any vector $(x_1, x_2) : 1 \leq x_1 \leq 5 - 2\sqrt{2},\ x_2 = -x_1 + 5 - 2\sqrt{2}$ is in the core $\mathscr{C}(v^{pes})$.

10.3.2 Expected Costs

If there is no additional information, we assume for each edge $j \in E$ that the cost is a random variable t uniformly distributed in the interval $[\underline{c}_j(y), \overline{c}_j(y)]$ with density $\varphi_j(t)$. Averaging between the lower cost and the upper cost, the expected installing cost is given by

$$C_j(y) = \frac{\underline{c}_j(y) + \overline{c}_j(y)}{2}$$

for any $y \geq 0$.

Example 3 Consider the network design situation $(N, G, h, OD, r, EIC)^{\phi}$ of the previous example, with uniform cost distribution. Now, the characteristic function is

$$v(\{1\}) = 2, v(\{2\}) = 1, v(\{1, 2\}) = 6 - 3/2\sqrt{2},$$

and any vector $(x_1, x_2) : 2 \leq x_1 \leq 5 - 3/2\sqrt{2}, x_2 = -x_1 + 6 - 3/2\sqrt{2}$ is in the core $\mathscr{C}(v)$. For a value of x_1 admissible in the pessimistic case and also in the average case, say $x_1 = 2.1$ we see that for the second player the share is, respectively, $x_2 = 0.08$ and $x_2 = 1.79$.

Proposition 2 *Let $(N, G, h, OD, r, EIC)^{\phi}$ be a network design situation with cost distribution ϕ, where $\underline{c}_j, \overline{c}_j$, for any $j \in E$, are concave cost functions. If costs follow a uniform distribution and the uncertainty is solved by mean of the expected cost, then the core of the cooperative game $< N, v >$ is not empty.*

Proof The network $(N, G, h, OD, r, EIC)^{\phi}$ is acyclic and if $t \sim U([\underline{c}_j(y), \overline{c}_j(y)])$, with $\underline{c}_j$ and $\overline{c}_j$ concave functions, then also the expected cost C_j is concave and the core is not empty.

10.3.3 Upper Bound Expansion

In real situations it can happen that the upper cost for installing a network is unknown when it is designed. To capture this possibility we consider a special case of the previous examples, considering an upper cost function $\overline{c}_j(y) = Ac_j(y)$, where A is an unknown parameter, i.e., for any edge $j \in E$ and transported quantity $y \geq 0$, the cost is a value in $[c_j(y), Ac_j(y)]$ and $A \geq 1$ is a real parameter describing an expansion effect on the costs $c_j(y)$.

Denoting with $\gamma(A)$ the density function of the parameter A, the expected installing cost for each edge j, $C_j(y)$, can be derived by the law of iterated expectations as follows:

$$C_j(y) = \int_{\Omega_A} \mathbb{E}[t|A]\gamma(A)dA, \tag{10.1}$$

where $\Omega_A = [1, +\infty[$ is the set of admissible values of A and $\mathbb{E}[t|A]$ is the conditional expected installing cost. If the installing cost is a random variable with uniform distribution in the interval $[c_j(y), Ac_j(y)]$, the conditional expected cost is given by

$$\mathbb{E}[t|A] = \frac{c_j(y) + Ac_j(y)}{2}. \tag{10.2}$$

To model the uncertainty on the parameter A, a shifted exponential density function is considered as follows:

$$\gamma(A) = \begin{cases} \lambda e^{-\lambda(A-1)}, & \text{if } A \geq 1, \\ 0, & \text{if } A < 1. \end{cases} \tag{10.3}$$

for a real positive parameter λ. Then, the expected installing cost is

$$C_j(y) = c_j(y) + \frac{c_j(y)}{2\lambda}.$$

As usual for exponential random variables, the expected cost C_j is a decreasing function of the parameter λ, given that $\partial C_j/\partial\lambda < 0$ for each $\lambda > 0$. The expected worth for each coalition is

$$v(S) = \sum_{i \in S} r_i - C(S),$$

where $C(S)$ is given by

$$C(S) = \sum_{j \in E} C_j\Big(\sum_{\substack{i:i \in S \\ p_i \in Q_j}} h_i\Big),$$

so that we guarantee core solutions for any choice of concave cost functions c_j.

Example 4 The network design situation $(N, G, h, OD, r, EIC)^{\phi}$ of the previous examples can be reconsidered assuming that the upper bound of cost function is unknown, i.e., taking $\bar{c}_j(x) = A\sqrt{x}$, were A is a random variable with a shifted exponential distribution. So, given $c_j(x) = \sqrt{x}$, the expected installing cost for each edge j is

$$C_j(x) = \sqrt{x} + \frac{\sqrt{x}}{2\lambda}.$$

The expected characteristic function is

$$v(\{1\}) = 5 - 2\left(\frac{1}{\lambda} + 1\right), \quad v(\{2\}) = 4 - 2\left(\frac{1}{\lambda} + 1\right)$$

$$v(\{1,2\}) = 9 - \left(\sqrt{2} + 2\right)\left(\frac{1}{\lambda} + 1\right).$$

The core of this game is a function of the parameter λ and is given by the system

$$\begin{aligned} \frac{3\lambda - 2}{\lambda} \leq x_1 \leq \frac{-\sqrt{2}\lambda + 5\lambda - \sqrt{2}}{\lambda}, \\ x_2 = \frac{-\sqrt{2}\lambda + 7\lambda + \lambda(-x_1) - \sqrt{2} - 2}{\lambda}. \end{aligned}$$

For $\lambda = 2$ we have the same solutions as in Example 2, for $\lambda \to +\infty$ we have the core of the lower game $\mathscr{C}(v^{opt})$ and for $\lambda = 1$ we have the core of the upper game $\mathscr{C}(v^{pes})$.

10.4 Conclusions

In this chapter, a multi-commodity network flow problem has been analyzed when some degrees of uncertainty affect the cost to realize it. Under some assumptions on the network itself (i.e., it has no cycles) and on the density functions that describe the randomness of costs, two cases were considered: a first one in which the costs lie within a real interval, and a second case in which the upper bound of the interval is a random variable itself. There is a clear link between this model and the literature that faces costs sharing problem with interval cooperative games. In this sense, the contribution of the chapter is to propose an approach that solve the ambiguity on preferences given by the interval core solutions.

The first limitations of the model is given by the assumption of acyclic network, that is, there is only one way to connect to nodes of the network. The reason beyond this choice lies in the fact that in case of acyclic network, convex costs function are sufficient condition to have a non-empty core.

The model with uncertainty could be deeply extended to the interesting case of networks with cycles, namely when at least a player has the possibility to use different paths to ship his commodity. In that case a minimum cost network can be defined as follows: given a network design situation (N, G, h, OD, r, IC), for any player $i \in N$ consider the set

$$P_i = \{\text{path connecting } o_i \text{ and } d_i\}$$

and define the cost of a coalition as

$$c(S) = \min_{p_i : p_i \in P_i \ \forall i \in S} \sum_{j \in E} c_j \Big(\sum_{\substack{i : i \in S \\ p_i \in Q_j}} h_i \Big)$$

and then

$$v(S) = \sum_{i \in S} r_i - c(S).$$

Unfortunately, the core of the cooperative game $< N, v >$ can be empty. Here:
(i) the profit sharing problem requires solution concepts different from the core solutions;
(ii) the minimum cost network has to be refined in cases where there exist many minimum cost networks, as in Example 1;

(iii) uncertainty can be considered also in the choice of the minimum cost network, besides on returns and/or on costs.

We address these considerations to future research.

Acknowledgements The work has been supported by STAR 2014 (linea 1) "Variational Analysis and Equilibrium Models in Physical and Social Economic Phenomena", University of Naples Federico II, Italy.

References

1. Alparslan Gök, S.Z.: On the interval Shapley value. Optimization **63**, 747–755 (2014)
2. Alparslan Gök, S.Z., Miquel, S., Tijs, S.: Cooperation under interval uncertainty. Math. Methods Oper. Res. **69**, 99–109 (2009)
3. Alparslan Gök, S.Z., Branzei, R., Tijs, S.: The interval Shapley value: an axiomatization. Cent. Eur. J. Oper. Res. **18**, 131–140 (2010)
4. Avrachenkov, K., Elias, J., Martignon, F., Neglia, Petrosyan, G.L.: A Nash bargaining solution for cooperative network formation games. In: Proceedings of Networking 2011, Valencia, Spain (2011)
5. Branzei, R., Dimitrov, D., Tijs, S.: Models in Cooperative Game Theory, vol. 556. Springer (2003)
6. Branzei, R., Branzei, O., Alparslan Gök, S.Z., Tijs, S.: Cooperative interval games: a survey. Cent. Eur. J. Oper. Res. **18**, 397–411 (2010)
7. Chen, H., Roughgarden, T., Valiant, G.: Designing networks with good equilibria. In: SODA '08/SICOMP '10 (2008)
8. D'Amato, E., Daniele, E., Mallozzi, L.: A network design model under uncertainty. In: Pardalos, P.M., Rassias, T.M. (eds.) Contributions in Mathematics and Engineering, In Honor of Constantin Caratheodory, pp. 81–93. Springer (2016)
9. Faigle, U., Nawijn, W.M.: Note on scheduling intervals on-line. Discret. Appl. Math. **58**, 13–17 (1995)
10. Gilles, R.P., Chakrabarti, S., Sarangi, S.: Nash equilibria of network formation games under consent. Math. Soc. Sci. **64**, 159–165 (2012)
11. Liu, X., Zhang, M., Zang, Z.: On interval assignment games. In: Zang, D. (ed.) Advances in Control and Communication, LNEE, vo. 137, pp. 611–616 (2012)
12. Mallozzi, L.: An application of optimization theory to the study of equilibria for games: a survey. Cent. Eur. J. Oper. Res. **21**, 523–539 (2013)
13. Mallozzi, L., Scalzo, V., Tijs, S.: Fuzzy interval cooperative games. Fuzzy Sets Syst. **165**, 98–105 (2011)
14. Mares, M., Vlach, M.: Fuzzy classes of cooperative games with transferable utility. Scientiae Mathematicae Japonica **2**, 269–278 (2004)
15. Marinakis, Y., Migdalas, A., Pardalos, P.M.: Expanding neighborhood search GRASP for the probabilistic traveling salesman problem. Optim. Lett. **2**, 351–361 (2008)
16. Monderer, D., Shapley, L.S.: Potential games. Games Econ. Behav. **14** 124–143 (1996)
17. Moulin, H.: Cost sharing in networks: some open questions. Int. Game Theory Rev. **15**, 134–144 (2013)
18. Moulin, H., Shenker, S.: Serial cost sharing. Econometrica **60**, 1009–1037 (1992)
19. Nisan, N., Roughgarden, T., Tardos, E., Vazirani, V.V.: Algorithmic Game Theory. Cambridge University Press, New York (2007)
20. Owen, G.: Game Theory. Academic Press, UK (1995)
21. Ozen, U., Slikker, M., Norde, H.: A general framework for cooperation under uncertainty. Oper. Res. Lett. **37**, 148–154 (2017)

22. Sharkey, W.W.: Network models in economics. In: Bali, M.O. et al., (eds.) Handbooks in OR & MS, vol. 8 (1995)
23. Tijs, S.: Introduction to Game Theory. Hindustan Book Agency (2003)
24. Topkis, D.: Supermodularity and Complementarity. Princeton University Press, Princeton (1998)
25. Trudeau, C., Vidal-Puga, J.: On the set of extreme core allocations for minimal cost spanning tree problems. J. Econ. Theory **169**, 425–452 (2017)

Chapter 11
Pricing Competition Between Cell Phone Carriers in a Growing Market of Customers

Andrey Garnaev and Wade Trappe

11.1 Introduction

Pricing is a core problem faced by communication markets. There is an extensive literature treating different aspects of the pricing problem. As a quick sampling, revenue sharing and pricing strategies for Internet Service Providers were studied by [16, 23]. Pricing was investigated for local and global WiFi markets by [6], under uncertainty related to the demand posed by users by [2], while pricing for video streaming in mobile networks was modeled by [19], and for uplink power in wideband cognitive radio networks by [1]. The difference between flat rate pricing and power-based pricing was studied by [11], while license virtual mobile network operators were investigated by [5] and competition between telecommunication service providers was modeled by [20].

In the United States, there are four major nationwide cellular carriers that cover the entire United States, and three smaller regional carriers (see, [13]). Choosing a cellular carrier is a tough problem for customers and, though there certainly is some aspect of non-rationality in the decision making, most customers nonetheless make their decision by comparing plan styles, prices, coverage, phone selection, speed, customer service quality and the future outlook for the provider (see, [13, 18]). A sophisticated customer even might adapt its selection of a carrier using the integrated analytical process and grey relational analysis algorithm suggested for network selection by [22].

In this paper, rather than exploring how customers choose carriers, we explore a complementary problem in which we consider all the customers as a market, which can be shared between the carriers based on an integrated characteristic incorporating

A. Garnaev (✉) · W. Trappe
WINLAB, Rutgers University, North Brunswick, USA
e-mail: garnaev@yahoo.com

W. Trappe
e-mail: trappe@winlab.rutgers.edu

S. K. Neogy et al. (eds.), *Mathematical Programming and Game Theory*,
Indian Statistical Institute Series, https://doi.org/10.1007/978-981-13-3059-9_11

its QoS and (service) prices. By the concept of service price, in this paper, we consider an abstracted, aggregate value incorporating such characteristics as plans and associated prices. Under QoS we consider an integrated characteristic incorporating such issues like coverage, data speed and customer services.

Trying to attract customers by better pricing, each of the carriers meet a dilemma to solve: on the one hand, by reducing prices the carrier can attract new customers, while on the other hand it yields a smaller profit from each customer. To deal with this dilemma a simple dynamic game-theoretical model associated with sharing a market of customers between carriers is presented in this paper, which allows one to find how the equilibrium pricing strategy depends on the customers's loyalty and the overall growth of the entire market of customers.

The organization of this paper is as follows: in Sect. 11.2, a game-theoretical model describing the competition between the carriers for the customers is given, as well as the existence and uniqueness of its solution. In Sect. 11.3, this solution is illustrated numerically in one-step- and multi-step scenarios. Finally, in Sect. 11.4, conclusions are provided related to the game.

11.2 Customers' Market Sharing Game

In this Section, we consider a non-zero sum game associated with the sharing of a growing market of customers between N carriers. At the beginning, let M_{0i} be the number of the customers signed up to the carrier i. Thus, the total number of the customers is $\sum_{i=1}^{N} M_{0i}$. It is expected that the market will be increased by M new customers. Depending on the price assigned by a carrier i, some of its customers could make a decision of whether to prolong their contract with the carrier or to look for a better option in the market with another carrier. Intuitively, higher prices lead to more customers leaving the carrier. We assume that

$$D_i = M_{0i} - a_i p_i \tag{11.1}$$

customers are inclined to keep the same carrier i, where p_i is the price assigned by the carrier i, and a_i is a sensitivity coefficient associated with likelihood of leaving given the price and the QoS the provider supplies the customer. D_i might be interpreted as the demand of loyal customers. We note that demand functions have found a wide applications in different economic models (see, [9]). So, if $p_i = 0$ then any customers are inclined to keep the same carrier i, but it does not bring any profit for the carrier. If $p_i = M_{0i}/a_i$, then all the customers lose loyalty to the carrier. Then, $a_i p_i$ is the number of the customers who are going to be disloyal to the carrier based on suggested price, and who are going to look for a better option in the market. Also, we assume that there is a random factor which finally could lead to some of the $M_{0i} - a_i p_i$ customers originally inclined to keep the same carrier i to ultimately change their mind. Let q_i be the probability that a customer belonging to carrier i, even in spite of there being a proper price, is going to go on the market. So, q_i can be

interpreted as a probability of disloyalty to the carrier, and $1 - q_i$ is the probability of loyalty. Thus, the carriers i expect

(a) to serve $(1 - q_i)(M_{0i} - a_i p_i)$ loyal customers,
(b) to compete for its share on the market consisting of $\overline{M}$ customers where

$$\overline{M} = M + \sum_{j=1}^{N} a_j p_j + \sum_{j=1}^{N} q_j (M_{0j} - a_j p_j).$$

We assume that the carriers share the customers' market according to the ratio form contest success function. Such function is commonly used for modelling share holders's attraction (see, [7]) or share goodwill levels (see, [8]), or even protection's level depending on applied efforts (see, [10, 14]). Namely, we assume that the carriers share the customers' market proportional to their contribution into the total demand of loyal customers $\sum_{j=1}^{N}(1 - q_j)D_j$. Thus, the carrier i gets the following share of the customers being on the market:

$$\xi_i(p_i, \boldsymbol{p}_{-i}) = \overline{M} \frac{(1 - q_i)D_i}{\sum_{j=1}^{N}(1 - q_j)D_j}, \tag{11.2}$$

where $\boldsymbol{p}_{-i} = (p_1, \ldots, p_{i-1}, p_{i+1}, \ldots, p_N)$ is the profile of strategies for all of the carriers excluding the carrier i.

The payoff π_i to the carrier i is the expected total profit gained from the loyal customers, and the ones he can attract from the market through competition with other carriers. Thus, the payoff is given as follows:

$$\begin{aligned}
\pi_i(p_i, \boldsymbol{p}_{-i}) &= (1 - q_i)(M_{0i} - a_i p_i)p_i + \xi_i(p_i, \boldsymbol{p}_{-i})p_i \\
&= (1 - q_i)(M_{0i} - a_i p_i)p_i \\
&\quad + \frac{\left(M + \sum_{j=1}^{N} a_j p_j + \sum_{j=1}^{N} q_j (M_{0j} - a_j p_j)\right)(1 - q_i)(M_{0i} - a_i p_i)}{\sum_{j=1}^{N}(1 - q_j)(M_{0j} - a_j p_j)} p_i,
\end{aligned} \tag{11.3}$$

with $p_i \in [0, M_{0i}/a_i]$ for $i = 1, \ldots, N$. Thus, $[0, M_{0i}/a_i]$ is the set of the all feasible strategies for carriers i.

We assume that the carriers have complete knowledge about the market's parameters, i.e. about the number of customers M, M_{0i}, probabilities of disloyalty q_i, and coefficients of sensitivity a_i.

Each carrier wants to maximize its profit, i.e. we are looking for a Nash equilibrium (see, [9]). Recall that $\boldsymbol{p}_*$ is a Nash equilibrium if and only if for each $\boldsymbol{p}$ the following inequalities hold:

$$\pi_i(p_i, \boldsymbol{p}_{-i*}) \leq \pi_i(p_{i*}, \boldsymbol{p}_{-i*}) \text{ for } i = 1, \ldots, N. \tag{11.4}$$

The best response strategy for the carrier i to a fixed strategy profile $\boldsymbol{p}_{-i}$ for the other carriers is

$$p_i = \mathrm{BR}_i(\boldsymbol{p}_{-i}) = \arg_{p_i} \max \pi_i(p_i, p_{-i}). \tag{11.5}$$

Then, $\boldsymbol{p}$ is an equilibrium if and only if it is a solution of the best response Eq. (11.5) with $i = 1, \ldots, N$.

Theorem 1 *The best response strategy BR_i for $i = 1, \ldots, N$ can be obtained in closed form as follows:*

$$p_i = BR_i(\boldsymbol{p}_{-i}) = \frac{(1-q_i)M_{0i} + s_i - \sqrt{((1-q_i)M_{0i} + s_i)s_i}}{a_i(1-q_i)}, \tag{11.6}$$

where

$$s_i = \sum_{j=1, j\neq i}^{N} (1-q_j)(M_{0j} - a_j p_j). \tag{11.7}$$

Proof First note that

$$\begin{aligned} \frac{\partial \pi_i}{\partial p_i} &= \frac{(1-q_i)(s_i + M_{0i} + M + f_i)}{((1-q_i)(M_{0i} - a_i p_i) + s_i)^2} \\ &\times \left(a_i^2(1-q_i)p_i^2 - 2a_i((1-q_i)M_{0i} + s_i)p_i + M_{0i}(M_{0i} + s_i)\right) \end{aligned} \tag{11.8}$$

and

$$\begin{aligned} \frac{\partial^2 \pi_i}{\partial p_i^2} &= -\frac{2a_i(1-q_i)((1-q_i)M_{0i} + s_i)}{((1-q_i)(M_{0i} - a_i p_i) + s_i)^3} \\ &\times (s_i + M_{0i} + M + f_i) < 0 \end{aligned} \tag{11.9}$$

with

$$f_i = \sum_{j=1, j\neq i}^{N} (a_j p_j + (1-q_j)(M_{0j} - a_j p_j)). \tag{11.10}$$

Thus, π_i is concave and the best response strategy can be obtained as the unique root in $[0, M_{0i}/a_i]$ of the quadratic equation:

$$a_i^2(1-q_i)p_i^2 - 2a_i((1-q_i)M_{0i} + s_i)p_i + M_{0i}(M_{0i} + s_i) = 0. \tag{11.11}$$

This implies (11.6), and the result follows.

Here we can observe a quite interesting phenomena that the best response strategies do not depend explicitly on the number of new customers M coming to the market. On one hand it is surprising, since an important parameter appears to not

be taken into account. On the other hand, it is quite natural, since making a better proposal in competing for re-sharing of users already existent in the market, the carriers understand that this re-sharing will impact the choices of the new customers, since they also try to choose better proposals. Thus, in a short-run price planning, the number of new customers coming on the market does not have an impact on price. Meanwhile, in terms of a long-run price planning, this number produces an impact on the price since after coming to the market, the customers also will sign up, and so they join the updated structure of the market of customers, which impacts pricing.

Theorem 2 *The considered game has an unique equilibrium* $\boldsymbol{p}$ *given as follows:*

$$p_i = \frac{(1-q_i)M_{0i} + x_i - \sqrt{((1-q_i)M_{0i} + x_i)x_i}}{a_i(1-q_i)} \text{ for } i = 1, \ldots, N, \tag{11.12}$$

where

$$x_i = \frac{-(1-q_i)M_{0i} + \sqrt{(1-q_i)^2 M_{0i}^2 + 4x^2}}{2}, \tag{11.13}$$

and x *is the unique positive root of the equation*

$$F(x) = 0 \tag{11.14}$$

with

$$F(x) := 2(N-1)x + \sum_{i=1}^{N}(1-q_i)M_{0i} - \sum_{i=1}^{N}\sqrt{(1-q_i)^2 M_{0i}^2 + 4x^2}. \tag{11.15}$$

Proof Due to (11.9), π_i is concave on p_i. Thus, an equilibrium exists by Nash Theorem [9].

Finding all of the equilibria is equivalent to finding all of the solutions of the best response Eq. (11.6), which are equivalent to

$$(1-q_i)(M_{0i} - a_i p_i) + s_i = \sqrt{((1-q_i)M_{0i} + s_i)s_i}. \tag{11.16}$$

Let us introduce an auxiliary notation

$$x = \sum_{j=1}^{N}(1-q_j)(M_{0j} - a_j p_j). \tag{11.17}$$

Then, by (11.7),

$$s_i = x - (1-q_i)(M_{0i} - a_i p_i). \tag{11.18}$$

Substituting this s_i into the left side of Eq. (11.16) implies

$$x = \sqrt{((1-q_i)M_{0i} + s_i)s_i}. \tag{11.19}$$

Solving this equation on positive s_i implies

$$s_i = \frac{-(1-q_i)M_{0i} + \sqrt{(1-q_i)^2 M_{0i}^2 + 4x^2}}{2}. \tag{11.20}$$

On one hand, summing up this equation by $i = 1, \ldots, N$ implies

$$\sum_{i=1}^{N} s_i = \sum_{i=1}^{N} \frac{-(1-q_i)M_{0i} + \sqrt{(1-q_i)^2 M_{0i}^2 + 4x^2}}{2}. \tag{11.21}$$

On the other hand, by (11.7),

$$\begin{aligned} \sum_{i=1}^{N} s_i &= \sum_{i=1}^{N} \sum_{j=1, j\neq i}^{N} (1-q_j)(M_{0j} - a_j p_j) \\ &= (N-1)x. \end{aligned} \tag{11.22}$$

Then, (11.21) and (11.22) imply that

$$(N-1)x = \sum_{i=1}^{N} \frac{-(1-q_i)M_{0i} + \sqrt{(1-q_i)^2 M_{0i}^2 + 4x^2}}{2}. \tag{11.23}$$

Thus, x has to be a positive root of the Eq. (11.14) F given by (11.15).

Note that for F given by (11.15) the following relations hold:

$$\frac{d^2 F}{d x^2} = -\sum_{i=1}^{N} \frac{4((1-q_i)M_{0i})^2}{(((1-q_i)M_{0i})^2 + 4x^2)^{3/2}} < 0 \tag{11.24}$$

and

$$\frac{d F}{d x}(0) = 2(N-1) > 0. \tag{11.25}$$

So, F is a continuous concave function, increasing at $x = 0$ such that $F(0) = 0$ and $\lim_{t\uparrow\infty} F(t) = -\infty$. Thus,(11.12) has a unique positive root. This allows us to obtain uniquely s_i by (11.20), as well as the strategy p_i by (11.16), and the result follows.

In particular, (11.12) yields that an increasing sensitivity coefficient a_i implies a decreasing equilibrium price. Also, it is interesting to observe a similarity between these equilibrium strategies and water-filling strategies [3, 12, 17]. Namely, both these strategies are given in closed form defined by a parameter which can be found as the unique solution of an auxiliary equation.

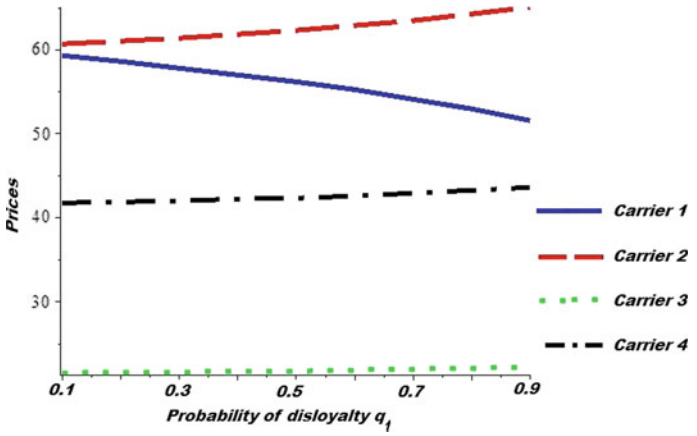


Fig. 11.1 Prices as function on probability of disloyalty q_1

11.3 Numerical Illustrations

As a numerical illustration, we consider a market consisting of four carriers, i.e. $N = 4$. Let the number of new, incoming customers be $M = 10{,}00{,}000$, the number of customers assigned to the carriers be given by $\boldsymbol{M}_0 = (20{,}000, 30{,}000, 10{,}000, 15{,}000)$, the sensitivity coefficients be given by $\boldsymbol{a} = (200, 300, 250, 200)$, the probabilities of the customer's disloyalty is given by $\boldsymbol{q} = (q_1, 0.3, 0.3, 0.3)$ while q_1 varying from 0.1 to 0.9 for carrier 1. Increasing this probability makes carrier 1 reduce its expected predictable income by reducing its share of the loyal customers, to compensate this loss, the carriers have to pay more attention to the market reducing its price (Fig. 11.1). In any case, this increase in probability leads to a provider reducing its share of the market (Fig. 11.2) and its payoff (Fig. 11.3). The other carriers gain from such increasing the market in increasing their shares and the payoffs. It also leads to a slight increase in their prices, which can be explained as a necessity to serve more customers, which, of course, is not free.

Another important issue that the customer's disloyalty could impact is the relative share of the market. To illustrate this, we consider the game played repeatedly over time slots $t = 0, 1, \ldots$ with a market that is growing at a rate of α percent per time slot. Let us describe the scenario in detail. Suppose, at the beginning of time slot t there are M_{0i}^t customers shared by the carriers. Then, the demand for the loyal customers is given by $D_i^t = M_{0i}^t - a_i p_i^t$ where p_i^t is the price assigned by the carrier i at time slot t. Due to the fact that the market grows at a fixed rate α, the number of (new) incoming customers is $M^t = \alpha \sum_{i=1}^{N} M_{0i}^t$. Thus, at time slot t, the carrier expects to compete in a market consisting of $\overline{M}^t$ customers, where

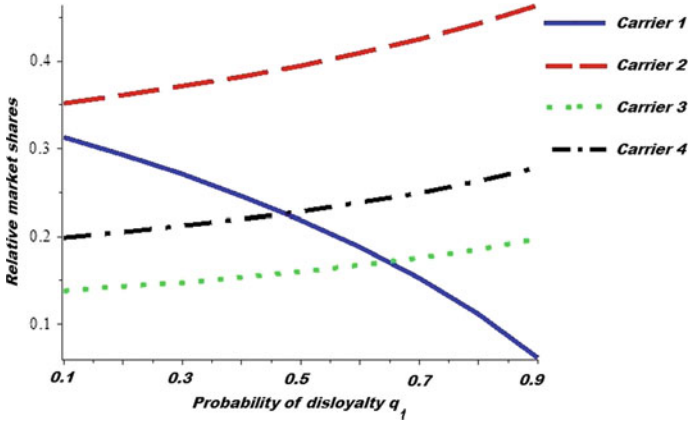


Fig. 11.2 Relative shares of the market as function on probability of disloyalty q_1

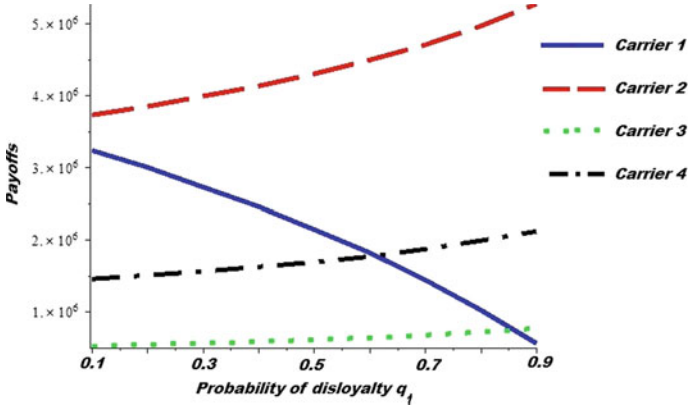


Fig. 11.3 Payoffs to the carriers as function on probability of disloyalty q_1

$$\begin{aligned}\overline{M}^t &= M^t + \sum_{j=1}^{N} a_j p_j^t + \sum_{j=1}^{N} q_j (M_{0j}^t - a_j p_j^t) \\ &= \alpha \sum_{i=1}^{N} M_{0i}^t + \sum_{j=1}^{N} a_j p_j^t + \sum_{j=1}^{N} q_j (M_{0j}^t - a_j p_j^t).\end{aligned} \tag{11.26}$$

This allows one to define payoffs for the carriers at time slot t by (11.3) with $M = M^t$, $M_{0i} = M_{0i}^t$ and $\boldsymbol{p} = \boldsymbol{p}^t$. For time slot t, we may find the unique Nash equilibrium $\boldsymbol{p}^t$ of this game. Then, (11.2) with $\overline{M} = \overline{M}^t$ and $D_i = D_i^t$ returns the shares of the customers obtained by the carriers at the end of time slot t. These shares serve as the beginning customer shares for the carriers (i.e. M_{0i}^{t+1}) at the beginning of the next time slot $t + 1$, and so on.

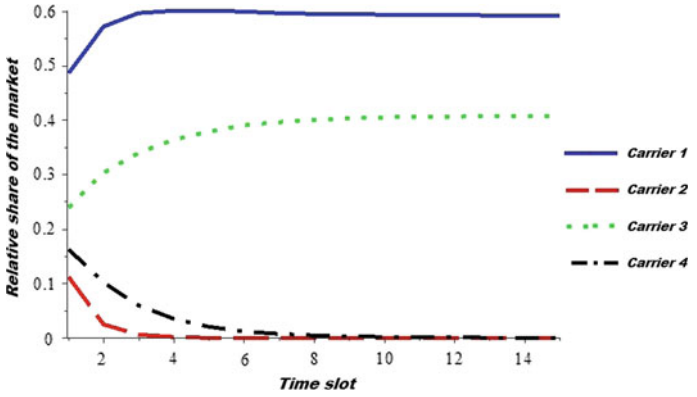


Fig. 11.4 Relative shares of the market by time slots for $\boldsymbol{q} = (0.1, 0.9, 0.3, 0.7)$

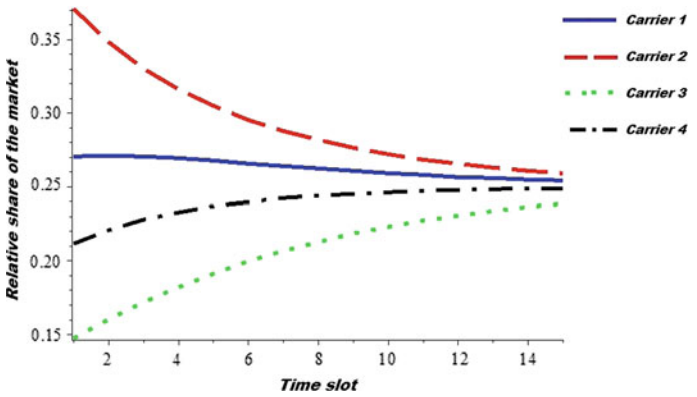


Fig. 11.5 Relative shares of the market by time slots for $\boldsymbol{q} = (0.9, 0.9, 0.9, 0.9)$

As a numerical illustration we consider $\alpha = 0.1$ and $\boldsymbol{q} = (0.1, 0.9, 0.3, 0.7)$ and $\boldsymbol{q} = (0.9, 0.9, 0.9, 0.9)$. Figures 11.4 and 11.5 illustrate the stabilization of the relative market shares associated with the carriers across time. In the case where there is significant switching tendency for the customers, the share is more fair compared with the situation when some of the carriers (1 and 3) has a large percent of loyal customers relative to disloyal customers.

11.4 Conclusions

In this paper a game-theoretical model for the competition between service providers, such as cell-phone carriers, in a market of customers that is growing was investigated. Solving this game allowed us to show how the loyalty factor associated with the

carriers might impact to the prices and relative market share between the carriers. Namely, higher loyalty leads to higher prices and obtaining a larger share of the market. Consequently, when considering regulatory mechanisms that can support price stability for consumers and a fair sharing of the customer market, it is desirable that regulatory agencies develop rules that simplify and encourage the ability for customers to be able to switch their carriers. It is important to note that for a growing market, we can observe numerically the stabilization of the relative shares of the market across the time slots for repeatedly played game scenarios, but this observation is one that we cannot prove analytically. One of the goals of our future research is to develop mathematical techniques to prove such stabilization in repeatedly played games. Another important issue about the model is that the carriers have complete knowledge about all of the parameters. Here problems arise (a) to estimate the demand functions and involved parameters, and verify model with reality, and (b) whether or not private information be beneficial for the carriers. To deal with first problem, a special branch of economics theory, *econometrics*, was developed which involves the application of statistical and mathematical theories in economics to test hypotheses, and then compare and contrast the results against real-life examples. As examples related to this, estimating characteristics such as the demand function and market power, we refer the readers to [4, 21] correspondingly. To deal with the second problem, a *Bayesian* approach has to be applied. As examples of such approach we refer the readers to textbook [15]. The goal of our future work is to investigate the second problem, namely, how the carriers could benefit from private information.

References

1. AlDaoud, A., Alpcan, T., Agarwal, S., Alanyali, M.: A Stackelberg game for pricing uplink power in wide-band cognitive radio networks. In: 47th IEEE Conference on Decision and Control (CDC), pp. 1422–1427 (2008)
2. Altman, E., Avrachenkov, K., Garnaev, A.: Taxation for green communication. In: 8th International Symposium on Modeling and Optimization in Mobile, Ad Hoc and Wireless Networks (WiOpt), pp. 108–112 (2010)
3. Altman, E., Avrachenkov, K., Garnaev, A.: Closed form solutions for water-filling problems in optimization and game frameworks. Telecommun. Syst. **47**, 153–164 (2011)
4. Berry, S., Levinsohn, J., Pakes, A.: Differentiated products demand systems from a combination of micro and macro data: the new car market. J. Polit. Econ. **112**, 68–105 (2004)
5. Dewenter, R., Haucap, J.: Incentives to licence virtual mobile network operators (MVNOs). In: Dewenter, R., Haucap, J. (eds.) Access Pricing: Theory and Practice, pp. 303–323. Elsevier BV, Amsterdam (2006)
6. Duan, L., Huang, J., Shou, B.: Optimal pricing for local and global with markets. In: IEEE Conference on Computer Communications (INFOCOM), pp. 1088–1096 (2013)
7. Federgruen, A., Yang, N.: Competition under generalized attraction models: applications to quality competition under yield uncertainty. Manag. Sci. **55**, 2028–2043 (2009)
8. Fershtman, C., Mahajan, V., Muller, E.: Market share pioneering advantage: a theoretical approach. Manag. Sci. **36**, 900–918 (1990)
9. Fudenberg, D., Tirole, J.: Game Theory. MIT Press, Cambridge (1991)
10. Garnaev, A., Baykal-Gursoy, M., Poor, H.V.: Security games with unknown adversarial strategies. IEEE Trans. Cybern. **46**, 2291–2299 (2016)

11. Garnaev, A., Hayel, Y., Altman, E.: Multilevel pricing schemes in a deregulated wireless network market. In: 7th International Conference on Performance Evaluation Methodologies and Tools (Valuetools), pp. 126–135 (2013)
12. Garnaev, A., Trappe, W.: Bargaining over the Fair Trade-off Between Secrecy and Throughput in OFDM Communications. IEEE Trans. Inf. Forensics Secur. **12**, 242–251 (2017)
13. German, K.: Quick guide to cell phone carriers. CNET 27 May 2014. http://www.cnet.com/news/quick-guide-to-cell-phone-carriers/
14. Guan, P., Zhuang, J.: Modeling resources allocation in attacker-defender games with "warm up" CSF. Risk Anal. **36**, 776–791 (2016)
15. Han, Z., Niyato, D., Saad, W., Basar, T., Hjrungnes, A.: Game Theory in Wireless and Communication Networks: Theory, Models, and Applications. Cambridge University Press, Cambridge (2011)
16. He, L., Walrand, J.: Pricing and revenue sharing strategies for internet service provider. IEEE J. Sel. Areas Commun. **24**, 942–951 (2006)
17. He, P., Zhao, L., Zhou, S., Niu, Z.: Waterlling: a geometric approach and its application to solve generalized radio resource allocation problems. IEEE Trans. Wirel. Commun. **12**, 3637–3647 (2013)
18. Lendono, J.: How to choose a cell phone carrier. PC 29 Aug 2011. http://www.pcmag.com/article2/0%2c2817%2c2368279%2c00.asp
19. Lin, W., Liu, K.: Game-theoretic pricing for video streaming in mobile networks. IEEE Trans. Image Process. **21**, 2667–2680 (2012)
20. Maille, P., Tuffin, B., Vigne, J.: Economics of technological games among telecommunication service providers. J. Commun. Netw. **21**, 65–82 (2011)
21. Perloff, J.M., Karp, L.S., Golan, A.: Estimating Market Power and Strategies. Cambridge University Press, Cambridge (2007)
22. Song, Q., Jamalipour, A.: Network selection in an integrated wireless LAN and UMTS environment using mathematical modeling and computing techniques. IEEE Wirel. Commun. **12**, 42–48 (2005)
23. Wu, Y., Kim, H., Hande, P., Chiang, M., Tsang, D.: Revenue sharing among ISPs in two-sided markets. In: IEEE Conference on Computer Communications (INFOCOM), pp. 596–600 (2011)

Chapter 12
Stochastic Games with Endogenous Transitions

Reinoud Joosten and Robin Meijboom

12.1 Introduction

We present and subsequently analyze a stochastic game in which transition probabilities at any point in time depend on the history of the play, i.e., players' past action choices, their current choices, and the current state. This development has been inspired by an ambition to incorporate certain empirical phenomena into *Small Fish Wars*[1] [37]. Here, agents possess the fishing rights on a body of water, and the resource can be in either of two states, *High* or *Low*. In the former, the fish are more abundant and therefore catches are larger than in the latter. The agents have two options, to fish with or without restraint. Fishing with restraint by both agents is (assumed to be) sustainable in the long run, as the resource will be (assumed to be) able to recover; unrestrained fishing by both yields higher immediate catches, but damages the resource significantly if continued for prolonged periods of time. This damage becomes apparent in the dynamics of the system as an increase in the probabilities that the system moves from *High* to *Low*, and simultaneously a decrease in the probabilities of the system to move from *Low* to *High*. This causes the system and hence the play, to spend a higher proportion of time in *Low*.

We additionally aim to incorporate hysteresis effects called poaching pits in the field of management of replenishable resources (e.g., Bulte [11], Courchamp et al. [13], Hall et al. [24]). Hysteresis may be caused by biological phenomena induced by the (nature of the) exploitation of the resource. For instance, full-grown cod

[1] A word play on Levhari and Mirman [50] who show that strategic interaction in a fishery may induce a "tragedy of the commons" [27].

I thank J. Flesch, F. Thuijsman, E. Solan and A. Laruelle for advice. Audiences in Enschede, Tilburg, Maastricht, Tel Aviv, Bilbao and Istanbul are also thanked for feedback. Last but by no means least, I thank the referee for extremely careful reading and for excellent suggestions for improvement.

R. Joosten (✉) · R. Meijboom
IEBIS, BMS, University of Twente, POB 217,
7500 AE Enschede, The Netherlands
e-mail: r.a.m.g.joosten@utwente.nl

S. K. Neogy et al. (eds.), *Mathematical Programming and Game Theory*,
Indian Statistical Institute Series, https://doi.org/10.1007/978-981-13-3059-9_12

spawn a considerably higher number of eggs than younger specimen: Oosthuizen and Daan [57], Armstrong et al. [3] find linear fecundity-weight relations, Rose et al. [63] report exponential fecundity-weight relations. As mature cod are targeted by modern catching techniques such as for instance gill netting, overfishing hurts mainly the cohorts most productive in providing offspring. To regain full reproductive capacity, younger cohorts must reach ages well beyond adulthood. Hence, it may take cod a long while to escape a poaching pit after a recovery plan or program to replenish the stock has been effectuated.

To achieve our goals we engineered a stochastic game[2] as follows. Nature (chance) may move the play from one state to the other dependent on the current action choices of the agents, but also on their past catching behavior. To achieve the above-formulated modeling aims we introduce *endogenously changing* stochastic variation,[3] the evolution of the transition probabilities reflects that the more frequently the agents exploit the resource without restraint, the more it deteriorates. Here, the probability of moving to *High* may decrease in time in each state and for each action combination if the agents show prolonged lack of restraint, i.e., overfish frequently.

Transition probabilities from *Low* to *High* may become zero, resulting in *Low* becoming a *temporarily* absorbing state. If the agents keep overexploiting the resource, this situation does not change in our model. Even if the agents revert to restraint in order to bring about the recovery of the resource, it may take a long time before *High* becomes accessible again. Thus, we endeavor to reproduce effects similar to the ones associated to hysteresis.

The agents are assumed to wish to maximize their long-term average catches. We adopt a Folk Theorem type analysis as in Joosten et al. [42], and validate relevant procedures in this new setting. First, we show how to establish the rewards for any pair of jointly convergent pure strategies. Then, we determine the set of jointly convergent pure-strategy rewards. A more complex issue is then to find for each player the threat point reward, i.e., the highest amount this player can guarantee himself if his opponent tries to minimize his rewards. Finally, we obtain a large set of rewards which can be supported by equilibria using threats, namely all jointly-convergent pure-strategy rewards giving each player more than the threat point reward.

In the model analyzed throughout the chapter for expository purposes, we gain insights relevant to the management of the resource. Our findings reveal a potential for compromise between ecological and economic maximalistic goals, thus overcoming the one-sidedness of management policies for natural resources as noted by e.g., Holden [33], Brooks et al. [10], and in turn improving their chances of success cf., e.g., BenDor et al. [5], Sanchirico et al. [64]. Full restraint, an ecological maximalistic goal, yields total rewards which are considerably higher than never-restraint rewards. Yet, a possible economic maximalistic goal, i.e., Pareto-efficient equilibrium rewards resulting from jointly convergent pure strategies with threats, yields a

[2] 'Engineered' as in Aumann [4]. Stochastic games were introduced by Shapley [69], see also Amir [1] for links to difference and differential games to which much work on fisheries belongs, cf., e.g., Haurie et al. [29], Long [51] for overviews.

[3] So, the Markov property of standard stochastic games [69] is lost.

sizeable increase of total rewards even over full restraint. We find that the proportion of time spent in such a poaching pit goes to zero in the long run under equilibrium behavior. A whole range of models should be analyzed to obtain general findings providing insights into the full range of fishery management games.

Next, we introduce our model with endogenous transition probabilities. In Sect. 12.3, we focus on strategies and restrictions desirable or resulting from the model. Section 12.4 treats rewards in a very general sense, and equilibrium rewards more specifically. Also some attention is paid to the complexity of computing threat point rewards. Section 12.5 concludes.

12.2 Endogenous Transition Probabilities

A *Small Fish War* is played by row player A and column player B at discrete moments in time called stages. Each player has two actions and at each stage $t \in \mathbb{N}$ the players independently and simultaneously choose an action. Action 1 for either player denotes the action for which some restriction exists allowing the resource to recover, e.g., catching with wide-mazed nets or catching a low quantity. Action 2 denotes the action with little restraint.

We assume catches to vary due to random shocks, which we model by means of a stochastic game with two states at every stage of the play. First, let us capture the past play until stage t, $t > 1$, by the following two matrices:

$$QH^t = \begin{bmatrix} q_1^t & q_2^t \\ q_3^t & q_4^t \end{bmatrix}, \text{ and } QL^t = \begin{bmatrix} q_5^t & q_6^t \\ q_7^t & q_8^t \end{bmatrix}.$$

Here, e.g., q_1^t is the relative frequency with which action pair top-left in *High* has occurred until stage t, and q_7^t is the relative frequency of action pair bottom-left in *Low* having occurred during past play. So, we must have $q^t = \left(q_1^t, \ldots, q_8^t\right) \in \Delta^7 = \{x \in \mathbb{R}^8 | x_i \geq 0 \text{ for all } i = 1, \ldots, 8 \text{ and } \sum_{j=1}^{8} x_j = 1\}$. We refer to such a vector as the **relative frequency vector**.

Let the interaction at stage t of the play be represented by the following:

$$H^t = H\left(q^t\right) = \begin{bmatrix} \theta_1, p_1\left(q^t\right) & \theta_2, p_2\left(q^t\right) \\ \theta_3, p_3\left(q^t\right) & \theta_4, p_4\left(q^t\right) \end{bmatrix},$$

$$L^t = L\left(q^t\right) = \begin{bmatrix} \theta_5, p_5\left(q^t\right) & \theta_6, p_6\left(q^t\right) \\ \theta_7, p_7\left(q^t\right) & \theta_8, p_8\left(q^t\right) \end{bmatrix}.$$

Here $H^t\left(q^t\right)$ ($L^t\left(q^t\right)$) indicates state *High* (*Low*) at stage t of the play if the play until then resulted in relative frequency vector q^t. Each entry of the two matrices has an ordered pair denoting the pair of payoffs to the players $\theta_i = \left(\theta_i^A, \theta_i^B\right)$ if the corresponding action pair is chosen and the probability $p_i\left(q^t\right)$ that the system moves

to *High* at stage $t+1$ (and to *Low* with the complementary probability). All functions $p_i : \Delta^7 \to [0,1]$ are assumed continuous. We now give an example.

Example 1 In this Small Fish War we assume that in both states Action 1, i.e., catching with restraint, is dominated by the alternative.[4] Let, for given relative frequency vector $q^t \in \Delta^7$, the transition functions $p_i : \Delta^7 \to [0,1]$, $i = 1, \ldots, 8$, governing the transition probabilities, be given by

$$\begin{aligned}
p_1(q^t) &= \left[\tfrac{8}{10} - \tfrac{11}{24}q_4^t - \tfrac{11}{12}q_8^t\right]_+ \\
p_2(q^t) &= p_3(q^t) = \left[\tfrac{6}{10} - \tfrac{11}{20}q_4^t - \tfrac{11}{10}q_8^t\right]_+ \\
p_4(q^t) &= \left[\tfrac{3}{10} - \tfrac{11}{16}q_4^t - \tfrac{11}{8}q_8^t\right]_+ \\
p_5(q^t) &= \left[\tfrac{6}{10} - \tfrac{11}{12}q_4^t - \tfrac{11}{6}q_8^t\right]_+ \\
p_6(q^t) &= p_7(q^t) = \left[\tfrac{4}{10} - \tfrac{11}{8}q_4^t - \tfrac{11}{4}q_8^t\right]_+ \\
p_8(q^t) &= \left[\tfrac{1}{10} - \tfrac{11}{4}q_4^t - \tfrac{11}{2}q_8^t\right]_+ .
\end{aligned}$$

Here, $[x]_+$ is short hand for $\max\{x, 0\}$. These equations capture the following deliberations. Two-sided full restraint is assumed to cause not more damage to the resource in both states than if exactly one player catches with restraint. Hence, the probability that during the next stage play is in *High* if the first case arises is at least equal to the corresponding probability in the second case. We also assume symmetry, hence $p_2\left(q^t\right) = p_3\left(q^t\right)$ and $p_6\left(q^t\right) = p_7\left(q^t\right)$. Furthermore, we assume that exactly one player catching without restraint is not more harmful to the resource than two players catching without restraint. The inequalities $p_i\left(q^t\right) \geq p_{i+4}\left(q^t\right)$ for $i = 1, \ldots, 4$, are assumed to hold because if the play is in *Low*, the system is assumed at least as more vulnerable to overfishing as in *High*. We refer to e.g., Kelly et al. [46] for an empirical underpinning of these modeling choices.

Now, we show that renewable resources may recuperate slowly after a program of recovery has been taken up. Suppose both agents play Action 1 twice followed by 2 for a sufficiently long period of time until stage t^*. Clearly, $q_4^{t^*} + q_8^{t^*} = \frac{t^*-2}{t^*}$. Now, for $t^* \to \infty$, $p_5\left(q^{t^*}\right) = p_6\left(q^{t^*}\right) = p_7\left(q^{t^*}\right) = p_8\left(q^{t^*}\right) = 0$, because

$$\begin{aligned}
&\tfrac{6}{10} - \tfrac{11}{12}q_4^{t^*} - \tfrac{11}{6}q_8^{t^*} = \tfrac{6}{10} - \tfrac{11}{12}\left(\tfrac{t^*-2}{t^*} - q_8^{t^*}\right) - \tfrac{11}{6}q_8^{t^*} = \\
&\tfrac{6}{10} - \tfrac{11}{12}\left(1 - \tfrac{2}{t^*} - q_8^{t^*}\right) - \tfrac{11}{6}q_8^{t^*} = -\tfrac{19}{60} + \tfrac{11}{6t^*} - \tfrac{11}{12}q_8^{t^*} < 0.
\end{aligned}$$

Then, $p_5\left(q^{t^*}\right) = 0$ and by the relation to the other transition probability functions, $p_6\left(q^{t^*}\right) = p_7\left(q^{t^*}\right) = p_8\left(q^{t^*}\right) = 0$ as well. Take $t^* = 16$, clearly

$$-\tfrac{19}{60} + \tfrac{11}{6t^*} - \tfrac{11}{12}q_8^{t^*} < -\tfrac{19}{60} + \tfrac{11}{6t^*} < 0.$$

If both agents switch to playing sequences of $(1, 1, 1, \ldots)$ from then on, it will take a while before $p_5\left(q^t\right)$ becomes positive again. Since

[4] Right now, we do not need the actual payoffs and focus on the transition probabilities.

$$\frac{6}{10} - \frac{11}{12} q_4^{t^*+k} - \frac{11}{6} q_8^{t^*+k} = \frac{6}{10} - \frac{11}{12} \frac{t^*-2}{t^*+k} - \frac{11}{12} q_8^{t^*+k} =$$
$$\frac{6}{10} - \frac{11}{12} \left(1 - \frac{k+2}{t^*+k}\right) - \frac{11}{12} q_8^{t^*+k} = -\frac{19}{60} + \frac{11}{12} \frac{k+2}{t^*+k} - \frac{11}{12} q_8^{t^*+k}$$
$$< -\frac{19}{60} + \frac{11}{12} \frac{k+2}{t^*+k},$$

the first expression cannot be positive for $k < \frac{19t^*-110}{36}$. So, for $t^* = 16$ it takes **at least** six stages for the play to be able to return to *High*.

12.3 Strategies and Restrictions

A strategy is a game plan for the entire infinite time horizon, allowing it to depend on any condition makes an extensive analysis of infinitely repeated games quite impossible. Most restrictions in the literature put requirements on what aspects the strategies are conditional upon. For instance, a *history-dependent* strategy prescribes a possibly mixed action to be played at each stage conditional on the current stage and state, as well as on the full history until then, i.e., all states visited and all action combinations realized before.

Less general strategies are for instance, *action independent* ones which condition on all states visited before, but not on the action combinations chosen [31]. Markov strategies condition on the current state and the current stage, and stationary strategies only condition on the present state (cf., e.g., Filar and Vrieze [20], Flesch [21]).

The challenge in the present framework is to find restrictions on strategies which are helpful in the analysis. Although Markov and stationary strategies have proven their value in the analysis of finite state stochastic games with fixed transition probabilities, it is quite unclear what their contribution can be in the present framework.

Essentially, (at least) two points of view can be adopted to analyze the present framework. The one we favor is the one in which *High* and *Low* are seen as the states with the transitions between these states being a function of the history of the play as captured by the relative frequency vector q^t. Stationary strategies are easily formulated here, but probably much too simple for analytical purposes as some link with q^t must be assumed to be useful. An alternative is to define the states according to the relative frequency vector in which there exist infinitely many states $H(q^t)$ and $L(q^t)$. Here, the practical problem is the enormity of the task of infinitely many stationary or Markov strategies to be defined.

Let $\mathcal{X}^k$ denote the set of history-dependent strategies of player $k = 1, 2$. A strategy is **pure**, if at *each* stage a **pure action** is chosen, i.e., an action is chosen with probability 1. The set of pure strategies for player k is $\mathcal{P}^k$, and $\mathcal{P} \equiv \mathcal{P}^A \times \mathcal{P}^B$. Let us define the following notions, introduced before in a rather informal manner, a bit more formally. For $j = 1, 2$, $t > 1$

$$
\begin{aligned}
q^t_j &\equiv \frac{\#\left\{\left(j_u^{A,H}, j_u^{B,H}\right) |\ j_u^{A,H}=1,\ j_u^{B,H}=j,\ 1\leq u<t\right\}}{t-1},\\
q^t_{j+2} &\equiv \frac{\#\left\{\left(j_u^{A,H}, j_u^{B,H}\right) |\ j_u^{A,H}=2,\ j_u^{B,H}=j,\ 1\leq u<t\right\}}{t-1},\\
q^t_{j+4} &\equiv \frac{\#\left\{\left(j_u^{A,L}, j_u^{B,L}\right) |\ j_u^{A,H}=1,\ j_u^{B,H}=j,\ 1\leq u<t\right\}}{t-1},\\
q^t_{j+6} &\equiv \frac{\#\left\{\left(j_u^{A,L}, j_u^{B,L}\right) |\ j_u^{A,H}=2,\ j_u^{B,H}=j,\ 1\leq u<t\right\}}{t-1}.
\end{aligned}
$$

Here, $j_u^{A,X}$ ($j_u^{B,X}$) denotes the action taken by player A (B) while being in state $X = H, L$ at stage u. So, for instance q_4^t is the relative frequency of action pair $(2, 2)$ in state H being chosen until stage t.

The strategy pair $(\pi, \sigma) \in \mathcal{X}^A \times \mathcal{X}^B$ is **jointly convergent** if and only if $q \in \Delta^7$ exists such that for all $\varepsilon > 0$, $i \in \{1, 2, \ldots, 8\}$:

$$\limsup_{t\to\infty} \Pr_{\pi,\sigma}\left[\left|q_i^t - q_i\right| \geq \varepsilon\right] = 0. \tag{12.1}$$

$\Pr_{\pi,\sigma}$ denotes the probability under strategy pair (π, σ). $\mathcal{JC}$ denotes the set of jointly convergent strategy pairs. Under such a pair of strategies, the relative frequency of each action pair in both states as play goes to infinity converges to a fixed number with probability 1 in the terminology of Billingsley [8, p. 274]). The **set of jointly-convergent pure-strategy rewards** $P^{\mathcal{JC}}$ is then the set of pairs of rewards obtained by using a pair of jointly-convergent pure strategies.

For a pair of jointly convergent pure strategies, let $p_i \equiv \lim_{t\to\infty} p_i\left(q^t\right) = p_i(q)$ for $i = 1, \ldots, 8$. These notions are well defined as the relevant functions are continuous (cf., e.g., Billingsley [8]). We distinguish the following restrictions to be explained below:

$$0 < \textstyle\sum_{i=1}^4 q_i\,(1 - p_i) = \sum_{i=5}^8 q_i\, p_i \text{ and } 0 < \sum_{i=1}^4 q_i < 1, \tag{12.2}$$

$$\textstyle\sum_{i=5}^8 q_i = 1, \text{ and } q_i > 0 \Longrightarrow p_i = 0,\ i = 5, \ldots, 8, \tag{12.3}$$

$$\textstyle\sum_{i=1}^4 q_i = 1, \text{ and } q_i > 0 \Longrightarrow p_i = 1,\ i = 1, \ldots, 4. \tag{12.4}$$

Restriction (12.2) is a conservation of flow equation: play takes place on both states infinitely often, therefore, due to the law of large numbers the actual instances of leaving *High* must be proportional to the long run probability of leaving it and the latter must be equal to the probability of returning.

If the long run play occurs in *Low* exclusively, (12.3) must hold. The former part is obvious, if $q_i\, p_i > 0$ for some $i = 5, \ldots, 8$, then play would visit the corresponding entry infinitely often as time goes to infinity, hence with probability at least $q_i\, p_i$ state *High* would occur. Similar reasoning applies to the other case that play occurs only in *High*, hence (12.4). We now show the implications for jointly-convergent pure strategies.

Example 2 Now, (12.3) can only hold if $p_i = 0$ or $q_i = 0$ for all $i = 5, \ldots, 8$. Similarly, (12.4) can only hold if $1 - p_i = 0$ or $q_i = 0$ for all $i = 1, \ldots, 4$. So, *if a state is absorbing, then positive mass on a component of the relative frequency vector q can only occur if the associated probability of leaving that state is zero*. Observe

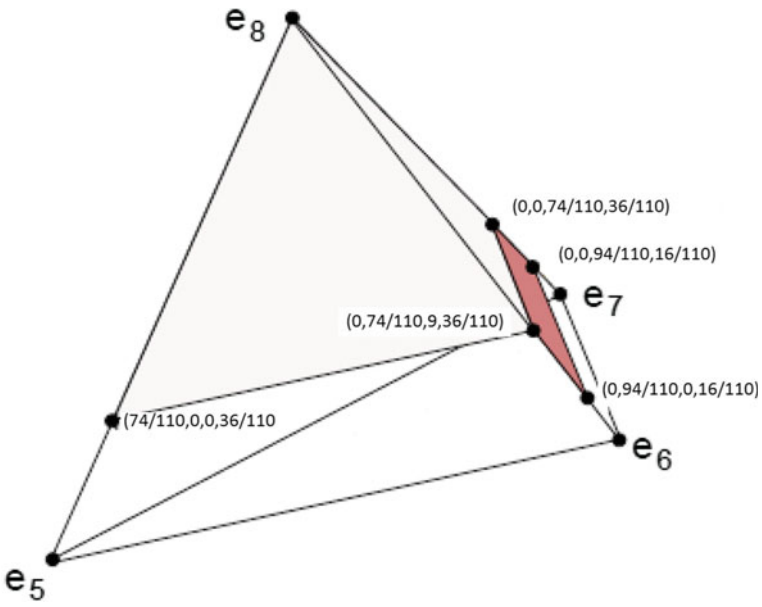


Fig. 12.1 If play concentrates on *Low*, $q_1 = \cdots = q_4 = 0$ and $q_5 + \cdots + q_8 = 1$. We depict this face of Δ^7 as a "projection" unto Δ^3. Extreme point e_i has component $i - 4$ equal to one. The admissible q's, are sketched as the three-dimensional set on top, and the two-dimensional boundary set

that therefore only *Low* can be absorbing. From the ranking of probabilities, we may distinguish the following three subcases.

$$\begin{array}{l} q_8 = 1 \text{ and } p_8 = \frac{1}{10} - \frac{11}{4}q_4 - \frac{11}{2}q_8 \leq 0 \text{ or} \\ \sum_{i=6}^{8} q_i = 1 \text{ and } p_6 = p_7 = \frac{4}{10} - \frac{11}{8}q_4 - \frac{11}{4}q_8 \leq 0 \text{ or} \\ \sum_{i=5}^{8} q_i = 1 \text{ and } p_5 = \frac{6}{10} - \frac{11}{12}q_4 - \frac{11}{6}q_8 \leq 0. \end{array}$$

Clearly, $q_4 = 0$. The first case is easily checked reducing analysis to

$$\begin{array}{l} \sum_{i=6}^{8} q_i = 1 \text{ and } \frac{16}{110} \leq q_8 \leq \frac{36}{110}, \text{ or} \\ \sum_{i=5}^{8} q_i = 1 \text{ and } q_8 \geq \frac{36}{110}. \end{array}$$

leading to $q_5 = 0$ and $\frac{16}{110} \leq q_8 \leq \frac{36}{110}$, and $q_8 \geq \frac{36}{110}$ and $q_5, \ldots, q_7 \geq 0$.

Figure 12.1 visualizes these restrictions for *Low* being absorbing. The upper three-dimensional subset of Δ^3, is connected to the final inequality; the parallelogram on the face of Δ^3 is connected to the former.

12.4 On Rewards and Equilibrium Rewards

The players receive an infinite stream of stage payoffs, they are assumed to wish to maximize their average rewards. For a given pair of strategies (π, σ), $R_t^k(\pi, \sigma)$ is the expected payoff to player k at stage t under strategy combination (π, σ), then player k's **average reward,** $k = A, B$, is $\gamma^k(\pi, \sigma) = \liminf_{T\to\infty} \frac{1}{T} \sum_{t=1}^{T} R_t^k(\pi, \sigma)$, and $\gamma(\pi, \sigma) \equiv \left(\gamma^A(\pi, \sigma), \gamma^B(\pi, \sigma)\right)$. Moreover, for vector $q \in \Delta^7$, the q**-averaged payoffs** $(x, y)_q$ are given by

$$(x, y)_q = \sum_{i=1}^{8} q_i \theta_i.$$

The strategy pair $(\pi^*, \sigma^*) \in \mathcal{X}^A \times \mathcal{X}^B$ is **an equilibrium** if and only if

$$\gamma^A(\pi^*, \sigma^*) \geq \gamma^A(\pi, \sigma^*) \text{ for all } \pi \in \mathcal{X}^A$$
$$\gamma^B(\pi^*, \sigma^*) \geq \gamma^B(\pi^*, \sigma) \text{ for all } \sigma \in \mathcal{X}^B.$$

The rewards $\gamma(\pi^*, \sigma^*)$ associated with an equilibrium (π^*, σ^*) will be referred to as equilibrium rewards.

In the analysis of repeated games, another helpful measure to reduce complexity is to focus on rewards instead of strategies. It is more rule than exception that one and the same reward combination can be achieved by several distinct strategy combinations. Here, we focus on rewards to be obtained by jointly-convergent pure strategies.

12.4.1 Jointly Convergent Pure-Strategy Rewards

The next result connects notions introduced in the previous sections.

Proposition 1 *Let strategy pair $(\pi, \sigma) \in \mathcal{JC}$ and let $q \in \Delta^7$ for which (12.1) is satisfied, then the average payoffs are given by $\gamma(\pi, \sigma) = (x, y)_q$.*

Proof Let $(\pi, \sigma) \in \mathcal{JC}$ and $E\{\theta_u^{\pi,\sigma}\} \equiv \left(R_u^1(\pi, \sigma), R_u^2(\pi, \sigma)\right)$, then

$$\lim_{t\to\infty} \tfrac{1}{t} \sum_{u=1}^{t} E\{\theta_u^{\pi,\sigma}\} = \lim_{t\to\infty} E\left\{\tfrac{1}{t} \sum_{u=1}^{t} \theta_u^{\pi,\sigma}\right\} =$$
$$\lim_{t\to\infty} E\left\{\sum_{i=1}^{8} q_i^t \theta_i\right\} = \lim_{t\to\infty} \sum_{i=1}^{8} E\left\{q_i^t\right\} \theta_i = \sum_{i=1}^{8} q_i \theta_i = (x, y)_q.$$

The second equality sign involves a change in counting: on the left-hand side we sum over all periods, on the right-hand side over all eight entries of the two bi-matrices weighed by their relative frequencies. Equalities one and three are standard, the penultimate one follows from (12.1), cf., e.g., Billingsley [8, p. 274], the final one by the definition given above. Since $\lim_{t\to\infty} \frac{1}{t} \sum_{u=1}^{t} E\{\theta_u^{\pi,\sigma}\}$ equals $(x, y)_q$, it follows that $\gamma(\pi, \sigma) = (x, y)_q$.

Example 3 To continue the example, we add stage payoffs

$$H\left(q^t\right) = \begin{bmatrix} (4,4)\,,\, p_1\left(q^t\right) & \left(\frac{7}{2}, 6\right)\,,\, p_2\left(q^t\right) \\ \left(6, \frac{7}{2}\right)\,,\, p_3\left(q^t\right) & \left(\frac{11}{2}, \frac{11}{2}\right)\,,\, p_4\left(q^t\right) \end{bmatrix},$$

$$L\left(q^t\right) = \begin{bmatrix} (2,2)\,,\, p_5\left(q^t\right) & \left(\frac{7}{4}, 3\right)\,,\, p_6\left(q^t\right) \\ \left(3, \frac{7}{4}\right)\,,\, p_7\left(q^t\right) & \left(\frac{11}{4}, \frac{11}{4}\right)\,,\, p_8\left(q^t\right) \end{bmatrix}.$$

Observe that $\theta_i = \frac{1}{2}\theta_{i-4}$ for $i = 5, \ldots, 8$. The specifics for the probabilities $p_1\left(q^t\right), \ldots, p_8\left(q^t\right)$ were already presented earlier. Note that in both states, the first action is dominated by the second for both players.

Figure 12.2 shows the rewards consistent with *Low* being absorbing and note that this hexagon is not convex.[5] The link between rewards in Fig. 12.2 and the strategy restrictions visualized in Fig. 12.1 is that the extreme points in Fig. 12.2 have the following coordinates (i.e., rewards)

$$\begin{aligned}
&\tfrac{74}{110}(2,2) + \tfrac{36}{110}\left(\tfrac{11}{4}, \tfrac{11}{4}\right) = \left(\tfrac{247}{110}, \tfrac{247}{110}\right) \\
&\tfrac{74}{110}\left(3, \tfrac{7}{4}\right) + \tfrac{36}{110}\left(\tfrac{11}{4}, \tfrac{11}{4}\right) = \left(\tfrac{642}{220}, \tfrac{457}{220}\right) \\
&\tfrac{74}{110}\left(\tfrac{7}{4}, 3\right) + \tfrac{36}{110}\left(\tfrac{11}{4}, \tfrac{11}{4}\right) = \left(\tfrac{457}{220}, \tfrac{642}{220}\right) \\
&\tfrac{94}{110}\left(3, \tfrac{7}{4}\right) + \tfrac{16}{110}\left(\tfrac{11}{4}, \tfrac{11}{4}\right) = \left(\tfrac{652}{220}, \tfrac{417}{220}\right) \\
&\tfrac{94}{110}\left(\tfrac{7}{4}, 3\right) + \tfrac{16}{110}\left(\tfrac{11}{4}, \tfrac{11}{4}\right) = \left(\tfrac{417}{220}, \tfrac{652}{220}\right).
\end{aligned}$$

The first three rewards coincide with the lower three vertices of the shaded simplex of dimension 3 within Δ^3 in Fig. 12.1. The latter two coincide with the lower two vertices of the quadrangle on the face of Δ^3 in Fig. 12.1. Finally, the reward $\left(\frac{11}{4}, \frac{11}{4}\right)$ coincides with the vertex e_8 in Fig. 12.1.

So, e_5 corresponds to the situation that in the long run the relative frequency of play on action pair (1, 1) in *Low* is 1 (*if that were possible*). The left-hand lowest vertex of the shaded simplex in Fig. 12.1 has coordinates (74/110, 00, 36/100), so the corresponding rewards are obtained by the linear combination of both (2, 2) and $\left(\frac{11}{4}, \frac{11}{4}\right)$ with the associated weights.

Similarly, all interior points of the shaded simplex in Fig. 12.1 correspond to the interior of the shaded parallelogram in Fig. 12.2. The interior points of the boundary quadrangle in Fig. 12.1 correspond to the interior of the trapezium in Fig. 12.2.

We must also find rewards such that (12.2) is satisfied. Figure 12.3 shows all jointly-convergent pure-strategy rewards. For instance, rewards $\left(\frac{7}{2}, \frac{7}{2}\right)$ correspond to mutual full restraint; furthermore, the Pareto-efficient line segment connecting $\left(\frac{22}{6}, \frac{23}{6}\right)$ and $\left(\frac{23}{6}, \frac{22}{6}\right)$ is achieved by playing Top-Right in *High* and by playing the off-diagonal action pairs in *Low* exclusively.

[5] Figures 12.2 and 12.3 are based on Matlab graphs generated by an algorithm yielding 6 million pairs of rewards which took several days. Memory restrictions corrupt image quality as we experienced. The algorithm and output are available on request.

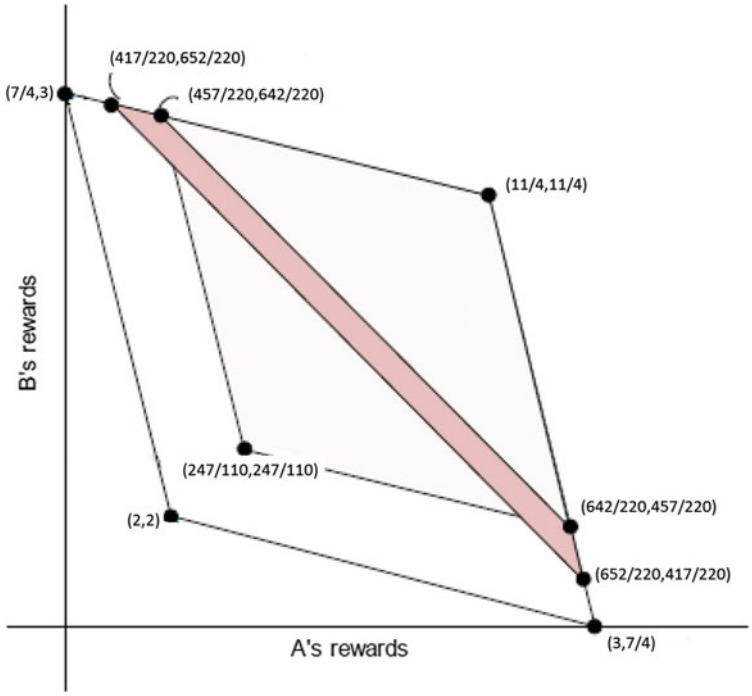


Fig. 12.2 A sketch of the hexagon being the union of the lightly shaded parallelogram and the darker trapezium. The former corresponds to the three-dimensional set, the latter to the two-dimensional boundary set in Fig. 12.1. The other rewards, corresponding to the convex hull of the four entries associated with *Low* are **not feasible by jointly-convergent pure strategies**

12.4.2 Equilibrium Rewards

We now focus on rewards from equilibria involving threats. Our approach is similar to a well-established one in the repeated games literature (cf., e.g., Hart [28], Forges [23]), linked to the Folk Theorem (see e.g., Van Damme [74]) and applied to stochastic games as well (cf., e.g., Thuijsman and Vrieze [71], Joosten et al. [42], Schoenmakers [67]).

We call $v = \left(v^A, v^B\right)$ the **threat point**, where $v^A = \min_{\sigma \in \mathcal{X}^B} \max_{\pi \in \mathcal{X}^A} \gamma^A(\pi, \sigma)$, and $v^B = \min_{\pi \in \mathcal{X}^A} \max_{\sigma \in \mathcal{X}^B} \gamma^B(\pi, \sigma)$. So, v^A is the highest amount A can get if B tries to minimize A's average payoffs. Under a pair of **individually rational** (feasible) rewards each player receives at least the threat-point reward.

Let $E = \left\{(x, y) \in P^{\mathcal{JC}} | \; x > v^A \text{ and } y > v^B\right\}$ be the set of all individually rational jointly convergent pure-strategy rewards giving each player strictly more than his threat point reward. We can now present the following formal result:

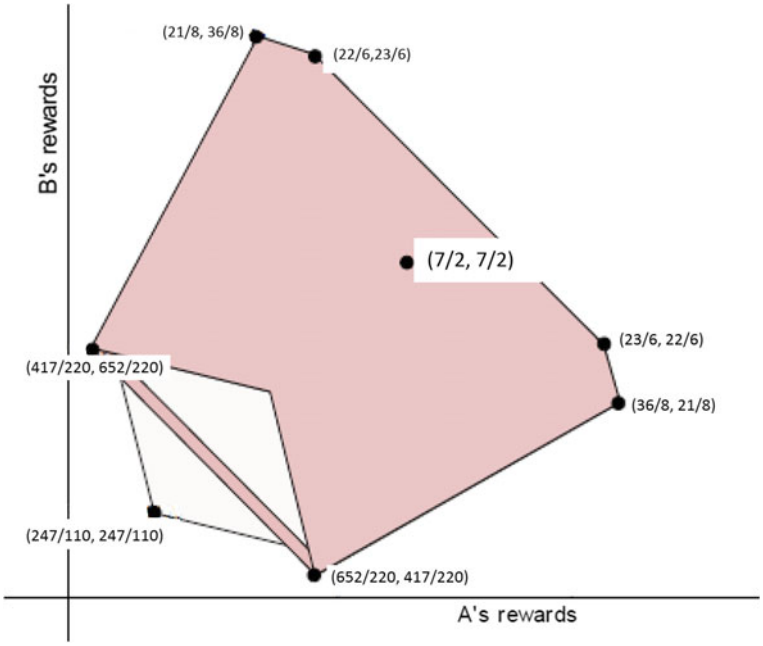


Fig. 12.3 The set $P^{\mathcal{JC}}$: on the lower left-hand side the hexagon of Fig. 12.2, for other rewards both states are visited infinitely often

Theorem 1 *Each pair of rewards in E can be supported by an equilibrium.*

Proof Let $(x, y) \in E$, then a pure-strategy combination $(\pi, \sigma) \in \mathcal{JC}$ exists such that $\gamma(\pi, \sigma) = (x, y)$. Let $\varepsilon = \frac{1}{2} \min\left(x - v^A, y - v^B\right)$ and let π^p (σ^p) be a punishment-strategy of A (B), i.e., a strategy holding his opponent to at most $v^B + \varepsilon$ $(v^A + \varepsilon)$. Let

$$\pi_t^* \equiv \begin{cases} \pi_t & \text{if } j_k = \sigma_k^* \text{ for all } k < t, \\ \pi_t^p & \text{otherwise.} \end{cases}$$

$$\sigma_t^* \equiv \begin{cases} \sigma_t & \text{if } i_k = \pi_k^* \text{ for all } k < t, \\ \sigma_t^p & \text{otherwise.} \end{cases}$$

Here, i_t (j_t) denotes the action taken by player A (B) at stage t of the play. Clearly, $\gamma(\pi^*, \sigma^*) = \gamma(\pi, \sigma) = (x, y)$. Suppose player A were to play π' such that $\pi'_k \neq \pi_k^*$ for some k, then player B would play according to σ^p from then on. Since, $\gamma^A\left(\pi', \sigma^p\right) \leq v^A + \varepsilon < x$, it follows immediately that player A cannot improve against σ^*. A similar statement holds in case player B deviates unilaterally. Hence, (π^*, σ^*) is an equilibrium.

Such a pair of strategies (π^*, σ^*) is called an equilibrium involving threats, e.g., Hart [28], Van Damme [74], Thuijsman and Vrieze [71].

Joosten et al. [42] prove by construction that each reward in the convex hull of E can be supported by an equilibrium, too. Equilibrium rewards in the convex hull of E *not* in E can be obtained by history-dependent strategies with threats, which are *neither* jointly-convergent, *nor* pure. The construction of Joosten et al. [42] involves a randomization phase which obviously violates the pure-strategy part. The randomization phase serves to identify and communicate to both players which equilibrium pair of jointly convergent pure strategies is to be played afterwards. So, this also violates the very notion of jointly convergent strategies. This construction need not work for every stochastic game, but for the present class of games it does as no state is absorbing (permanently).

Whether equilibria exist yielding rewards that are not in the convex hull of E, is an open question. Such equilibria then must be associated with strategies which are not jointly convergent. For instance, in the example here, it can be shown by construction that rewards in the convex hull of $\left(\frac{417}{220}, \frac{417}{220}\right)$ and $P^{\mathcal{JC}}$ can be obtained for the average reward criterion using the limes inferior. Similarly, although this is out of the scope of this chapter, one can obtain the convex hull of $\left(\frac{7}{4}, \frac{7}{4}\right)$ and $P^{\mathcal{JC}}$ as feasible rewards for the average reward criterion using the limes superior. For the latter criterion all additional rewards Pareto dominate all equilibrium rewards in $P^{\mathcal{JC}}$. Therefore, these rewards can be supported by equilibria as well for this alternative evaluation criterion.

Theorem 1 hinges on the possibility of punishing unilateral deviations, as in e.g., Hämäläinen et al. [25]. So, we cannot restrict ourselves to Markov or stationary strategies as these types of strategies do not offer the strategic richness to allow punishing. History-dependent strategies do offer the required flexibility, but it is an open question whether less general classes of strategies might suffice. What is clear though, is that action independent strategies do not.

There is no contradiction between strategy pairs being both jointly-convergent and history-dependent, or for that matter cooperative, e.g., Tołwinski [72], Tołwinski et al. [73], Krawczyk and To łwinski [48], or incentive strategies, or combinations, e.g., Ehtamo and Hämäläinen [15–18].

12.4.3 On Computing Threat Points

We illustrate Theorem 1 and the notions introduced. Moreover, we use the examples to show the scope of the problem of computing threat points. The next example shows that linear programs may not suffice.

Example 4 Assume that player B uses his second action at all stages of the play. Now, consider the (nonlinear) program

$$
\begin{aligned}
&\min_{q_2,q_4,q_6,q_8} 6q_2 + \tfrac{11}{2}q_4 + 3q_6 + \tfrac{11}{4}q_8\\
&\text{s.t. } 1 = q_2 + q_4 + q_6 + q_8\\
&0 = (1-p_2)q_2 + (1-p_4)q_4 - p_6q_6 - p_8q_8\\
&p_2 = \left[\tfrac{6}{10} - \tfrac{11}{20}q_4 - \tfrac{11}{10}q_8\right]_+\\
&p_4 = \left[\tfrac{3}{10} - \tfrac{11}{16}q_4 - \tfrac{11}{8}q_8\right]_+\\
&p_6 = \left[\tfrac{4}{10} - \tfrac{11}{8}q_4 - \tfrac{11}{4}q_8\right]_+\\
&p_8 = \left[\tfrac{1}{10} - \tfrac{11}{4}q_4 - \tfrac{11}{2}q_8\right]_+\\
&0 \le q_2, q_4, q_6, q_8.
\end{aligned}
$$

Clearly, $q_8 = 1$ yields rewards equal to $\frac{11}{4}$; all other feasible rewards involve $q_8 < 1$ yielding a reward strictly higher than $\frac{11}{4}$. Evidently, player B can guarantee himself at least 2.75. This implies $v^B \ge 2.75$.

Next, we aim to show that player A can hold his opponent to at most 2.75 by using his second action at all stages of the play. First, we argue that the best reply of player B resulting in a pair of jointly convergent strategies yields at most 2.75. Then, we argue that if B uses a strategy resulting in a pair of strategies which is not jointly convergent, then this cannot yield more than 2.75. We do not provide the lengthy computations underlying our findings,[6] only intuitions.

For the first part, since we assume that the pair of strategies is jointly convergent, we may consider the (nonlinear) program

$$
\begin{aligned}
&\max_{q_3,q_4,q_7,q_8} \tfrac{7}{2}q_3 + \tfrac{11}{2}q_4 + \tfrac{7}{4}q_7 + \tfrac{11}{4}q_8\\
&\text{s.t. } 1 = q_3 + q_4 + q_7 + q_8\\
&0 = (1-p_3)q_3 + (1-p_4)q_4 - p_7q_7 - p_8q_8\\
&p_3 = \left[\tfrac{6}{10} - \tfrac{11}{20}q_4 - \tfrac{11}{10}q_8\right]_+\\
&p_4 = \left[\tfrac{3}{10} - \tfrac{11}{16}q_4 - \tfrac{11}{8}q_8\right]_+\\
&p_7 = \left[\tfrac{4}{10} - \tfrac{11}{8}q_4 - \tfrac{11}{4}q_8\right]_+\\
&p_8 = \left[\tfrac{1}{10} - \tfrac{11}{4}q_4 - \tfrac{11}{2}q_8\right]_+\\
&0 \le q_3, q_4, q_7, q_8.
\end{aligned}
$$

Observe that if $p_7 = 0$, then $p_8 = 0$ as well, hence $q_3 = q_4 = 0$. Then, the maximization program implies $q_8 = 1$ and the value of the objective function is $\frac{11}{4}$. Let us define $e_k = (q_3, q_4, q_7, q_8)$ by $q_k = 1$, $q_j = 0$ for $j \ne k$. Now, $p_7 = 0$ if the relative frequency vector (q_3, q_4, q_7, q_8) is in

$$
\begin{aligned}
S^0 = conv\Big\{\Big\{&e_4, e_8, \left(\tfrac{78}{110}, \tfrac{32}{110}, 0, 0\right), \left(0, \tfrac{32}{110}, \tfrac{78}{110}, 0\right)\Big\}\\
&\cup \Big\{\left(0, 0, \tfrac{94}{110}, \tfrac{16}{110}\right), \left(\tfrac{94}{110}, 0, 0, \tfrac{16}{110}\right)\Big\}\Big\},
\end{aligned}
$$

where $conv\ S$ denotes the convex hull of set S. Possible higher rewards are only to be found for $(q_3, q_4, q_7, q_8) \in \Delta^3 \backslash S^0$.

Furthermore, $1 - p_4 > 1 - p_3 \ge \frac{4}{10} \ge p_7 > p_8$, hence $q_3 + q_4 \le \frac{1}{2} \le q_7 + q_8$. So, only tuples (q_3, q_4, q_7, q_8) in

[6] They are available on request, of course.

$$S^1 = conv\left\{\left\{\left(\tfrac{1}{2}, 0, \tfrac{1}{2}, 0\right), \left(\tfrac{1}{2}, 0, \tfrac{39}{110}, \tfrac{16}{110}\right), \left(\tfrac{23}{110}, \tfrac{32}{110}, \tfrac{1}{2}, 0\right)\right\} \cup \left\{e_7, \left(0, 0, \tfrac{94}{110}, \tfrac{16}{110}\right), \left(0, \tfrac{32}{110}, \tfrac{78}{110}, 0\right)\right\}\right\}.$$

may yield higher rewards than $\frac{11}{4}$. This follows from the observation that the sum of the probabilities to move to (from) *Low* is always above (below) $\frac{4}{10}$, hence the (long term) proportion of the play spent in *Low* is at least $\frac{1}{2}$.

The points in S^1 satisfying the restriction

$$0 = (1 - p_3)q_3 + (1 - p_4)q_4 - p_7q_7 - p_8q_8$$

form a two-dimensional manifold, say M, and the restriction is clearly violated in a neighborhood of the plane

$$P = conv\left\{\left(\tfrac{1}{2}, 0, \tfrac{39}{110}, \tfrac{16}{110}\right), \left(\tfrac{23}{110}, \tfrac{32}{110}, \tfrac{1}{2}, 0\right), \left(0, 0, \tfrac{94}{110}, \tfrac{16}{110}\right), \left(0, \tfrac{16}{55}, \tfrac{39}{55}, 0\right)\right\}$$

which is the facet of S^1 opposite the line segment $\left(\frac{1}{2} - x, 0, \frac{1}{2} + x, 0\right)$, $x \in [0, \frac{1}{2}]$. Hence, M does not intersect P. The following defines for $\alpha \in [0, \frac{16}{55}]$ a family of two-dimensional planes in S^1:

$$S(\alpha) = \left\{(q_3, q_4, q_7, q_8) \in S^1 | q_4 + q_8 = \alpha\right\}.$$

For increasing α, we establish whether $S(\alpha) \cap M \neq \varnothing$, and in that case the intersection is either a point, a line segment or a two-dimensional subset of $S(\alpha)$. Any unique point in this intersection with the highest weight on q_4 clearly maximizes the objective function for $S(\alpha)$; otherwise a one-dimensional set of points exist with highest weights on q_4, then the point with the highest weight on q_3 is the solution with respect to $S(\alpha)$. So, for fixed α one observes immediately that $q_4 = \alpha$ and $q_8 = 0$ for any solution with respect to $S(\alpha)$.

Take $q_3 + q_7 = 1$, then $1 - p_3 = p_7 = \frac{4}{10}$ which in turn implies $q_3 = q_7 = \frac{1}{2}$. In this case, $\frac{7}{2}q_3 + \frac{11}{2}q_4 + \frac{7}{4}q_7 + \frac{11}{4}q_8 = \frac{21}{8}$. To obtain higher values of the objective function q_4 should be increased from zero while keeping $q_8 = 0$. The final point is that the one-dimensional set of solutions restricted to such $S(\alpha)$ for $\alpha \in [0, \frac{16}{55}]$ "beginning at" $\left(\frac{1}{2}, 0, \frac{1}{2}, 0\right)$ does not lead to higher values of the objective function than $\frac{21}{8}$.

As no solution satisfying the restrictions of the maximization problem, yields more than $\frac{11}{4}$ in $\Delta^3 \backslash S^0$, the solution is located in S^0, so the global solution is $q_8 = 1$; the connected reward to player B is 2.75. As player A can hold B to this amount, we have $v^B \leq 2.75$. Hence, under the assumption that the outcome of the maximization problem of player B against his opponent using his second action in any state and at any stage, is a jointly convergent pair of strategies, we find $v^B = 2.75$.

Now, we continue our reasoning with the **assumption that the maximization problem does not** result in a pair of jointly-convergent strategies. First, note that the latter expression in the present framework means that B uses a strategy σ against

player A playing $\pi^* = (2, 2, 2, \ldots)$, such that $q^t = \left(q_3^t, q_4^t, q_7^t, q_8^t\right)$ never converges, i.e., q^t must move around in the three-dimensional unit simplex forever.

Note that if for some (non-jointly convergent) pair of strategies (π^*, σ) and some T, it holds that[7] $\{q^t\}_{t \geq T} \subset S^0$, then $\lim_{t\to\infty} q_3^t = \lim_{t\to\infty} q_4^t = 0$. This follows from the circumstance that $p_7\left(q^t\right) = p_8\left(q^t\right) = 0$ for all $t \geq T$. So, the long-term average payoffs at point t in time for t sufficiently large satisfy

$$\frac{7}{4}q_7^t + \frac{11}{4}q_8^t = \frac{7}{4}q_7^t + \frac{11}{4}\left(1 - q_7^t\right) = \frac{11}{4} - q_7^t < \frac{11}{4}.$$

This means that $\gamma^B(\pi^*, \sigma) < \frac{11}{4}$.

Furthermore, let $S^2 = conv\{e_7, e_8, \left(\frac{4}{7}, \frac{3}{7}, 0, 0\right), \left(0, \frac{11}{15}, \frac{4}{15}, 0\right)\}$. Then it is easily confirmed that $\frac{7}{2}q_3 + \frac{11}{2}q_4 + \frac{7}{4}q_7 + \frac{11}{4}q_8 \leq \frac{11}{4}$ for all $q \in S^2$. Hence, if for some (π^*, σ) it holds that

$$\limsup_{T\to\infty}\left[\Pr_{\pi^*,\sigma}\left[\frac{\#\left\{q^t \subset S^2 | t \leq T\right\}}{T}\right] \geq \varepsilon\right] > 0 \text{ for all } \varepsilon > 0,$$

then $\gamma^B(\pi^*, \sigma) \leq \frac{11}{4}$.

Let $S^3 = \Delta^3 \backslash \left(S^0 \cup S^2\right)$ and note that $\frac{7}{2}q_3 + \frac{11}{2}q_4 + \frac{7}{4}q_7 + \frac{11}{4}q_8 \geq \frac{11}{4}$ for $q \in S^3$. By choosing a set of convenient (but not even tight) upper and lower bounds it takes quite some effort to confirm that if for some (π^*, σ)

$$\limsup_{T\to\infty}\left[\Pr_{\pi^*,\sigma}\left[\frac{\#\left\{q^t \subset S^3 | t \leq T\right\}}{T}\right] \geq \varepsilon\right] = 0 \text{ for all } \varepsilon > 0,$$

then $\gamma^B(\pi^*, \sigma) < \frac{11}{4}$. This contradiction implies that it is impossible to guarantee play such that the resulting relative frequencies vectors stay in S^3 (hence out of $S^0 \cup S^2$) almost forever.

So, candidates to yield a limiting average reward higher than $\frac{11}{4}$ must induce play such that relative frequency vectors stay forever in $\left(S^0 \backslash S^2\right) \cup S^3$. However, in $\left(S^0 \backslash S^2\right) \cup S^3$ there is persistent drift away from $conv\{e_3, e_4\}$ because the transition probabilities from Low to High are small and the transition probabilities from High to Low are large. Away from $conv\{e_3, e_4\}$ means towards $conv\{e_7, e_8\}$ which implies that the play will induce relative frequency vectors in S^2. Note that due to the assumption that (π^*, σ) is not jointly convergent means e_8 can only be approached infinitely often by relative frequency vectors from S^2 or returning to S^2, yielding limiting average rewards below $\frac{11}{4}$.

The negative results above imply that the maximization problem can be solved in jointly convergent strategies in this example. Hence, $v^B = 2.75$ (Fig. 12.4).

[7] Hordijk et al. [34] show that a stationary strategy suffices as a best reply against a fixed stationary strategy, and we may write the next sequence as a deterministic one.

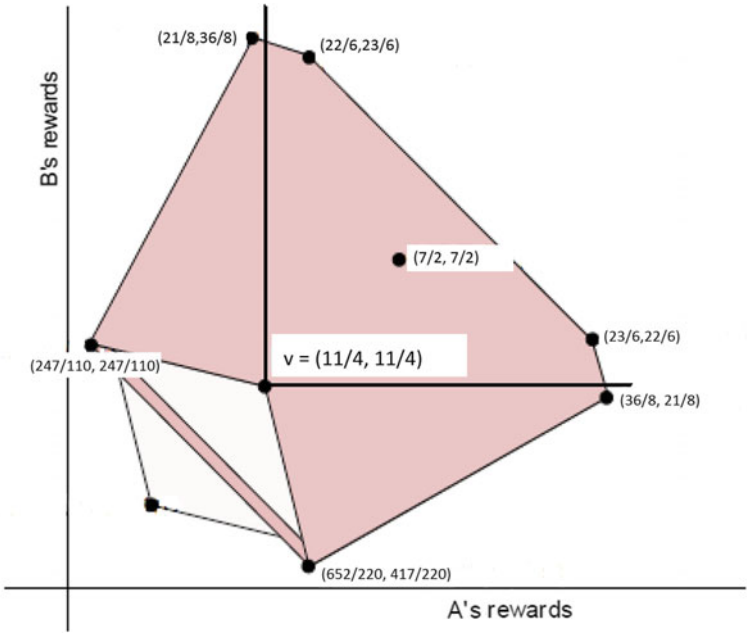


Fig. 12.4 Each pair of jointly convergent pure-strategy rewards to the "north-east" of $v = (2.75, 2.75)$ can be supported by an equilibrium involving threats

Example 4 illustrates that finding threat points may be cumbersome as it requires at least a nonlinear program. Our approach was to alternate a minimization and a maximization program against sequences of stationary strategies to obtain lower and upper bounds for the threat point. If solutions coincide, as in the example above after two steps, we are done. Otherwise, all rewards yielding more than the lowest upper bound established can be associated to equilibria involving threats.

We can interpret every minimization and maximization program as a single controller stochastic game (cf., e.g., Parthasarathy and Raghavan [60]). However, the circumstance that the number of states captured in the relative frequency vectors (please recall our remarks on this issue in Sect. 12.3) is not finite takes our problem out of the scope of the algorithms implied to compute the associated values (e.g., Filar and Raghavan [19], Vrieze [75], see Raghavan and Filar [62] for a survey). Hordijk et al. [34] show that a stationary strategy suffices as a best reply against a fixed stationary strategy, and the optimization problems mentioned reduce to Markov decision problems (cf., e.g., Filar and Vrieze [20]). We used these results partially above,[8] but found not much help in them otherwise.

[8] In earlier versions of our paper we were too quick to conclude that the associated optimization problems yield jointly convergent strategies. A referee pointed out a flaw in our reasoning, which by the way, makes to problem of finding an optimal strategy against a fixed strategy even much harder to solve. If jointly convergent strategies do not yield a solution, play never settles down measured in

The *general* problem is equivalent to finding the value of a zero-sum stochastic game. Well-known techniques from standard stochastic game theory, e.g., Bewley and Kohlberg [6, 7] and Mertens and Neyman [56], offer insufficient solace because of the state space which is not finite but denumerable.

12.5 Conclusions

We added an innovation to the framework of Small Fish Wars (e.g., Joosten [37, 38, 41]) by allowing endogeneity in the transition structure: transition probabilities depend on the actions taken by the agents currently in the current state and on the history of the play. In this new setting states may become absorbing *temporarily*. Here, this feature is used to model the phenomenon that, even if the agents turn to ecologically sound exploitation policies, it may take a long time before the first transition to a state yielding higher outcomes occurs if the state *Low* turns out to have become temporarily absorbing. Thus, we capture hysteresis, called a poaching pit in the management of natural resources literature (cf., e.g., Bulte [11]). Hysteresis is an empirical phenomenon and may be observed in the slow recovery of coastal cod stocks in Canada after a moratorium on cod fishing since 1992 (cf., Rose et al. [63]). More recent estimates of stocks show a less bleak picture due to recent developments unrelated to resource management, but the stocks are still far removed from high historical levels.

Our approach generalizes standard stochastic games,[9] too. We propose methods of analysis originally introduced in Joosten et al. [42] inspired by Folk Theorems for stochastic games e.g., Thuijsman and Vrieze [71], Joosten [35, 36] and Schoenmakers [67], and developed further in for instance Joosten [37, 38, 41]. Crucial notion is that of jointly-convergent strategies which justify the necessary steps in creating analogies to the Folk Theorem. In our view, it is convenient that the complex model arising from endogenous transition probabilities may be solved quite analogously to repeated games.[10]

---

the space of the relative frequency vectors and the sequence of relative frequency vectors induced is essentially stochastic.

[9] At several presentations the question was raised whether our games should not be presented as stochastic games with infinitely many states. We agree that our games fall into this class, as they can be rewritten as such. We prefer our presentation because of its simplicity and the circumstance that we were able to generate a number of results. Moreover, we are very sceptic about which known results from the analysis of stochastic games with infinitely many states would be helpful to obtain results for ours.

[10] We like our rather complex model to resemble repeated games for psychological reasons and for reasons of ease of communication for instance with less mathematically inclined people (politicians, civil servants). Many people have learned about the repeated prisoners' dilemma in educational programs, so offering our model in a simple fashion may offer windows of opportunity for communication with the general public. To present our model as a stochastic game with infinitely many states might scare researchers but more likely less mathematically inclined people away.

Our analysis of a special example with hysteresis shows that a "tragedy of the commons" can be averted by sufficiently patient rational agents[11] maximizing their utilities non-cooperatively. All equilibrium rewards yield more than the amounts associated to the permanent ruthless exploitation of the resource. Pareto optimal equilibrium rewards correspond to strategy pairs involving a considerable amount of restraint on the part of the agents, and are considerably higher than no-restraint rewards and slightly higher than perfect-restraint rewards.

To present a tractable model and to economize on notations, we kept the fish stock fixed yet stochastic, i.e., the variation in stock size and catches is only due to random effects; we imposed symmetry and used the three "twos": two states, two players and two actions. Two distinct states allow to model the kind of transitions we had in mind; two agents are minimally required to model strategic interaction; two stage-game actions leave something to choose. In order to capture additional real-life phenomena observed, such as seasonalities or other types of correlations, a larger number of states may be required. Furthermore, more levels or dimensions of restraining measures may be necessary. Adding states, (asymmetric) players or actions changes nothing to our approach conceptually.

By keeping the model and its analysis relatively simple, hence presumably more tractable, further links to and comparisons with contributions in the social dilemma literature, cf., e.g., Komorita and Parks [1994], Heckathorn [30], Marwell and Oliver [53] where dyadic choice is predominant, may be facilitated. Our resource game is to be associated primarily with a social trap, see e.g., Hamburger [26], Platt [61], Cross and Guyer [14] of which the 'tragedy of the commons' cf., e.g., Hardin [27], Messick et al. [55], Messick and Brewer [54]) is a special notorious example.

Ongoing related research focusses on designing algorithms improving computational efficiency of existing ones to generate large sets of jointly-convergent pure-strategy rewards. The algorithms used to find the rewards visualized in consecutive figures in this chapter are unacceptably slow. This was an unpleasant surprise as they were in fact modifications of algorithms working extremely rapidly in models within the same and related frameworks (e.g., Joosten [37, 39–41]). The new algorithms not only generate the desired sets within acceptable computing times here, but also seem much more efficient than our algorithms used before when applied to certain repeated games, stochastic games and games with frequency dependent stage payoffs (cf., Joosten and Samuel [43, 44]).

Related ongoing research is devoted to computing threat points with spin-offs of the algorithms of Joosten and Samuel [43, 44] for the same models as mentioned in the previous paragraph. This is a solution born out of necessity because very little

[11] Our agent is not the individual fisherman, but rather countries, regions, villages or cooperatives. Whether or not the latter care for the future sufficiently to induce sustainability (see e.g., Ostrom [58], Ostrom et al. [59] for optimistic views), individual fisherman's preferences seem too myopic (cf., e.g., Hillis and Wheelan [32]). Next to impatience of the agents, their number, communication, punishment possibilities and the observability of actions taken influence the likelihood that the tragedy of the commons can be averted (cf., e.g., Komorita and Parks [47], Ostrom [58, 59], Steg [70]).

is known on finding threat points in this new framework. Future research should address this knowledge gap.

Future research should combine the various modifications and extensions of the original Small Fish Wars [37] with the innovation presented here. Joosten [41] adds various price-scarcity feedbacks to the model, as well as another low-density phenomenon called the Allee effect. For the majority of results and our methods of analysis we anticipate to need no more than the notion of jointly convergent strategies and continuity of stage payoff functions and transition probability functions involved.

We envision applications of stochastic games with endogenous transitions where hysteresis-like phenomena occur, for instance shallow lakes (e.g., Scheffer [65], Carpenter et al. [12], Mäler et al. [52]), labor markets (e.g., Blanchard and Summers [9]), climate change (e.g., Lenton et al. [49]), or more general, where tipping or regime shifts may occur [2, 66]. We also see possible extensions of earlier models on (un)learning by (not) doing, cf., Joosten et al. [42, 45], and related work, e.g., Schoenmakers et al. [68], Schoenmakers [67], Flesch et al. [22].

References

1. Amir, R.: Stochastic games in economics and related fields: an overview. In: Neyman, A., Sorin, S. (eds.) Stochastic Games and Applications. NATO Advanced Study Institute, Series D, pp. 455–470. Kluwer, Dordrecht (2003)
2. Anderson, T., Carstensen, J., Hernández-Garcia, E., Duarte, C.M.: Ecological thresholds and regime shifts: approaches to identification. Trends Ecol. Evol. **24**, 49–57 (2008)
3. Armstrong, M.J., Connolly, P., Nash, R.D.M., Pawson, M.G., Alesworth, E., Coulahan, P.J., Dickey-Collas, M., Milligan, S.P., O'Neill, M., Witthames, P.R., Woolner, L.: An application of the annual egg production method to estimate spawning biomass of cod (Gadus morhua L.), plaice (Pleuronectes platessa L.) and sole (Solea solea L.) in the Irish Sea. ICES J. Mar. Sci. **58**, 183–203 (2001)
4. Aumann, R.: Game engineering. In: Neogy, S.K., Bapat, R.B., Das, A.K., Parthasarathy, T. (eds.) Mathematical Programming and Game Theory for Decision Making, pp. 279–286. World Scientific, Singapore (2008)
5. BenDor, T., Scheffran, J., Hannon, B.: Ecological and economic sustainability in fishery management: a multi-agent model for understanding competition and cooperation. Ecol. Econ. **68**, 1061–1073 (2009)
6. Bewley, T., Kohlberg, E.: The asymptotic theory of stochastic games. Math Oper Res. **1**, 197–208 (1976)
7. Bewley, T., Kohlberg, E.: The asymptotic solution of a recursive equation occuring in stochastic games. Math. Oper. Res. **1**, 321–336 (1976)
8. Billingsley, P.: Probability and Measure. Wiley, New York (1986)
9. Blanchard, O., Summers, L.: Hysteresis and the European unemployment problem. In: Fisher, S. (ed.) NBER Macroecon. Annu., pp. 15–78. MIT Press, Cambridge (1986)
10. Brooks, S.E., Reynolds, J.D., Allison, A.E.: Sustained by snakes? seasonal livelihood strategies and resource conservation by Tonle Sap fishers in Cambodia. Hum. Ecol. **36**, 835–851 (2008)
11. Bulte, E.H.: Open access harvesting of wildlife:the poaching pit and conservation of endangered species. Agric. Econ. **28**, 27–37 (2003)
12. Carpenter, S.R., Ludwig, D., Brock, W.A.: Management of eutrophication for lakes subject to potentially irreversible change. Ecol. Appl. **9**, 751–771 (1999)

13. Courchamp, F., Angulo, E., Rivalan, P., Hall, R.J., Signoret, L., Meinard, Y.: Rarity value and species extinction: the anthropogenic Allee effect. PLoS Biol. **4**, 2405–2410 (2006)
14. Cross, J.G., Guyer, M.J.: Social Traps. University of Michigan Press, Ann Arbor (1980)
15. Ehtamo, H., Hämäläinen, R.P.: On affine incentives for dynamic decision problems. In: Başar, T. (ed.) Dynamic Games and Applications in Economics, pp. 47–63. Springer, Berlin (1986)
16. Ehtamo, H., Hämäläinen, R.P.: Incentive strategies and equilibria for dynamic games with delayed information. JOTA **63**, 355–369 (1989)
17. Ehtamo, H., Hämäläinen, R.P.: A cooperative incentive equilibrium for a resource management problem. J. Econ. Dyn. Control. **17**, 659–678 (1993)
18. Ehtamo, H., Hämäläinen, R.P.: Credibility of linear equilibrium strategies in a discrete-time fishery management game. Group Decis. Negot. **4**, 27–37 (1995)
19. Filar, J., Raghavan, T.E.S.: A matrix game solution to a single-controller stochastic game. Math. Oper. Res. **9**, 356–362 (1984)
20. Filar, J., Vrieze, O.J.: Competitive Markov Decision Processes. Springer, Berlin (1996)
21. Flesch, J.: Stochastic games with the average reward. Ph.D. thesis, Maastricht University, ISBN 90-9012162-5 (1998)
22. Flesch, J., Schoenmakers, G., Vrieze, O.J.: Loss of skills in coordination games. Int. J. Game Theory **40**, 769–789 (2011)
23. Forges, F.: An approach to communication equilibria. Econometrica **54**, 1375–1385 (1986)
24. Hall, R.J., Milner-Gulland, E.J., Courchamp, F.: Endangering the endangered: the effects of perceived rarity on species exploitation. Conserv. Lett. **1**, 75–81 (2008)
25. Hämäläinen, R.P., Haurie, A., Kaitala, V.: Equilibria and threats in a fishery management game. Optim. Control. Appl. Methods **6**, 315–333 (1985)
26. Hamburger, H.: N-person prisoner's dilemma. J. Math. Psychol. **3**, 27–48 (1973)
27. Hardin, G.: The tragedy of the commons. Science **162**, 1243–1248 (1968)
28. Hart, S.: Nonzero-sum two-person repeated games with incomplete information. Math. Oper. Res. **10**, 117–153 (1985)
29. Haurie, A., Krawczyk, J.B., Zaccour, G.: Games and Dynamic Games. World Scientific, Singapore (2012)
30. Heckathorn, D.D.: The dynamics and dilemmas of collective action. Am. Sociol. Rev. **61**, 250–277 (1996)
31. Herings P.J.J., Predtetchinski, A.: Voting in collective stopping games, working paper Maastricht University (2012)
32. Hillis, J.F., Wheelan, J.: Fisherman's time discounting rates and other factors to be taken into account in planning rehabilitation of depleted fisheries. In: Antona, M., et al. (eds.) Proceedings of the 6th Conference of the International Institute of Fisheries Economics Trade, pp. 657–670. IIFET-Secretariat, Paris (1994)
33. Holden, M.: The Common Fisheries Policy: Origin, Evaluation and Future. Fishing News Books, Blackwell (1994)
34. Hordijk, A., Vrieze, O.J., Wanrooij, L.: Semi-Markov strategies in stochastic games. Int. J. Game Theory **12**, 81–89 (1983)
35. Joosten, R.: Dynamics, Equilibria, and Values. Ph.D. thesis, Faculty of Economics and Business Administration, Maastricht University (1996)
36. Joosten, R.: A note on repeated games with vanishing actions. Int. Game Theory Rev. **7**, 107–115 (2005)
37. Joosten, R.: Small Fish Wars: a new class of dynamic fishery-management games. ICFAI J. Manag. Econ. **5**, 17–30 (2007a)
38. Joosten, R.: Small Fish Wars and an authority. In: Prinz, A. (ed.) The Rules of the Game: Institutions, Law, and Economics, pp. 131–162. LIT, Berlin (2007)
39. Joosten, R.: Strategic advertisement with externalities: a new dynamic approach. In: Neogy, S.K., Das, A.K., Bapat, R.B. (eds.) Modeling, Computation and Optimization. ISI Platinum Jubilee Series, vol. 6, pp. 21–43. World Scientific Publishing Company, Singapore (2009)
40. Joosten, R.: Long-run strategic advertisement and short-run Bertrand competition. Int. Game Theory Rev. **17**, 1540014 (2015). https://doi.org/10.1142/S0219198915400149

41. Joosten, R.: Strong and weak rarity value: resource games with complex price-scarcity relationships. Dyn. Games Appl. **16**, 97–111 (2016)
42. Joosten, R., Brenner, T., Witt, U.: Games with frequency-dependent stage payoffs. Int. J. Game Theory **31**, 609–620 (2003)
43. Joosten, R., Samuel, L.: On stochastic fishery games with endogenous stage payoffs and transition probabilities. In: Proceedings of 3rd Joint Chinese-Dutch Workshop on Game Theory and Applications and 7th China Meeting on Game Theory and Applications. CCIS-series. Springer, Berlin (2017)
44. Joosten, R., Samuel, L.: On the computation of large sets of rewards in ETP-ESP-games with communicating states. Research memorandum, Twente University, The Netherlands (2017)
45. Joosten, R., Thuijsman, F., Peters, H.: Unlearning by not doing: repeated games with vanishing actions. Games Econ. Behav. **9**, 1–7 (1993)
46. Kelly, C.J., Codling, E.A., Rogan, E.: The Irish Sea cod recovery plan: some lessons learned. ICES J. Mar. Sci. **63**, 600–610 (2006)
47. Komorita, S.S., Parks, C.D.: Social Dilemmas. Westview Press, Boulder (1996)
48. Krawczyk, J.B., Tołwinski, B.: A cooperative solution for the three nation problem of exploitation of the southern bluefin tuna. IMA J. Math. Appl. Med. Biol. **10**, 135–147 (1993)
49. Lenton, T.M., Livina, V.N., Dakos, V., Scheffer, M.: Climate bifurcation during the last deglaciation? Clim. Past **8**, 1127–1139 (2012)
50. Levhari, D., Mirman, L.: The great fish war: an example using a dynamic Cournot-Nash solution. Bell J. Econ. **11**, 322–334 (1980)
51. Long, N.V.: A Survey of Dynamic Games in Economics. World Scientific, Singapore (2010)
52. Mäler, K.-G., Xepapadeas, A., de Zeeuw, A.: The economics of shallow lakes. Environ. Resour. Econ. **26**, 603–624 (2003)
53. Marwell, G., Oliver, P.: The Critical Mass in Collective Action: A Micro-Social Theory. Cambridge University Press, Cambridge (1993)
54. Messick, D.M., Brewer, M.B.: Solving social dilemmas: a review. Annu. Rev Pers. Soc. Psychol. **4**, 11–43 (1983)
55. Messick, D.M., Wilke, H., Brewer, M.B., Kramer, P.M., Zemke, P.E., Lui, L.: Individual adaptation and structural change as solutions to social dilemmas. J Pers. Soc. Psychol. **44**, 294–309 (1983)
56. Mertens, J.F., Neyman, A.: Stochastic games. Int. J. Game Theory. **10**, 53–66 (1981)
57. Oosthuizen, E., Daan, N.: Egg fecundity and maturity of North Sea cod, gadus morhua. Neth. J. Sea Res. **8**, 378–397 (1974)
58. Ostrom, E.: Governing the Commons. Cambridge University Press, Cambridge (1990)
59. Ostrom, E., Gardner, R., Walker, J.: Rules, Games, and Common-Pool Resources. Michigan University Press, Ann Arbor (1994)
60. Parthasarathy, T., Raghavan, T.E.S.: An orderfield property for stochastic games when one player controls the transition probabilities. J. Optim. Theory Appl. **33**, 375–392 (1981)
61. Platt, J.: Social traps. Am. Psychol. **28**, 641–651 (1973)
62. Raghavan, T.E.S., Filar, J.: Algorithms for stochastic games, a survey. Z. Oper. Res. **33**, 437–472 (1991)
63. Rose, G.A., Bradbury, I.R., de Young, B., Fudge, S.B., Lawson, G.L., Mello, L.G.S., Robichaud, D., Sherwood, G., Snelgrove, P.V.R., Windle, M.J.S.: Rebuilding Atlantic Cod: Lessons from a Spawning Ground in Coastal Newfoundland. In: Kruse, G.H., et al. (eds.) 24th Lowell Wakefield Fisheries Symposium on Resiliency of gadid stocks to fishing and climate change, pp. 197–219 (2008)
64. Sanchirico, J.N., Smith, M.D., Lipton, D.W.: An empirical approach to ecosystem-based fishery management. Ecol. Econ. **64**, 586–596 (2008)
65. Scheffer, M.: The Ecology of Shallow Lakes. Chapman & Hall, London (1998)
66. Scheffer, M., Carpenter, S., Foley, J.A., Folke, C., Walker, B.: Catastrophic shifts in ecosystems. Nature **413**, 591–596 (2001)
67. Schoenmakers, G.M.: The profit of skills in repeated and stochastic games. Ph.D. thesis Maastricht University (2004)

68. Schoenmakers, G.M., Flesch, J., Thuijsman, F.: Coordination games with vanishing actions. Int. Game Theory Rev. **4**, 119–126 (2002)
69. Shapley, L.: Stochastic games. Proc. Natl. Acad. Sci. USA **39**, 1095–1100 (1953)
70. Steg, L.: Motives and behavior in social dilemmas relevant to the environment. In: Hendrickx, L., Jager, W., Steg, L. (eds.) Human Decision Making and Environmental Perception. Understanding and Assisting Human Decision Making in Real-Life Settings, pp. 83–102 (2003)
71. Thuijsman, F., Vrieze, O.J.: The power of threats in stochastic games. In: Bardi, M., et al. (eds.) Stochastic and Differential Games, Theory and Numerical Solutions, pp. 343–358. Birkhauser, Boston (1998)
72. Tołwinski, B.: A concept of cooperative equilibrium for dynamic games. Automatica **18**, 431–441 (1982)
73. Tołwinski, B., Haurie, A., Leitmann, G.: Cooperative equilibria in differential games. JOTA **119**, 182–202 (1986)
74. Van Damme, E.E.C.: Stability and Perfection of Nash Equilibria. Springer, Berlin (1992)
75. Vrieze, O.J.: Linear programming and undiscounted games in which one player controls transitions. OR Spektrum **3**, 29–35 (1981)