MODELS, ALGORITHMS, AND DISTRIBUTIONAL ROBUSTNESS IN NASH GAMES AND RELATED PROBLEMS

by

ATSUSHI HORI

Submitted in partial fulfillment of
the requirement for the degree of
DOCTOR OF INFORMATICS
(Applied Mathematics and Physics)

KYOTO UNIVERSITY
KYOTO 606-8501, JAPAN
FEBRUARY 2023

Preface

Game theory has been employed across various domains of the economic discipline and has emerged as a fundamental methodology in the field. Furthermore, the game-theoretic constructs have been expanded to encompass diverse domains beyond economics, including computer science, political science, and psychology, up until the present time.

The game theory originated from the “Theory of Games and Economic Behavior” by J. von Neumann and O. Morgenstern in 1944, which aims to elucidate the behavior of economic agents using a rigorous mathematical methodology. The book mainly covers the theory of a cooperative game in which players can choose strategies based on agreements with other players. Meanwhile, in the 1950s J.F. Nash considered a *noncooperative game* in which all players simultaneously and independently try to maximize their profits without cooperation. An equilibrium point of the noncooperative game, referred to as *Nash equilibrium*, is a tuple of all players’ strategies where no one can improve the profit by changing their strategy unilaterally (other players keep their current strategies); we refer to the noncooperative game to find Nash equilibrium as *Nash game*.

Nash equilibrium has played a central role in describing player optimality in noncooperative game theory to date, and the equilibrium concept unifies other equilibrium concepts that appear in economics, such as Cournot competition, Stackelberg competition, and Bertrand competition. Besides, the condition for the Nash equilibrium technically coincides with the first-order optimality condition, such as a variational inequality, for each player’s payoff maximization problem. Against this background, Nash games have also been extensively studied in terms of continuous optimization theory.

Nevertheless, there is room for research on more complex decision-making situations in which each player makes a strategy under uncertainty, which is one of the motivations of this thesis to establish mathematical methodologies to find a ‘better’ strategy for each player under information uncertainty. In addition, we address Nash games with a bilevel structure referred to as a *multi-leader–follower game*, in which two or more of the players, called *leaders*, take actions first, and the rest of the players, called *followers*, take actions after observing the leaders’ decisions. This game has been much attention in recent years.

Variational inequality, another interest of this thesis, was first systematically studied in the 1960s by G. Stampacchia and his collaborators, who used the model to analyze a free boundary problem. The class of problem has also been extensively studied over the years and has many applications, such as game theory, engineering, traffic, physics, and so on. Although there is also a long history of research on variational inequalities, few studies have considered a variational inequality involving random vectors with an uncertain probability distribution. This thesis addresses this issue through an approach called *expected residual minimization*.

This thesis provides models, algorithms, and distributional robustness in Nash games and variational inequalities, and the contributions of this thesis are summarized as follows:

1. We propose an algorithm for finding a (stationary) equilibrium for the multi-leader–follower game and discuss the convergence of the algorithm. Then we report some numerical results to illustrate the behavior of the algorithm;

2. We consider a Nash game involving random vectors with uncertain probability distributions, in which each player makes two-stage decisions in response to changes in the conditions. We analyze this game from the perspective of distributionally robust optimization and demonstrate the existence of a Nash equilibrium under certain assumptions;

3. We address a variational inequality involving random vectors with an uncertain probability distribution and propose a distributionally robust expected residual minimization to find an approximate solution for the model. We also provide a numerically tractable reformulation for the minimization model to avoid numerical integration to evaluate the expected residual functions.

Atsushi Hori
February 2023, Kyoto, Japan

Acknowledgments

First, I would like to express my sincerest gratitude to Professor Nobuo Yamashita for his guidance and insightful advice. He kindly and patiently provided sophisticated yet easily understandable explanations to every single question I had. Additionally, I received valuable advice on my career path from him. Second, I am grateful to Assistant Professor Yuya Yamakawa for his valuable advice on my research and for his guidance in my career. He also taught me a wide range of subjects, from mathematics to the attitudes of a researcher, with great enthusiasm. Third, I am deeply thankful to Associate Professor Ellen Hidemi Fukuda. She was always friendly, and I enjoyed discussing various topics with her. I am also grateful to her for introducing me to Assistant Professor Jie Jiang of Chongqing University, which was an inspiring experience for me. I extend my sincere appreciation to Jie Jiang for his helpful and insightful comments on my study, which greatly improved the quality of my paper. I would also like to express my appreciation to Associate Professor Hiroyuki Sato for his practical advice and comments on my study during the workshop. Although we were unable to collaborate on research during my doctoral program, I learned a great deal from him, particularly in regard to his attitude as an educator. I aspire to emulate his efforts in fostering a welcoming and supportive environment for students seeking guidance. I am deeply grateful to Mr. Hardik Tankaria for the wide-ranging interactions we have had as friends and researchers. In addition, I am also thankful to Ms. Yuiko Igura, a secretary of the laboratory, for her support and management, as well as to the members of the System Optimization Laboratory.

My first official contact with Professor Nobuo Yamashita was through an introduction by Professor Emeritus Masao Fukushima of Kyoto University, my previous supervisor at Nanzan University, when we participated in a regular conference of the Operations Research Society of Japan in 2018. At that time, I was working as a software engineer for a company after completing my Master's degree at Nanzan University. I am greatly appreciative of Professor Masao Fukushima for taking the time to consult with me on my career. Moreover, he continued to encourage me via email during my doctoral program, and I always looked forward to sharing my completed papers with him.

I also frequently participated in an operations research seminar held online at Nanzan University. Professor Mihiro Sasaki of Nanzan University first invited me to join the seminar and give a talk, for which I am deeply grateful. I am also grateful to Professors Atsuo Suzuki, Hidetoshi Miura, and Shungo Koichi of Nanzan University, who also provided guidance and support during my time as a master's student at Nanzan University. I received insightful comments on my studies from them and enjoyed discussing case studies of operations research, which broadened my interests.

I would also like to express my appreciation for the emotional support provided by friends who visited me in Kyoto after I left Nagoya.

Furthermore, I am grateful for the funding provided by the Kyoto University Division of Graduate Studies and the JST SPRING (JPMJSP2110), as well as the Iwadare Scholarship Foundation for the scholarship that supported my living expenses, allowing me to fully dedicate myself to my studies.

I could not have completed this program without the support of all of these individuals and organizations.

Finally, I extend my heartfelt thanks to my family for their unwavering support.

Contents

List of Figures

List of Tables

Acronyms

The following acronyms are given in full when they appear for the first time in the text.

CQ	Constraint Qualification
DRERM	Distributionally Robust Expected Residual Minimization
DRNE	Distributionally Robust Nash Equilibrium
DRO	Distributionally Robust Optimization
EPEC	Equilibrium Problem with Equilibrium Constraints
ERM	Expected Residual Minimization
FB-function	Fischer–Burmeister function
KKT	Karush–Kuhn–Tucker
LCP	Linear Complementarity Problem
LICQ	Linear Independence Constraint Qualification
MFCQ	Mangasarian–Fromovitz Constraint Qualification
MPCC	Mathematical Program with Complementarity Constraints
MPEC	Mathematical Program with Equilibrium Constraints
NCP	Nonlinear Complementarity Problem
NSDP	Nonlinear Semidefinite Programming
PMPCC	Parametrized Mathematical Program with Complementarity Constraints
SIP	Semi-infinite Programming
SVI	Stochastic Variational Inequality
TSDRNE	Two-Stage Distributionally Robust Nash Equilibrium
TSDRVI	Two-Stage Distributionally Robust Variational Inequality
TSSVI	Two-Stage Stochastic Variational Inequality
VI	Variational Inequality
VIP	Variational Inequality Problem

Glossary of notation

Sets

$\in$, $\notin$	element membership, non-membership in a set
$\emptyset$, $\subset$, $\subsetneq$	empty set, set inclusion, proper set inclusion
$\cup$, $\cap$, $\times$	union, intersection, Cartesian product
ΠS_i	Cartesian product of sets S_i
$S_1 \setminus S_2$	difference of two sets S_1 and S_2
$S_1 + S_2$	$:= \{a + b \mid a \in S_1, b \in S_2\}$; Minkowski sum of two sets S_1 and S_2

Spaces

$\mathbb{R}^n$	real n-dimensional space
$\mathbb{R}^n_+$	nonnegative orthant of $\mathbb{R}^n$
$\mathbb{R}^n_{++}$	positive orthant of $\mathbb{R}^n$
$\mathbb{R}^{n\times m}$	space of $n \times m$ real matrices
$\mathbb{S}^n$	space of symmetric matrices in $\mathbb{R}^{n\times n}$
$\mathbb{S}^n_+$	cone of symmetric positive semidefinite matrices of order n
$\mathbb{S}^n_{++}$	cone of positive definite matrices in $\mathbb{S}^n$

Scalars

$\mathbb{R}$	real line
$[\cdot]_+$	$:= \max(0, \cdot)$; the nonnegative part of a scalar

Vectors

$x^\top$	$:= (x_1, \dots, x_n)$; transpose of a vector x with components x_i
$[x]_+$	$:= (\max(0, x_1), \dots, \max(0, x_n))$; componentwise nonnegative part of a vector x
$\{x^k\}$	sequence of vectors $x^1, x^2, \dots$
$\langle x, y\rangle$	$:= x^\top y = x_1y_1 + \cdots + x_ny_n$; standard inner product of vectors in $\mathbb{R}^n$
$\Vert x\Vert_p$	$:= (\sum_{i=1}^n \lvert x_i\rvert^p)^{1/p}$; ℓ_p-norm of a vector $x \in \mathbb{R}^n$
$\Vert x\Vert$	ℓ_2-norm of $x \in \mathbb{R}^n$, unless otherwise specified
$\Vert x\Vert_\infty$	$:= \max_{i=1,\dots,n} \lvert x_i\rvert$; ℓ_∞-norm of $x \in \mathbb{R}^n$
$x \geq y$	(usual) partial ordering $x_i \geq y_i$, $i = 1, \dots, n$
$x \circ y$	$:= (x_1y_1, \dots, x_ny_n)$; Hadamard product of x and y
$x \perp y$	x and y are perpendicular
$\mathbf{1}_n$	n-vector of all ones
e_i	$:= (0, \dots, 0, 1, 0, \dots, 0) \in \mathbb{R}^n$; a unit vector along the x_i-axis

Matrices

A	$:= (a_{ij})$; a matrix with entries a_{ij}
$\det A$	determinant of a matrix $A \in I\!R^{n\times n}$
$\operatorname{tr} A$	$:= \sum_{i=1}^{n} a_{ii}$; trace of a matrix $A \in I\!R^{n\times n}$
$A^\top$	transpose of a matrix A
A^{-1}	inverse of a matrix $A \in I\!R^{n\times n}$
$\langle A, B\rangle$	Frobenius product of two matrices
I_m	identity matrix of order m (subscript often omitted)
$\operatorname{diag}(a)$	diagonal matrix with diagonal elements equal to the components of the vector a

Functions

$f\colon D \to I\!R$	a mapping with domain D and range $I\!R$
$f'(\cdot;\cdot)$	directional derivative of the mapping f
∇f	$:= \left(\frac{\partial f}{\partial x_j}\right)_{j=1}^{n}$; gradient of a function $f\colon I\!R^n \to I\!R$
$\nabla_x f(x,y)$	partial gradient of $f\colon I\!R^{n+m} \to I\!R$ with respect to $x \in I\!R^n$
$\nabla^2 f$	$:= \left(\frac{\partial^2 f}{\partial x_i \partial x_j}\right)_{i,j=1,\dots,n}$; Hessian matrix of $f\colon I\!R^n \to I\!R$
$\partial_B f$	Bouligand subdifferential of $f\colon I\!R^n \to I\!R$
∂f	$:= \operatorname{co} \partial_B f$; Clarke subdifferential of $F\colon I\!R^n \to I\!R$
$\mathcal{J}F$	$:= \left(\dfrac{\partial F_i}{\partial x_j}\right)$; $m \times n$ Jacobian of a mapping $F\colon I\!R^n \to I\!R^m$ $(m \geq 2)$
$o(t)$	any function such that $\lim\limits_{t\downarrow 0} \dfrac{o(t)}{t} = 0$
$O(t)$	any function such that $\limsup\limits_{t\downarrow 0} \dfrac{\lvert O(t)\rvert}{t} < \infty$
$\mathbf{proj}_S\,[x]$	Euclidean projection of $x \in I\!R^n$ onto a convex set $S \subset I\!R^n$
$\operatorname{dist}(x, S)$	distance between a point $x \in I\!R^n$ and a set $S \subset I\!R^n$
$\inf_{x\in S} f(x)$	infimum of f on S
$\sup_{x\in S} f(x)$	supremum of f on S

List of publications

This thesis is based on the following articles:

(P1) Atsushi Hori and Masao Fukushima, Gauss–Seidel method for multi-leader–follower games, *Journal of Optimization Theory and Applications*, **180** (2019), 651–670;

(P2) Atsushi Hori and Nobuo Yamashita, Two-stage distributionally robust noncooperative games: Existence of Nash equilibrium and its application to Cournot–Nash competition, *Journal of Industrial and Management Optimization*, to appear;

(P3) Atsushi Hori, Yuya Yamakawa and Nobuo Yamashita, Distributionally robust expected residual minimization for stochastic variational inequality problem, *Optimization Methods and Software*, to appear.

- Chapters 3, 4, and 5 in the thesis are based on the papers (P1), (P2), and (P3), respectively;
- In Chapter 3, we slightly extended the result of (P1), and the updates are summarized as follows:
  - While the paper only discussed the convergence of the sequence generated by our proposed algorithm to a Bouligand stationary point (equivalently a strong stationary point), we further show the convergence to a Clarke stationary point and Mordukhovich stationary point;
  - In addition, to show the convergence to a Bouligand stationary point, the paper assumed that the points obtained by solving a penalized problem ($\bar{P}_\rho^\nu(x^{-\nu})$ in this thesis) are a locally optimal solution to $\bar{P}_\rho^\nu(x^{-\nu})$ for every iteration in the proposed algorithm, which is difficult to confirm in practice. This thesis shows the same convergence result under weaker assumptions;
  - We have added an application of a multi-leader–follower game to a wholesale electricity market in Section 3.5.

Chapter 1

Introduction

In this introductory chapter, we present the scope, motivations, and contributions of the thesis. To this end, we begin by providing a broad perspective on the key concepts and context, as well as a brief explanation of the problems being studied, to understand which areas this thesis is closely related. The key technical concepts behind the thesis are Nash games, variational inequalities, and distributionally robust optimization. We first describe a brief summary of each concept and its relations.

Nash game Nash game is a subset of noncooperative game theory and originated from J.F. Nash in the 1950s. Noncooperative game theory is vast in scope, and a variety of games have been devised. Our highest interest in this thesis is a *strategic form game*: Defined players, each player's strategy set, and payoff, all players competitively maximize their profits. The Nash game is a mathematical model to find an equilibrium point, referred to as *Nash equilibrium*, at which no player has an incentive to gain more payoff. Unless otherwise specified, the noncooperative game is often used as a synonym instead of the Nash game throughout this thesis.

Variational inequality Variational inequality (VI) is a mathematical model, its rigorous definition will be given later, and is regarded as a general class of continuous optimization in a certain sense. VI appears in many real-world applications such as physics, chemistry, engineering, traffic design, finance, game theory (our highest interests), and others. The relation with Nash games is that the condition of Nash equilibrium is equivalently reformulated as VI under suitable assumptions. Although we mainly deal with this class of problem involving random variables in Chapter 5, the concept of VI plays an important role in analyzing and establishing a numerical method for finding the equilibrium of Nash games; several variants of VI will appear as well in Chapters 3 and 4.

Distributionally robust optimization Distributionally robust optimization (DRO) is one of the modelling frameworks in optimization under uncertainty. The concept is "maximize (minimize) the *worst expected value* of payoff (disutility) from a set of probability measures," under which the exact distribution of random variables which appears in a stochastic optimization problem cannot exactly be estimated; for example, due to the lack of observation data. A subclass of DRO is stochastic optimization and (scenario-based) robust optimization. The former is the case in which the distribution of random variables is exactly known; that

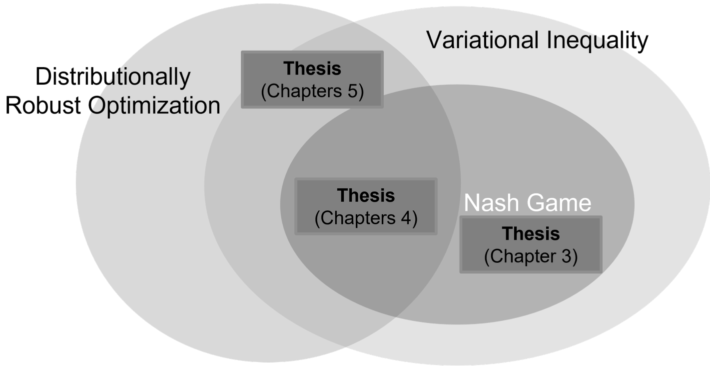


Figure 1.1: Structure of this thesis: This figure is a sketch to give a rough idea of the structure of this thesis. In the context of this thesis, Nash games can be analyzed in terms of variational inequalities, and hence the diagram is drawn so that the class of variational inequalities includes Nash games.

is, the set of probability measures is a singleton. The latter is 'distribution-free' optimization that "maximizes the *worst-case* payoff," and this corresponds to the case when the set of probability measures consists of all the measures satisfying the probability axioms[1]. Hence, the distributionally robust optimization allows us to consider more advanced uncertainty for each player's decision-making in Nash games.

Figure 1.1 depicts the relationship among the above technical concepts. Particularly, in Chapter 3 we focus on a bilevel-structured Nash game (see Figure 1.3), and in Chapter 4 we delve into the intersection of Nash games and distributionally robust optimization. In Chapter 5, we explore the intersection of distributionally robust optimization and variational inequalities. As we can see from the figure, the concept of Nash games and variational inequalities appears in Chapters 3, 4, and 5 and constitutes the core of this thesis. Henceforth, we introduce the overview of Nash games and variational inequalities in detail.

1.1 Overview of Nash games

Game theory is a branch of mathematics to analyze a strategic interaction among self-interested decision makers, say *players*. It has been widely applied in various fields, including economics, political science, psychology, computer science, and many others, along with social changes. The powerful mathematical tool helps us analyze and understand social structures that appear in our real lives; this is demonstrated by the history of game theory, which has consistently been developing since its inception without decline.

The general theory of games was first introduced by J. von Neumann and O. Morgenstern in their seminal book *Theory of Games and Economic Behavior* [87], just before the end of World War II. J.F. Nash [86] then considered an N-person noncooperative game of mixed strategies and established the equilibrium of the game, called *Nash equilibrium*, using

[1] This explanation is not strictly correct. In fact, if the uncertainty set consists of an uncountable set, such as an interval and ellipsoid, a continuous probability distribution of the "random variable" may be defined. However, the probability of taking only the worst-case scenario (a single point) is "0" in terms of measure theory. The explanation here is based on the case where probabilities are assigned to each scenario in an uncertainty set composed of finitely many scenarios, and any one of the scenarios has probability "1".

Kakutani's fixed point theorem [65]. An 'equilibrium' in this context means that all players have no incentive to change their strategies unilaterally to gain more profit. In other words, Nash equilibrium simultaneously achieves the global optimality for each player's optimization problem parametrized by the rivals' strategies. The solution concept derived from his work has had a profound impact on various fields and is still used in fields such as microeconomics, management science, and computer science [24, 26, 38, 43, 46, 69, 70, 73].

Game theory not only helps us understand the interactions between humans and organizations, but also leads to surprising discoveries through the modeling of strategic competition between artificial decision makers since the recent growth of *algorithmic game theory* [108] in computer science and machine learning. For example, adversarial models, which appear in generative adversarial networks [39], adversarial training [22], and multi-agent reinforcement learning [129], are some of the most successful implementations of game theory in recent years. In summary, it has made significant contributions as a prescriptive mathematical modeling paradigm and is expected to continue playing a role in analyzing more complex competitive situations in numerous fields.

In this thesis, we focus on finding a Nash equilibrium of the noncooperative game; we call such a game a *Nash game*. According to Li et al. [72], methods for finding an equilibrium of the Nash game are divided into two approaches. The first method is identifying some specific structures of the game, such as the supermodularity [121] or the property in which the profit of each player is characterized by a potential function, called a potential game [84]. The second approach is reformulating the game as a variational inequality [30, 101], which has been extensively studied and is efficient in finding the Nash equilibria of games with continuous decision variables in a unified way; to name a few, see [44, 52, 64, 69, 88, 100, 114]. To our best knowledge, the variational inequality reformulation approach was first proposed by Bensoussan [6].

Now we begin by introducing a mathematically fundamental model of the Nash game, and then we introduce its generalization models that are related to our thesis. Throughout the thesis, due to the consistency with the convention in optimization literature, we suppose that the payoff is replaced by $-1\times$payoff or a disutility, we refer to it as a *cost function*, and each player tries to 'minimize' the cost function subject to a strategy set. Note that this does not change the properties of the original model. Models of Nash games that we particularly study in the thesis are marked with $*$ at the beginning of the paragraph.

Nash game Suppose that there are N players who compete for minimizing the cost function $\theta_\nu\colon \mathbb{R}^n \to \mathbb{R}$ subject to the strategy (constraint) set $X^\nu \subset \mathbb{R}^{n_\nu}$, where $\nu \in \{1, \dots, N\}$ denotes a label to distinguish the player, and n denotes the sum of all players' dimensions for strategy vectors, i.e., $n \coloneqq n_1 + \cdots + n_N$. Let $x^\nu \in X^\nu \subset \mathbb{R}^{n_\nu}$ be a strategy vector of player ν. Consider that player ν solves the following optimization problem:

$$\min_{x^\nu \in \mathbb{R}^{n_\nu}} \theta_\nu(x^\nu, x^{-\nu}) \quad \text{s.t. } x^\nu \in X^\nu, \tag{1.1}$$

where $x^{-\nu} \coloneqq (x^1, \dots, x^{\nu-1}, x^{\nu+1}, \dots, x^N)$ denotes all strategy vector except x^ν. Throughout the thesis, we often write $x \coloneqq (x^1, \dots, x^N) \in \mathbb{R}^n$ as $(x^\nu, x^{-\nu})$ to emphasize the strategy of player ν. A *Nash equilibrium* is to find a tuple of strategies $x^* \coloneqq (x^{*,1}, \dots, x^{*,N})$ such that $x^{*,\nu}$ (globally) solves (1.1) for all $\nu \in \{1, \dots, N\}$; that is,

$$x^{*,\nu} \in \arg\min_{x^\nu \in X^\nu} \theta_\nu(x^\nu, x^{*,-\nu}) \quad \forall \nu \in \{1, \dots, N\}.$$

The *Nash game* is a mathematical model to find the Nash equilibrium in the noncooperative game.

Generalized Nash game While the classical formulation of the Nash game requires that the strategy set of each player is unaffected by rival players, in many real-world applications the strategy set of each player also depends on the other rival players' strategies; that is, the strategy set X^ν is replaced by $X^\nu(x^{-\nu})$ in (1.1). For example, the players share some common resources or limitations, such as an electrical transmission line or a common limit on the total resources for production [55]. Such a Nash game is referred to as a *generalized Nash game*, introduced by Arrow and Debreu [4], and has also been studied extensively, along with the development of a quasi-variational inequality [28,42,92]. In general, this class is more difficult to construct a numerical method to find the equilibrium. Hence, as we will see in Chapter 3, one of the techniques to overcome such difficulties is the reformulation of the generalized Nash games into a Nash game by utilizing a penalization approach [36,48,66].

Stackelberg game A *Stackelberg game* [115] is a bilevel-structured Nash game in which a player, referred to as a leader, takes action first, and then another player, referred to as a follower, takes action after observing the leader's action. Let $\theta\colon I\!R^{n+m} \to I\!R$ be a cost function of the leader and $\gamma\colon I\!R^{n+m} \to I\!R$ be that of the follower. The vectors $x \in I\!R^n$ and $y \in I\!R^n$ denote the leader and follower's strategies, respectively, and $X \subset I\!R^n$ and $Y(x) \subset I\!R^m$ denote their strategy sets, respectively. Note that the follower's strategy set $Y(x)$ depends on the leader's strategy x in general. The Stackelberg game can be formulated as the following bilevel optimization:

$$\begin{aligned} \min_{x\in X} \quad & \theta(x,y) \\ \text{s.t.} \quad & y \in \arg\min_{z\in Y(x)} \gamma(x,y). \end{aligned} \tag{1.2}$$

We say that $(x^*, y^*) \in X \times Y(x^*)$ is a *leader–follower equilibrium* if it solves (1.2). This model was originally proposed in microeconomics to analyze a market equilibrium between a large enterprise (leader) and a small firm (follower). However, in recent years this model has been much more active in machine learning, such as hyperparameter tuning and meta-learning [32–34, 51].

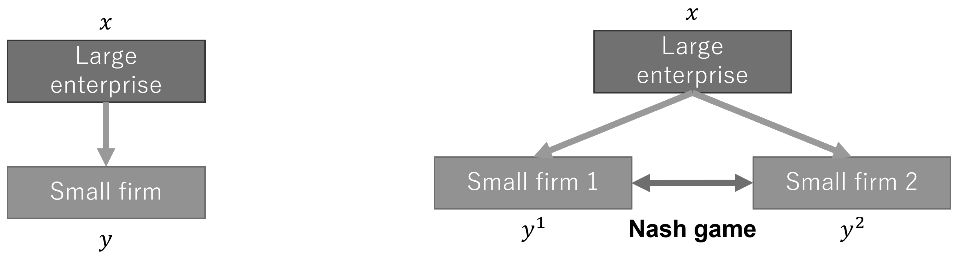
The game in which the number of followers is two or more is called a *single-leader–multi-follower game*; see Figure 1.2. In the context of this thesis, the single-leader–follower game is considered to be in the class of Stackelberg games in order to emphasize the difference from a multi-leader-follower game, which will be discussed later.

In general, for a given leader's strategy $x \in X$, the set of Nash equilibria for the lower-level (followers') game is not always a singleton. In such a case, the leader may estimate the potential responses of the followers whether they will make a beneficial or disadvantageous strategy to the leader: The former concept is called *optimistic*, and the latter is *pessimistic*. The *optimistic bilevel optimization* is formulated as

$$\min_{x\in X} \min_{y\in\mathcal{S}(x)} \theta(x,y),$$

where $\mathcal{S}(x)$ is the set of Nash equilibria in the followers' game, while the *pessimistic bilevel optimization* is formulated as

$$\min_{x\in X} \max_{y\in\mathcal{S}(x)} \theta(x,y).$$



(a) Stackelberg game (single-leader–single-follower) **(b)** Single-leader–multi-follower game

Figure 1.2: Structure of a Stackelberg game.

***Multi-leader–follower game (Chapter 3)** A *multi-leader-follower game* is a generalization of the Stackelberg and single-leader–follower games, involving two or more leaders and one or more followers. Consider a multi-leader–follower game consisting of N leaders and M followers. The leaders are labeled $\nu \in \{1, \ldots, N\}$, and the followers are labeled $\omega \in \{1, \ldots, M\}$. Let $x^\nu \in \mathbb{R}^{n_\nu}$, $X^\nu \subset \mathbb{R}^{n_\nu}$, and $\theta_\nu \colon \mathbb{R}^{n+m} \to \mathbb{R}$ denote the strategy vector, strategy set, and cost function of leader ν, respectively. Let $y^\omega \in \mathbb{R}^{m_\omega}$, $Y^\omega(x) \subset \mathbb{R}^{m_\omega}$, and $\gamma_\omega \colon \mathbb{R}^{n+m} \to \mathbb{R}$ denote the strategy vector, strategy set, and cost function of follower ω, respectively. Here, $n := n_1 + \cdots + n_N$ and $m := m_1 + \cdots + m_M$. For given $x^{-\nu} := (x^1, \ldots, x^{\nu-1}, x^{\nu+1}, \ldots, x^N) \in \mathbb{R}^{n-n_\nu}$ and $y := (y^1, \ldots, y^M) \in \mathbb{R}^m$, leader ν solves the following optimization problem:

$$\min_{x^\nu \in \mathbb{R}^{n_\nu}} \theta_\nu(x^\nu, x^{-\nu}, y) \quad \text{s.t. } x^\nu \in X^\nu. \tag{1.3}$$

For a given tuple of leaders' strategies $x := (x^1, \ldots, x^N) \in X := X^1 \times \cdots \times X^N \subset \mathbb{R}^n$ and the other followers' strategies $y^{-\omega} := (y^1, \ldots, y^{\omega-1}, y^{\omega+1}, \ldots, y^M) \in \mathbb{R}^{m-m_\omega}$, follower ω solves the following optimization problem:

$$\min_{y^\omega \in \mathbb{R}^{m_\omega}} \gamma_\omega(x, y^\omega, y^{-\omega}) \quad \text{s.t. } y^\omega \in Y^\omega(x). \tag{1.4}$$

Figure 1.3 shows an example of a multi-leader–follower game in which two large enterprises ($N = 2$) act as leaders, and two small firms ($M = 2$) act as followers.

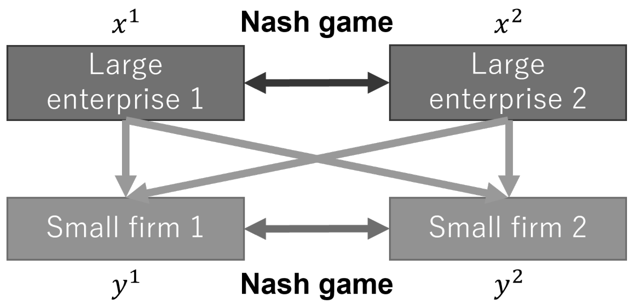


Figure 1.3: Structure of a multi-leader–follower game (two leaders and two followers).

Analogous to the discussion in the Stackelberg game, there may be non-unique solutions in the followers' Nash game. The optimistic and pessimistic formulation of the Stackelberg game can also be extended to the multi-leader–follower game; that is, each leader's problem is formulated as an optimistic or pessimistic bilevel optimization. In this thesis, we will focus on the case where the followers' response is unique, but potentially the model may be extended to an optimistic case.

As a real-world application, the multi-leader–follower game has been used to investigate the behaviors of producers (leaders) and regulators (followers) in the energy and telecommunication markets [24,47,57,92]. In Chapter 3, we will examine a wholesale electricity market, with energy companies serving as leaders and a market maker, known as an independent system operator (ISO), acting as the follower to regulate the market.

Now we introduce Nash games under uncertainty.

Stochastic Nash game In this thesis, the definition of a stochastic Nash game is twofold: an almost sure formulation and an expected value formulation. Let $\xi\colon \Omega \to \Xi$ be a random vector on the probability space $(\Omega, \mathcal{F}, P)$[2]. We say a stochastic Nash game in an *almost sure* (a.s.) formulation is that player ν solves the following optimization problem:

$$\min_{x^\nu \in \mathbb{R}^{n_\nu}} \quad \theta_\nu(x^\nu, x^{-\nu}, \xi) \quad \text{s.t. } x^\nu \in X^\nu.$$

We consider this game in Chapter 5. A stochastic Nash game in an *expected value* (EV) formulation is given by

$$\min_{x^\nu \in \mathbb{R}^{n_\nu}} \quad \mathbb{E}[\theta_\nu(x^\nu, x^{-\nu}, \xi)] \quad \text{s.t. } x^\nu \in X^\nu, \tag{1.5}$$

where $\mathbb{E}[f(\xi)]$ denotes an expected value of $f(\xi)$ with respect to the random variable ξ. Although we will not explicitly deal with the EV formulation, two-stage stochastic and distributionally robust Nash games, which will be considered in Chapter 4, are generalizations of this game in a certain sense. See Lei and Shanbhag [69] for a survey of the stochastic Nash game in the EV formulation. Note that we can further consider the case where X^ν also depends on ξ, but this is beyond the scope of this thesis.

The stochastic Nash game in the EV formulation imposes that each player should be risk-neutral, but players could be risk-averse. The risk-averse stochastic Nash games can be considered that the objective function of player ν is defined by a risk measure [78,93,98]; for example, the conditional value at risk [100,104] (at level α):

$$\mathsf{CV@R}_\alpha[f(\xi)] \coloneqq \min_{t \in \mathbb{R}} \left[t + \frac{1}{1-\alpha} \mathbb{E}[[f(\xi) - t]_+] \right],$$

where $[\cdot]_+ \coloneqq \max(\cdot, 0)$. Player ν solves the following minimization of the conditional value at risk of $\theta_\nu(x^\nu, x^{-\nu}, \xi)$:

$$\min_{x^\nu \in \mathbb{R}^{n_\nu}} \quad \mathsf{CV@R}_\alpha[\theta_\nu(x^\nu, x^{-\nu}, \xi)].$$

Figure 1.4 indicates a graphical illustration of an expected value and conditional value at risk.

[2] Note that the random variable is independent of the decisions by players.

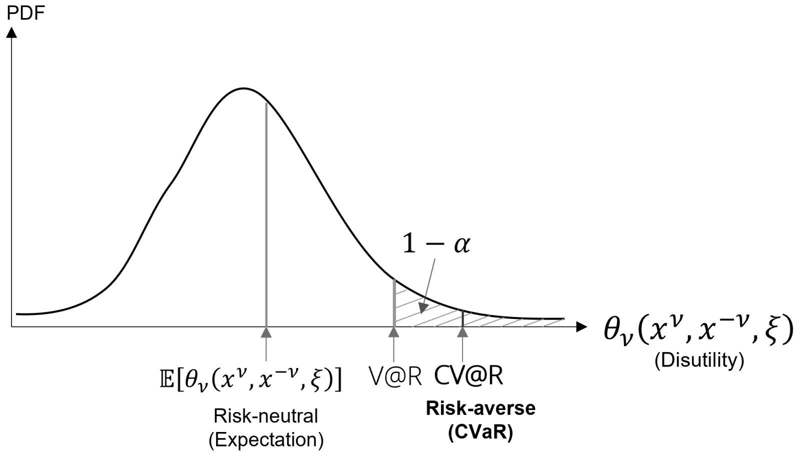


Figure 1.4: Graphical illustrations of expectation (risk-neutral) and conditional value at risk (CVaR, risk-averse).

Robust Nash game Aghassi and Bertsimas [2] have proposed a distribution-free model of incomplete-information games, with/without private information, in which the players use a *robust optimization* approach to contend with payoff uncertainty. In this model, a cost function θ_ν and/or a strategy set X^ν depends on an uncertainty parameter, and the distribution of the uncertainty parameter may be unknown. In a robust Nash game, player ν solves the following robust optimization problem:

$$\min_{x^\nu \in \mathbb{R}^{n_\nu}} \max_{\hat{u}^\nu \in \hat{\mathcal{U}}^\nu} \theta_\nu(x^\nu, x^{-\nu}, \hat{u}^\nu) \quad \text{s.t. } x^\nu \in X^\nu(\tilde{u}^\nu) \quad \forall \tilde{u}^\nu \in \tilde{\mathcal{U}}^\nu, \tag{1.6}$$

where $\hat{u}^\nu$ and $\tilde{u}^\nu$ are uncertainty parameters but the player knows that $\hat{u}^\nu$ and $\tilde{u}^\nu$ respectively belong to the sets $\hat{\mathcal{U}}^\nu$ and $\tilde{\mathcal{U}}^\nu$ of uncertainty.

The advantage of this model is that players do not need to know the probability distribution of uncertainty parameters, and also (1.6) can be recast as a second-order cone programming under suitable assumptions, which may allow us to reduce the computation time to obtain the equilibrium of the game. As an independent work of [2], Hayashi et al. [44] then considered the concept of robust Nash equilibria for bimatrix games to reformulate it to a second-order cone complementarity problem. Although [2] considered the mixed strategy games in which each player intrinsically solves a linear programming problem, Nishimura et al. [88] systematically analyzed a nonlinear case and presented sufficient conditions for the existence and uniqueness of equilibrium.

Distributionally robust stochastic Nash game Consider a stochastic Nash game in both the almost sure and expected value formulations. When the probability distribution of random vectors cannot exactly be identified due to the lack of sample data, and the observation may contain noise, players may not be able to estimate the expected value of the cost function in (1.5). In such a case, players may incur a significant loss when the true probability is different from what the players estimate. Meanwhile, as is well known in terms of robust optimization, an equilibrium obtained from robust Nash models may be too conservative because each player acts with the utmost importance on the tragedies that rarely occur probabilistically.

In response to those issues and taking advantage of the methodology of distributionally robust optimization, Sun and Xu [119], and Liu et al. [77] recently considered a Nash game in which player ν solves the following distributionally robust optimization:

$$\min_{x^\nu \in X^\nu} \max_{P \in \mathscr{P}^\nu} \mathbb{E}_P[\theta_\nu(x^\nu, x^{-\nu}, \xi)],$$

where $\mathscr{P}^\nu$ denotes an *ambiguity set*; that is, the set of probability distributions that may be constructed from observation data. The game may be regarded as a generalization of the stochastic and robust Nash games: When $\mathscr{P}^\nu$ is a singleton for all ν, this game corresponds to the stochastic Nash game in the EV form. Meanwhile, when Ξ is compact, and $\mathscr{P}^\nu$ is the set of all probability measures over the support Ξ, $\max_{P\in\mathscr{P}^\nu} \mathbb{E}_P[\theta_\nu(x^\nu, x^{-\nu}, \xi)] = \max_{\xi\in\Xi} \theta_\nu(x^\nu, x^{-\nu}, \xi)$, which corresponds to the formulation of robust Nash games. Liu et al. [77] further considered a hierarchical capacity competition problem in supply chain management as an application. This problem is characterized by a distributionally robust Stackelberg game in which either a leader (large enterprise) or followers (suppliers) solves distributionally robust optimization.

As is well known in the literature on distributionally robust optimization, the robust counterpart (duality form of the inner maximization) of each player's optimization problem is numerically tractable under suitable $\mathscr{P}^\nu$. For example, if $\mathscr{P}^\nu$ is constructed based on the class of measures called ϕ-divergence, the robust counterpart is known to be a convex conic optimization [97].

Two-stage stochastic Nash game In recent years, as two- and multistage stochastic variational inequalities have been extensively studied [16,17,59–62,103,107,120], a two-stage stochastic Nash game has also been studied [17,61,93,132]. In this game, each player takes actions in two stages: In the first stage, before a future event (scenario) occurs, each player solves a stochastic optimization by evaluating the expected cost for the second stage. In the second stage, after observing the realization, the player solves a 'deterministic' optimization problem for each scenario. Figure 1.5 shows an example of the two-stage stochastic Nash game in which two firms compete in the first (production) and second stages (supplying products).

Inherently, in the two-stage stochastic Nash game each player's optimization problem is given by a two-stage stochastic programming parameterized by other players' strategies. For a more detailed explanation of the formulation, see Section 4.2. Indeed, according to our knowledge, Haurie et al. [43] have already considered a similar model in the 1990s to analyze an oligopoly market. However, no systematical methodology has been conducted until recent years.

In the second stage at a scenario $\xi \in \Xi$ (Ξ is known in advance), player ν solves the optimization problem for a given first-stage strategy $x = (x^\nu, x^{-\nu})$:

$$Q_\nu(x^\nu, x^{-\nu}, \xi) := \min_{y^\nu(\xi) \in Y^\nu(x^\nu, \xi)} \gamma_\nu(y^\nu(\xi), y^{-\nu}(\xi), x^\nu, x^{-\nu}, \xi), \tag{1.7}$$

where $y^\nu(\xi)$, $\gamma_\nu(\cdot,\cdot,\cdot,\cdot,\xi)$, and $Y^\nu(\cdot,\xi)$ are the second-stage strategy vector, cost function, and strategy set at the scenario $\xi \in \Xi$, respectively. Here, $y^{-\nu}(\xi) := (y^1(\xi), \dots, y^{\nu-1}(\xi), y^{\nu+1}(\xi), \dots, y^N(\xi))$ denotes a tuple of the other rivals' strategies in the second stage at the scenario $\xi \in \Xi$. In the first stage, since all players do not know which scenario will realize in the

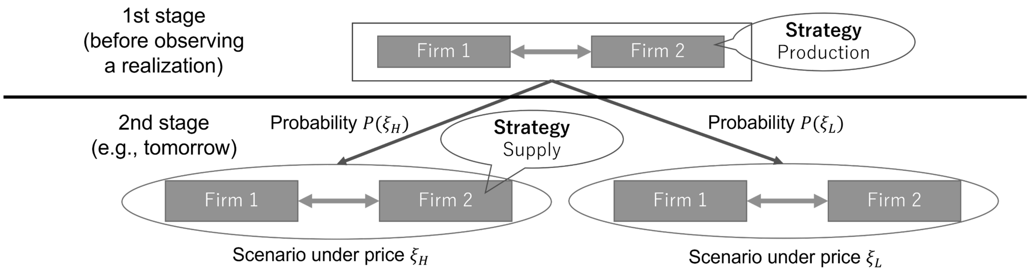


Figure 1.5: Structure of a two-stage stochastic Nash game (duopoly market): In this game, the random variable is a future demand, which is characterized by an inverse demand function (price). There are two scenarios: ξ_H and ξ_L represent prices at high and low, with their probabilities $P(\xi_H)$ and $P(\xi_L)$, respectively. In the first stage before a future demand is observed, players try to minimize the sum of the first stage cost ($\theta_\nu(x^\nu, x^{-\nu})$) and the expected value of the second stage cost ($\mathbb{E}[Q_\nu(x^\nu, x^{-\nu}, \xi)]$). Then after players observed either one of two scenarios, two firms play a (classical) Nash game under the given scenario (i.e., compute $Q_\nu(x^\nu, x^{-\nu}, \xi_L)$ and $Q_\nu(x^\nu, x^{-\nu}, \xi_H)$).

future, they evaluate the expected value of $Q_\nu(x^\nu, x^{-\nu}, \xi)$, and player ν solves the following stochastic optimization problem:

$$\min_{x^\nu \in X^\nu} \quad \{\theta_\nu(x^\nu, x^{-\nu}) + \mathbb{E}[Q_\nu(x^\nu, x^{-\nu}, \xi)]\},$$

where θ_ν is a first-stage cost function.

***Two-stage distributionally robust stochastic Nash game (Chapter 4)** In two-stage stochastic Nash games, the probability distribution of the random variable in the second stage is known in advance; that is, in Figure 1.5 all players know the probability distribution of the future demand. Then all players can evaluate the future expected cost. However, as we have mentioned in the (first-stage) distributionally robust stochastic Nash game, the probability distribution of random variables is often unknown in real-world applications.

To tackle this issue, a two-stage distributionally robust Nash game has been considered by Li et al. [72] more recently. Player ν solves the following distributionally robust optimization in the first stage:

$$\min_{x^\nu \in X^\nu} \quad \{\theta_\nu(x^\nu, x^{-\nu}) + \max_{P \in \mathscr{P}^\nu} \mathbb{E}_P[Q_\nu(x^\nu, x^{-\nu}, \xi)]\}.$$

Note that since all players have already observed a scenario in the second stage, the formulation of the second-stage problem is the same as (1.7). When $\mathscr{P}^\nu$ is a singleton for all ν, the game is reduced to the two-stage stochastic Nash game. Similar to Jiang et al. [61] in the case of the two-stage stochastic Nash game, Chen et al. [17], and Hori and Yamashita [50] considered a distributionally robust Cournot–Nash competition as an application of the game. The former analyzed the competition from the perspective of ex-post (distribution-free) equilibrium, which intrinsically corresponds to the concept proposed by Aghassi and Bertsimas [2], and the latter first established a concept of an equilibrium explicitly in the sense of Nash equilibrium.

1.2 Overview of variational inequality

The subject of variational inequalities originates in the calculus of variations associated with the minimization of infinite-dimensional functionals. The systematic study of variational inequalities began in the early 1960s with the seminal work of the Italian mathematician G. Stampacchia and his collaborators, who used the variational inequality as an analytic tool for studying free boundary problems defined by nonlinear partial differential operators arising from unilateral problems in elasticity and plasticity theory and in mechanics [30]. Since variational inequalities have been studied so extensively over the years, it is difficult even to summarize them here. The reader refers to a two-volume monograph by Facchinei and Pang [30] for a detailed history of variational inequalities and research trends.

Given a certain convex set S and a vector-valued function $F\colon \mathbb{R}^n \to \mathbb{R}^n$, a *variational inequality* (VI) is to find $x^* \in S$ such that

$$\langle F(x^*), x - x^* \rangle \geq 0 \qquad \forall x \in S. \tag{1.8}$$

By introducing a normal cone of S at $x \in S$ defined as

$$\mathcal{N}_S(x) := \{z \in \mathbb{R}^n \mid \langle z, y - x\rangle \leq 0 \quad \forall y \in S\},$$

VI (1.8) can also be denoted as the generalized equation:

$$0 \in F(x^*) + \mathcal{N}_S(x^*).$$

When $S = \mathbb{R}^n_+ := \{x \in \mathbb{R}^n \mid x_i \geq 0,\ i = 1, \ldots, n\}$, VI (1.8) is reduced to a *complementarity problem* (CP): Find x^* such that

$$0 \leq x^* \perp F(x^*) \geq 0.$$

This class also plays an important role in optimization theory, engineering, traffic, and so on. We will see the detailed mathematical properties of the VI and CP in Chapter 2. More specifically, when $S = \mathbb{R}^n$, VI (1.8) corresponds to a nonlinear equation $F(x^*) = 0$.

Hereafter, to identify the scope of this thesis, we begin by introducing the relationship between VI and Nash games, and then we introduce some classes of stochastic variational inequalities that appeared in Chapters 4 and 5.

Relation between VI and Nash games Consider the Nash game where each player solves (1.1). When the strategy set X^ν is closed convex, and the cost function $\theta_\nu(\cdot, x^{-\nu})$ is convex for any fixed $x^{-\nu}$, $x^* = (x^{*,1}, \ldots, x^{*,N})$ is the Nash equilibrium if and only if x^* solves variational inequality (1.8) by setting

$$F(x) = \begin{bmatrix} \nabla_{x^1}\theta_1(x^1, x^{-1}) \\ \vdots \\ \nabla_{x^N}\theta_N(x^N, x^{-N}) \end{bmatrix}, \quad S = X^1 \times \cdots \times X^N.$$

Because of such affinity between VI and Nash games in terms of formulation, the analysis of Nash games is often attributed to VI [6, 64].

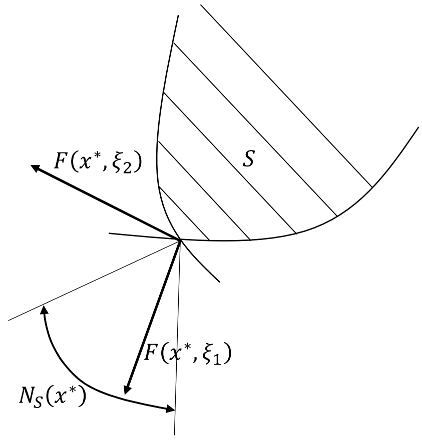


Figure 1.6: Stochastic variational inequality in the almost sure formulation: A point x^* satisfies the inequality in (1.9) when $\xi = \xi_1$, but not when $\xi = \xi_2$.

Stochastic variational inequality For a random vector ξ, a *stochastic variational inequality* (SVI) in an *almost sure* (a.s.) formulation is defined by

$$\langle F(x^*, \xi), x - x^* \rangle \geq 0 \qquad \forall x \in S, \quad \text{for } \xi \in \Xi, \text{ almost surely.} \tag{1.9}$$

SVI in an *expected value* (EV) formulation is defined as

$$\langle \mathbb{E}[F(x^*, \xi)], x - x^* \rangle \geq 0 \qquad \forall x \in S. \tag{1.10}$$

The stochastic Nash games in the almost sure and expected value formulations defined in the previous section can also be recast as the SVI in each formulation, respectively, under a certain assumption.

In general, (1.9) has no solution that satisfies the inequality for almost all $\xi \in \Xi$, as depicted in Figure 1.6. Therefore, SVI in the almost sure formulation is often considered as an *expected residual minimization* [14, 18, 19, 49, 80, 81]: Let $f\colon S \times \Xi \to I\!R_+$ be a merit function for the SVI, i.e., for a fixed $\xi \in \Xi$, $f(x, \xi) = 0$ if x satisfies the inequality in (1.9) and $f(x, \xi) > 0$ if x does not satisfy the inequality. The expected residual minimization model is to minimize the expected value of the merit function:

$$\min_{x \in I\!R^n} \mathbb{E}[f(x, \xi)] \qquad \text{s.t. } x \in S. \tag{1.11}$$

Chen and Fukushima [14] compared the solutions to ERM (1.11) and EV (1.10) in the class of stochastic linear complementarity problems, and then they found that if the variance of the random variable ξ is not small, a solution to (1.10) may violate the complementarity conditions more than a solution to (1.11) for many realizations of the random variables. Then Chen et al. [19] quantitatively analyzed the robustness of the solutions to (1.11) and (1.10) for monotone stochastic linear complementarity problems.

Two- and multistage stochastic variational inequality As an extension of 'single-stage' SVI in the EV form (1.10), Rockafellar and Wets [107] first explicitly proposed a multistage SVI. In particular the formulation of a two-stage SVI is given as follows: Find $(x^*, y^*(\cdot)) \in X \times \mathcal{Y}$ such that

$$\begin{aligned}
&0 \in \mathbb{E}[F(x^*, y^*(\xi), \xi)] + \mathcal{N}_X(x^*), \\
&0 \in G(x^*, y^*(\xi), \xi) + \mathcal{N}_{Y(\xi)}(y^*(\xi)), \quad \text{almost every } \xi \in \Xi,
\end{aligned}$$

where $F\colon I\!R^n \times I\!R^m \times \Xi \to I\!R^n$, $G\colon I\!R^n \times I\!R^m \times \Xi \to I\!R^m$, $X \subset I\!R^n$, $Y: \Xi \to I\!R^m$, and $\mathcal{Y} := \{y(\cdot) \mid y(\cdot) \in Y(\cdot)\}$. Then Rockafellar and Sun [103] proposed a progressive hedging algorithm to solve the multistage SVI with discrete random variables. Particularly, the two-stage SVI has been studied mainly to date. Chen et al. [15] proposed an expected residual minimization for the two-stage SVI. Chen et al. [17] proposed a discrete approximation method for the two-stage stochastic linear complementarity problem with continuously underlying random variables, and Jiang et al. [62] considered its extension to the multistage SVI. Chen et al. [16] proposed a sample average approximation method for a two-stage stochastic generalized equation. See a survey on the two-stage SVI by Sun and Chen [120] for more detailed research trends.

Distributionally robust variational inequality As another extension of SVI in the EV formulation (1.10), Sun et al. [118] recently proposed a distributionally robust VI: Find $(x^*, P^*) \in S \times \mathscr{P}$ such that

$$\begin{aligned}
&\langle \mathbb{E}_{P^*}[F(x^*, \xi)], x - x^* \rangle \geq 0 \qquad \forall x \in S, \\
&P^* \in \arg\max_{Q \in \mathscr{P}} \mathbb{E}_Q[f(x^*, \xi)],
\end{aligned} \tag{1.12}$$

where $f\colon S \times \Xi \to I\!R$ and $\mathscr{P}$ is an ambiguity set. One of the examples of the model is that the optimality condition for a distributionally robust optimization

$$\min_{x \in S} \max_{P \in \mathscr{P}} \mathbb{E}_P[f(x, \xi)],$$

is written as (1.12) by setting $F := \nabla_x f(x, \xi)$ under certain assumptions. In [118], they considered a case with continuously underlying random vectors in (1.12) and proposed a discrete approximation method for a monotone distributionally robust VI. They showed that if a discrete approximated problem of (1.12) can successfully be solved, the solution converges to the solution to (1.12) as the number of samples increases. However, the convergence of the numerical algorithm to solve the discretized distributionally robust VI is still open. Hori and Yamashita [50] introduced a two-stage distributionally robust VI to reformulate the two-stage distributionally robust Nash game with discrete random variables.

1.3 Research issues and contributions of the thesis

The fundamental goal of this thesis is to generalize the concepts of previous studies to more advanced decision-making situations in deterministic/stochastic Nash games and variational inequalities. Particularly, we address the following research issues:

(RI1) Algorithms for multi-leader–follower games with inequality constraints in followers' optimization problems have not been established: When each follower's optimization problem (1.4) contains inequality constraints, the multi-leader–follower game becomes much more difficult because the complementarity condition appears. However, there are only a handful of studies that systematically address this case in multi-leader–follower games, and more concrete clarification is needed;

(RI2) Few studies on stochastic Nash games and stochastic variational inequalities when the probability distribution of random vectors is uncertain: Recent developments in distributionally robust optimization have led to a move toward extensions to various models. However, few studies have considered distributionally robust stochastic Nash games [72, 77] and variational inequalities [118].

In response to the above research issues, the contributions to the above research issues are summarized as follows:

(C1) In Chapter 3, we address (RI1) in a nonlinear multi-leader–follower game. By the well-known reformulation approach for the bilevel optimization, we first reformulate the multi-leader–follower game to an equilibrium problem with equilibrium constraints. Then we propose a Gauss–Seidel type algorithm with a penalization technique for solving the equilibrium problem with equilibrium constraints. Using this approach, the resultant problem is a standard Nash game, and we can solve it with an off-the-shelf nonlinear solver. We discuss the convergence of the algorithm and report some numerical results to illustrate the behavior of the algorithm. We also suggest a refinement procedure to obtain more accurate solutions. Finally, we consider an application of the multi-leader–follower game, a wholesale electricity market. This game consists of electricity firms (leaders) and a market maker (follower). The leaders determine how much they sell electricity to each demand node (consumer), and the follower corrects the balance of demand and supply of electricity by paying the bid costs.

(C2) In Chapter 4, to tackle (RI2) in stochastic Nash games, we consider the two-stage distributionally robust stochastic Nash game. Existing studies on this model have been limited to strict assumptions, such as linear decision rules, and supposes that each player solves a two-stage linear distributionally robust optimization with a specifically structured ambiguity set. This motivated us to generalize and analyze the game in a nonlinear case. The contributions of this study are (i) demonstrating the conditions for the existence of two-stage Nash equilibria under convexity and compactness assumptions, and (ii) consideration of a two-stage distributionally robust Cournot–Nash competition as an application, as well as an investigation into the conditions for the existence of market equilibria in an economic sense. We also report some results of numerical experiments to illustrate how distributional robustness affects the decision of each player in the Cournot–Nash competition.

(C3) In Chapter 5, we address (RI2) in stochastic variational inequalities in the almost sure formulation (1.9) when the exact probability distribution of random variables ξ may not be known. We propose a distributionally robust expected residual minimization for the stochastic variational inequalities:

$$\min_{x\in \mathbb{R}^n} \max_{P\in\mathscr{P}} \mathbb{E}_P[f(x,\xi)] \quad \text{s.t. } x \in S. \tag{1.13}$$

This approach may be regarded as the extension of ERM method (1.11) [14, 18, 80, 81]. In general, solving (1.13) is quite computationally expensive because it includes the computation of the expected value (integration) and the maximization with respect to the probability. However, under suitable assumptions we demonstrate that the distributionally robust expected residual minimization can be reformulated as a deterministic nonlinear semidefinite programming problem to avoid numerical integration. We also show a sufficient condition that the deterministic optimization problem is convex. Finally, we conduct some numerical experiments to verify the robustness against the perturbation of probability distributions by comparing with the existing ERM method.

Chapter 2

Preliminaries

This chapter provides some basic mathematical concepts used throughout the thesis.

2.1 Mathematical terms and notations

Vectors Let $I\!R^n$ be the n-dimensional Euclidean space. A vector $x \in I\!R^n$ is denoted as

$$x \coloneqq \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} = (x_1, \ldots, x_n)^\top = (x_i)_{i=1}^n.$$

where the superscript $^\top$ denotes the transpose operation. For simplicity, we often omit $\top$ to represent column vectors. We use the following notations:

$$e_i \coloneqq (0, \ldots, 0, 1, 0, \ldots, 0) \in I\!R^n,$$
$$\mathbf{1} \coloneqq (1, 1, \ldots, 1) \in I\!R^n,$$

where $e_i \in I\!R^n$ represents a unit vector along the x_i-axis.

For any vectors $x \in I\!R^n$ and $y \in I\!R^n$, the inner product $\langle x, y\rangle$ or $x^\top y$ is defined by

$$\langle x, y\rangle = x^\top y \coloneqq x_1 y_1 + \cdots + x_n y_n.$$

The norms $\|x\|_1$, $\|x\|_2$, and $\|x\|_\infty$ of $x \in I\!R^n$ are defined as follows:

$$\|x\|_1 \coloneqq |x_1| + \cdots + |x_n|, \ \|x\|_2 \coloneqq \sqrt{\langle x, x\rangle}, \ \|x\|_\infty \coloneqq \max_{i=1,\ldots,n} |x_i|.$$

In particular, let $\|\cdot\|$ denote $\|\cdot\|_2$ unless otherwise specified.

We say that a finite list of vectors $x^1, \ldots, x^k$ in a vector space V is *linearly independent* if and only if for scalars $\alpha_1, \ldots, \alpha_k \in I\!R$, the equation $\alpha_1 x^1 + \cdots + \alpha_k x^k = 0$ can only be satisfied by $\alpha_i = 0$ for $i = 1, \ldots, k$. A finite list of vectors $x^1, \ldots, x^k$ in a vector space V is *linearly dependent* if and only if it is not linearly independent.

Matrices Let $I\!R^{m\times n}$ be the $m \times n$-dimensional Euclidean space. A matrix $A \in I\!R^{m\times n}$ is denoted by

$$A := \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{bmatrix}.$$

We say that the matrix $A \in I\!R^{n\times n}$ is *symmetric* if and only if $A^\top = A$, and the space of n-dimensional symmetric matrices denotes $\mathbb{S}^n$.

A symmetric matrix $A \in \mathbb{S}^n$ is *positive semidefinite* if and only if $x^\top Ax \geq 0$ for all $x \in I\!R^n$, and the space of symmetric positive semidefinite matrices is denoted by $\mathbb{S}^n_+$. A symmetric matrix $A \in \mathbb{S}^n$ is *positive definite* if and only if $x^\top Ax > 0$ for all $x \in I\!R^n \setminus \{0\}$, and the space of symmetric positive definite matrices is denoted by $\mathbb{S}^n_{++}$. For a matrix $A \in \mathbb{S}^n$, $A \in \mathbb{S}^n_+$ and $A \in \mathbb{S}^n_{++}$ are often denoted by $A \succeq O$ and $A \succ O$, respectively. For two matrices $A \in \mathbb{S}^n$ and $B \in \mathbb{S}^n$, $A \succeq (\succ) B$ means $A - B \succeq (\succ) O$. We say a symmetric matrix $A \in \mathbb{S}^n$ is *negative (semi)definite* if and only if $-A$ is positive (semi)definite, and we write $A \preceq (\prec) O$ to indicate that A is negative (semi)definite.

For any matrices $X \in \mathbb{S}^m$ and $Y \in \mathbb{S}^m$, the matrix inner product $\langle X, Y\rangle$ is defined by

$$\langle X, Y\rangle := \mathrm{tr}(XY) = \sum_{i,j=1}^{m} X_{ij}Y_{ij}.$$

Differentiablility Let $h\colon I\!R^n \to (-\infty, +\infty]$ be finite-valued in a certain neighborhood of $x \in I\!R^n$. If h has the partial derivative

$$\frac{\partial h(x)}{\partial x_i} := \lim_{\delta\to 0} \frac{h(x+\delta e_i) - h(x)}{\delta}, \quad i = 1, \dots, n,$$

and if

$$h(x+\varepsilon) = h(x) + \langle \nabla h(x), \varepsilon\rangle + o(\|\varepsilon\|) \quad \forall \varepsilon \in I\!R^n$$

with $o\colon [0, +\infty) \to I\!R$ satisfying $\lim_{\delta\to 0} o(\delta)/\delta = 0$ and

$$\nabla h(x) := \begin{bmatrix} \frac{\partial h(x)}{\partial x_1} \\ \vdots \\ \frac{\partial h(x)}{\partial x_n} \end{bmatrix},$$

then h is *differentiable* at x, where $\nabla h(x) \in I\!R^n$ is called a *gradient* of h at x. If $\nabla h(x)$ is continuous at x, we say that h is *continuously differentiable* at x. Likewise if h has second-order derivatives and

$$h(x+\varepsilon) = h(x) + \langle \nabla h(x), \varepsilon\rangle + \frac{1}{2}\langle \varepsilon, \nabla^2 h(x)\varepsilon\rangle + o(\|\varepsilon\|^2)$$

with

$$\nabla^2 h(x) := \begin{bmatrix} \frac{\partial^2 h(x)}{\partial x_1 \partial x_1} & \cdots & \frac{\partial^2 h(x)}{\partial x_1 \partial x_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial^2 h(x)}{\partial x_n \partial x_1} & \cdots & \frac{\partial^2 h(x)}{\partial x_n \partial x_n} \end{bmatrix},$$

then h is *twice differentiable* at x, and $\nabla^2 h(x)$ is referred to as the *Hessian* of h at x. When $\nabla^2 h(x)$ is continuous at x, we say h is *twice continuously differentiable* at x, and $\nabla^2 h(x)$ is symmetric. For a vector-value function $h\colon \mathbb{R}^n \to \mathbb{R}^m$ with $h := (h_1, \dots, h_m)^\top$, $\mathcal{J}h(x)$ denotes the *Jacobian matrix* of h at x; that is,

$$\mathcal{J}h(x) := \begin{bmatrix} \frac{\partial h_1(x)}{\partial x_1} & \cdots & \frac{\partial h_1(x)}{\partial x_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial h_m(x)}{\partial x_1} & \cdots & \frac{\partial h_m(x)}{\partial x_n} \end{bmatrix} = [\nabla h_1(x), \dots, \nabla h_m(x)]^\top \in \mathbb{R}^{m\times n}.$$

Henceforth, we denote $\mathcal{J}h(x)^\top$, transposed Jacobian matrix, as $\nabla h(x)$ for convenience, and we say it is the Jacobian matrix of h at x.

First- and second-order directional derivative Let f be any function from $\mathbb{R}^n$ to $\mathbb{R}$, and let x be a point where f is finite. The *first-order directional derivative* of f at x in the direction d is defined to be the limit

$$f'(x; d) := \lim_{\delta \downarrow 0} \frac{f(x + \delta d) - f(x)}{\delta},$$

if it exists. If f is continuously differentiable, $f'(x; d) = \langle \nabla f(x), d\rangle$. Suppose that $f'(x; d)$ exists. The *second-order directional derivative* of f at x in the directions d and v is defined to be the limit

$$f''(x; d, v) := \lim_{\delta \downarrow 0} \frac{f(x + \delta d + \delta^2 v) - f(x) - \delta f'(x; d)}{\delta^2},$$

whenever this limit exists [9]. If f is twice continuously differentiable, then

$$f''(x; d, v) = \langle \nabla f(x), v\rangle + \frac{1}{2}\langle d, \nabla^2 f(x) d\rangle.$$

For a function $f\colon \mathbb{R}^n \to \mathbb{R}$, an *effective domain* $\operatorname{dom} f$ of f is defined as

$$\operatorname{dom} f := \{x \mid f(x) < +\infty\}.$$

In this thesis, any functions are assumed to be *proper* unless otherwise specified; that is, for a given function $f\colon \mathbb{R}^n \to [-\infty, +\infty]$, $f(x) > -\infty$ for all x, and $\operatorname{dom} f \neq \emptyset$. A *graph* $\operatorname{gph}\Phi$ of a set-valued function $\Phi\colon \mathbb{R}^n \rightrightarrows \mathbb{R}^m$ is defined by

$$\operatorname{gph}\Phi := \{(x, y) \in \mathbb{R}^n \times \mathbb{R}^m \mid y \in \Phi(x)\}.$$

Definition 2.1.1 (convex set)**.** *A set $S \subset \mathbb{R}^n$ is said to be* convex *if and only if*

$$(1 - \lambda)x + \lambda y \in S \quad \forall x, y \in S,\ \lambda \in (0, 1).$$

Definition 2.1.2 (cone)**.** *A set $C \subset \mathbb{R}^n$ is called a cone if*

$$x \in C,\ \alpha \in [0, \infty) \implies \alpha x \in C.$$

Definition 2.1.3 (polar cone)**.** *Given a nonempty set $C \subset \mathbb{R}^n$, the polar cone C^* of C is given by*

$$C^* := \{y \in \mathbb{R}^n \mid \langle y, x\rangle \leq 0 \quad \forall x \in C\}.$$

Definition 2.1.4 (tangent cone)**.** *Given a nonempty set $S \subset I\!R^n$, a* tangent cone *of S at $x \in S$ is given by*

$$\mathcal{T}_S(x) := \left\{ y \in I\!R^n \;\middle|\; y = \lim_{k\to\infty} \alpha_k(x^k - x),\ \lim_{k\to\infty} x^k = x,\ x^k \in S,\ \alpha_k \geq 0,\ k = 1, 2, \dots \right\}.$$

Definition 2.1.5 (normal cone)**.** *Given a nonempty set $S \subset I\!R^n$, a* normal cone *of S at $x \in S$ is given by the polar cone of the tangent cone $\mathcal{T}_S(x)$; that is,*

$$\mathcal{N}_S(x) := \mathcal{T}_S^*(x) = \{z \in I\!R^n \mid \langle z, y\rangle \leq 0 \quad \forall y \in \mathcal{T}_S(x)\}.$$

Particularly, when $S \subset I\!R^n$ is convex, the normal cone $\mathcal{N}_S$ is equivalent to

$$\mathcal{N}_S(x) = \{z \in I\!R^n \mid \langle z, y - x\rangle \leq 0 \quad \forall y \in S\}.$$

Definition 2.1.6 (convex function)**.** *Let $C \subset \operatorname{dom} f$ be a convex set. A function $f\colon C \to (-\infty, +\infty]$ is said to be*

(a) convex *if and only if*

$$f((1-\lambda)x + \lambda y) \leq (1-\lambda)f(x) + \lambda f(y) \quad \forall x, y \in C, \lambda \in (0,1);$$

(b) strictly convex *if and only if*

$$f((1-\lambda)x + \lambda y) < (1-\lambda)f(x) + \lambda f(y) \quad \forall x, y \in C, \lambda \in (0,1);$$

(c) strongly convex *with modulus $\sigma > 0$ if and only if*

$$f((1-\lambda)x + \lambda y) \leq (1-\lambda)f(x) + \lambda f(y) - \frac{\sigma}{2}(1-\lambda)\lambda\|x - y\|^2 \quad \forall x, y \in C, \lambda \in (0,1).$$

It is obvious that [strongly convex $\Rightarrow$ strictly convex $\Rightarrow$ convex]. In addition, we say that a function f is *concave* if $-f$ is convex. If a convex function is (twice) differentiable, we have the following properties.

Theorem 2.1.7 (e.g., Boyd and Vandenberghe [13, Section 3.1])**.** *Let $C \subset \operatorname{dom} f$ be an open convex set. Suppose that $f\colon C \to I\!R$ is differentiable on C. Then a function f is convex on C if and only if*

$$f(y) \geq f(x) + \langle \nabla f(x), y - x\rangle \quad \forall x, y \in C.$$

Moreover, the function f is strictly convex on C if and only if the above inequality is strict whenever $x \neq y$.

Theorem 2.1.8 (e.g., Boyd and Vandenberghe [13, Section 3.1])**.** *Let $C \subset \operatorname{dom} f$ be an open convex set. Suppose that $f\colon C \to I\!R$ is twice continuously differentiable on C. Then the function f is convex if and only if*

$$\nabla^2 f(x) \succeq O \quad \forall x \in C,$$

Moreover, if $\nabla^2 f(x)$ is positive definite for all $x \in C$, then f is strictly convex[1].

[1] The converse is not true.

The (orthogonal) projection operator $\mathbf{proj}_S\,[x]$ from $x \in I\!R^n$ onto a convex set $S \subset I\!R^n$ is defined as follows:

$$\mathbf{proj}_S\,[x] := \arg\min_{z\in S} \|z - x\|,$$

which has the following properties.

Proposition 2.1.9 (e.g., Facchinei and Pang [30, Theorem 1.5.5]). *For each $x \in I\!R^n$, $\mathbf{proj}_S\,[\cdot]$ satisfies*

$$\langle x - \mathbf{proj}_S\,[x], y - \mathbf{proj}_S\,[x]\rangle \le 0 \quad \forall y \in S.$$

Proposition 2.1.10 (e.g., Facchinei and Pang [30, Theorem 1.5.5]). *The projection operator is nonexpansive; i.e.,*

$$\|\mathbf{proj}_S\,[x] - \mathbf{proj}_S\,[y]\| \le \|x - y\| \quad \forall x, y \in I\!R^n.$$

Next the monotonicity of a set-valued function is given as follows.

Definition 2.1.11 (monotone mapping). *A set-valued function $F\colon I\!R^n \rightrightarrows I\!R^n$ is said to be*

(a) monotone *if*

$$x, y \in I\!R^n, \xi \in F(x), \eta \in F(y) \implies \langle \xi - \eta, x - y\rangle \ge 0;$$

(b) strictly monotone *if*

$$x, y \in I\!R^n, x \ne y, \xi \in F(x), \eta \in F(y) \implies \langle \xi - \eta, x - y\rangle > 0;$$

(c) strongly monotone *with modulus $\sigma > 0$, or σ-strongly monotone, if*

$$x, y \in I\!R^n, \xi \in F(x), \eta \in F(y) \implies \langle \xi - \eta, x - y\rangle \ge \sigma\|x - y\|^2.$$

Among the above monotonicity properties, the following relations hold: [strongly monotone $\Rightarrow$ strictly monotone $\Rightarrow$ monotone].

Definition 2.1.12 (maximal monotone mapping). *A monotone map $\Phi\colon I\!R^n \rightrightarrows I\!R^n$ is* maximal monotone *if no enlargement of its graph is possible in $I\!R^n \times I\!R^n$ without destroying monotonicity; that is, no monotone map Ψ exists such that* $\operatorname{gph}\Phi \subset \operatorname{gph}\Psi$.

The word 'maximal' is derived from the maximal maps with respect to set value inclusion. Another rephrasing of maximal monotonicity is the following: A monotone map Φ is maximal monotone if and only if every solution $(y, \eta) \in I\!R^n \times I\!R^n$ of the system of inequalities

$$\langle \eta - \xi, y - x\rangle \ge 0 \quad \forall (x, \xi) \in \operatorname{gph}\Phi,$$

belongs to $\operatorname{gph}\Phi$.

Proposition 2.1.13 (Facchinei and Pang [30, Proposition 2.3.2]). *Let $F\colon D \to I\!R^n$ be continuously differentiable on the open convex set $D \subset I\!R^n$. The following statements are valid:*

(a) *F is monotone on D if and only if the Jacobian matrix $\nabla F(x)$ is positive semidefinite for all x in D;*

(b) *F is strictly monotone on D if $\nabla F(x)$ is positive definite for all x in D;*

(c) *F is strongly monotone on D with modulus $\sigma > 0$ if and only if $\nabla F(x)$ is uniformly positive definite for all x in D; that is,*

$$d^\top \nabla F(x) d \geq \sigma \|d\|^2 \quad \forall d \in I\!R^n,$$

for all $x \in D$.

We next introduce a subgradient and subdifferential of a convex real-valued function.

Definition 2.1.14 (subgradient)**.** *Let $f\colon I\!R^n \to I\!R$ be convex and let $x \in \operatorname{dom} f$. A vector $\xi \in I\!R^n$ is called a* subgradient *of f at x if*

$$f(y) \geq f(x) + \langle \xi, y - x\rangle \quad \forall y \in \operatorname{dom} f.$$

The set of all subgradients of f at x is called the *subdifferential* of f at x, denoted by $\partial f(x)$. When $x \notin \operatorname{dom} f$, we define $\partial f(x) = \emptyset$.

Definition 2.1.15 (locally Lipschitz continuous function)**.** *For a given function $f\colon I\!R^n \to I\!R$ and a bounded set $E \subset I\!R^n$, f is* locally Lipschitz continuous *if there exists $L_E > 0$ such that*

$$|f(x) - f(y)| \leq L_E \|x - y\| \quad \forall x, y \in E.$$

Particularly, when L_E is independent of E, f is said to be *(globally) Lipschitz continuous.* It is known that continuously differentiable functions and finite-valued convex functions are locally Lipschitz continuous. A locally Lipschitz continuous function is differentiable almost everywhere in the sense of Lebesgue measure from Rademacher's theorem. Using this property, for a locally Lipschitz continuous function $f\colon I\!R^n \to I\!R$, the subdifferential of f at x is given by

$$\partial f(x) = \operatorname{conv}\left\{ \lim_{k\to\infty} \nabla f(x^k) \;\middle|\; \lim_{k\to\infty} x^k = x,\ \{x^k\} \subset \mathcal{D}_f \right\} \subset I\!R^n, \tag{2.1}$$

where $\mathcal{D}_f \subset I\!R^n$ is the set of points at which f is differentiable, and conv denotes the convex hull of a given set [21]. Similar to (2.1), the subdifferential of a vector-valued function $F\colon I\!R^n \to I\!R^m$ is obtained as follows:

$$\partial F(x) = \operatorname{conv}\left\{ \lim_{k\to\infty} \nabla F(x^k) \;\middle|\; \lim_{k\to\infty} x^k = x,\ \{x^k\} \subset \mathcal{D}_F \right\} \subset I\!R^{n\times m},$$

where $\mathcal{D}_F \subset I\!R^n$ is the set of points at which F is differentiable. We say that an element of $\partial F(x)$ is a *generalized Jacobian matrix* of F at x.

Lemma 2.1.16. *For a locally Lipschitz continuous function $F\colon I\!R^n \to I\!R^m$, we have*

$$\partial F(x) \subset [\partial F_1(x) \ldots \partial F_m(x)],$$

where the right-hand side denotes the set of matrices on $I\!R^{n\times m}$ whose ith column is a vector in $\partial F_i(x) \subset I\!R^n$.

Now we introduce an optimality condition for the following problem:

$$\min_{x\in \mathbb{R}^n} \quad f(x) \qquad \text{s.t. } x \in S, \tag{2.2}$$

where $f\colon \mathbb{R}^n \to \mathbb{R}$ is locally Lipschitz continuous, and $S \subset \mathbb{R}^n$ is nonempty. For problem (2.2), a vector x is said to be a *feasible solution* if $x \in S$; the set S is called a *feasible set* of problem (2.2).

Definition 2.1.17 (optimality). *A feasible solution $x^* \in S$ is said to be*

(a) *a* globally optimal solution *of problem* (2.2) *if*

$$f(x^*) \le f(x) \quad \forall x \in S;$$

(b) *a* locally optimal solution *of problem* (2.2) *if there exists $\delta > 0$ such that*

$$f(x^*) \le f(x) \quad \forall x \in \{x \in \mathbb{R}^n \mid \|x - x^*\| \le \delta\} \cap S.$$

The first-order necessary condition for optimality in (2.2) is given as follows.

Theorem 2.1.18 (e.g., Clarke [21, Corollary of Proposition 2.4.3]). *Consider problem* (2.2)*. If a feasible solution $x^* \in S$ is a locally optimal solution, then it satisfies*

$$0 \in \partial f(x^*) + \mathcal{N}_S(x^*). \tag{2.3}$$

Moreover, if (2.2) *is a convex optimization, i.e., f and S are convex, the converse is true.*

In general, a point satisfying (2.3) is called a *stationary point* of problem (2.2).

In the remainder of this section, we provide another first-order necessary condition for optimality. Suppose that the feasible set $S \subset \mathbb{R}^n$ is defined by the collection of equality and inequality constraints:

$$S := \{x \in \mathbb{R}^n \mid g(x) \le 0,\ h(x) = 0\},$$

where $g\colon \mathbb{R}^n \to \mathbb{R}^l$ and $h\colon \mathbb{R}^n \to \mathbb{R}^m$ are continuously differentiable on $\mathbb{R}^n$. Suppose also that f is continuously differentiable.

Definition 2.1.19 (Karush–Kuhn–Tucker point). *A point $x^* \in S$ is called a* Karush–Kuhn–Tucker (KKT) point*, along with a Lagrange multiplier $(\lambda^*, \mu^*) \in \mathbb{R}^l_+ \times \mathbb{R}^m$ if they satisfy*

$$\begin{aligned} &\nabla f(x^*) + \nabla g(x^*)\lambda^* + \nabla h(x^*)\mu^* = 0, \\ &\lambda_i^* \ge 0,\ g_i(x^*) \le 0,\ \lambda_i^* g_i(x^*) = 0,\ i = 1, \dots, l, \\ &h(x^*) = 0. \end{aligned} \tag{2.4}$$

The following definition indicates so-called *constraint qualifications* to guarantee that the KKT condition is the first-order necessary condition for optimality.

Definition 2.1.20 (constraint qualification). *We say that*

(a) *the* linear independence constraint qualification *(LICQ) holds at x if vectors $\nabla g_i(x)$, $i \in \mathcal{I}(x)$ and $\nabla h_j(x)$, $j = 1, \dots, m$ are linearly independent, where $\mathcal{I}(x) := \{i \mid g_i(x) = 0,\ i = 1, \dots, l\}$.*

(b) *the* Slater's constraint qualification *holds if* g_i *is convex for all* $i \in \{1, \ldots, l\}$*,* h_j *is affine for all* $j \in \{1, \ldots, m\}$*, and there exists* $x^0 \in \mathbb{R}^n$ *such that* $g_i(x^0) < 0$*,* $i = 1, \ldots, l$ *and* $h_j(x^0) = 0$*,* $j = 1, \ldots, m$*.*

Theorem 2.1.21. *Let* $x^* \in \mathbb{R}^n$ *be a locally optimal solution of* (2.2)*. Then if either of the LICQ* (a) *or the Slater's CQ* (b) *in Definition* 2.1.20 *holds, there exists* $(\lambda^*, \mu^*) \in \mathbb{R}^l_+ \times \mathbb{R}^m$ *such that the KKT condition* (2.4) *holds. Moreover, if the LICQ holds at* $x^* \in \mathbb{R}^n$*, then the Lagrange multiplier* (λ^*, μ^*) *satisfying* (2.4) *is unique.*

2.2 Noncooperative games and their equilibria

Consider that there are N self-interested decision makers, and $\nu \in \{1, \ldots, N\}$ denotes the label of the player. Let $x^\nu \in \mathbb{R}^{n_\nu}$, $X^\nu \subset \mathbb{R}^{n_\nu}$, and $\theta_\nu \colon \mathbb{R}^n \to \mathbb{R}$ be a strategy vector, strategy set, and cost function of player ν, respectively, where $n := n_1 + \cdots + n_N$ denotes the sum of dimensions of all players' strategy vectors. Suppose that player ν solves the following optimization problem:

$$\min_{x \in \mathbb{R}^{n_\nu}} \theta_\nu(x^\nu, x^{-\nu}) \qquad \text{s.t. } x^\nu \in X^\nu, \tag{2.5}$$

where $x^{-\nu} := (x^1, \ldots, x^{\nu-1}, x^{\nu+1}, \ldots, x^N) \in \mathbb{R}^{n-n_\nu}$ denotes a tuple of strategies of rival players. In order to emphasize the strategy x^ν of player ν, we often write $x := (x^1, \ldots, x^N) \in \mathbb{R}^n$ as $(x^\nu, x^{-\nu})$.

Definition 2.2.1 (Nash equilibrium)**.** *A* Nash equilibrium $x^* := (x^{*,1}, \ldots, x^{*,N}) \in X := X^1 \times \cdots \times X^N$ *solves* (2.5) *for all* $\nu \in \{1, \ldots, N\}$*; that is,*

$$x^{*,\nu} \in \arg \min_{x^\nu \in X^\nu} \theta_\nu(x^\nu, x^{*,-\nu}) \quad \forall \nu \in \{1, \ldots, N\}.$$

Definition 2.2.1 indicates that no player has an incentive to change the strategy unilaterally.

In the same manner as Rosen's proof [109] in a Nash game with a common strategy set among all players, we show the existence of Nash equilibrium in which player ν solves (2.5).

We define the following function:

$$\psi(x, z) := \sum_{\nu=1}^{N} \theta_\nu(z^\nu, x^{-\nu}).$$

Lemma 2.2.2. *A point* $x^* \in X$ *is a Nash equilibrium if and only if* $x^* \in \arg\min_{z \in X} \psi(x^*, z)$*.*

Proof. First we show the "only if" part. Suppose that x^* is a Nash equilibrium. Then, by the definition, we have

$$\theta_\nu(x^{*,\nu}, x^{*,-\nu}) \le \theta_\nu(z^\nu, x^{*,-\nu}) \qquad \forall z^\nu \in X^\nu,\ \forall \nu \in \{1, \ldots, N\}.$$

Summing up with respect to ν from 1 to N on both sides of the inequality leads that x^* must be a globally optimal solution of the minimization of $\psi(x^*, \cdot)$ over X.

Next we show the "if" part by contradiction. Suppose that x^* is not a Nash equilibrium. There exists a player $\nu' \in \{1, \ldots, N\}$ such that they can reduce the cost unilaterally; that is, there exists $\bar{x}^{\nu'} \in X^{\nu'}$ such that

$$\theta_{\nu'}(\bar{x}^{\nu'}, x^{*,-\nu'}) < \theta_{\nu'}(x^{*,\nu'}, x^{*,-\nu'}),$$

which contradicts that x^* is a globally optimal solution of the minimization of $\psi(x^*, \cdot)$. We have completed the proof. $\square$

Proposition 2.2.3. *Suppose that for all ν, $X^\nu \subset I\!R^{n_\nu}$ is compact convex on $I\!R^{n_\nu}$. Suppose also that $\theta_\nu \colon I\!R^n \to I\!R$ is continuous on $I\!R^n$, and for each fixed $x^{-\nu}$, the function $\theta_\nu(\cdot, x^{-\nu})$ is convex in x^ν. Then there exists a Nash equilibrium $x^* \in X$.*

Proof. By Lemma 2.2.2, x^* is a Nash equilibrium if and only if $x^* \in \arg\min_{z\in X} \psi(x, z)$. Then, for a defined mapping $\Gamma \colon X \rightrightarrows X$, $x \mapsto \arg\min_{z\in X} \psi(x, z)$, it suffices to show the existence of a fixed point of Γ. Since θ_ν is continuous, so is $\psi(x, \cdot)$. By [106, Theorem 2.6, Theory 7.41], Γ is upper semi-continuous that maps each point of the convex compact set X into a closed convex subset of X. Therefore, Kakutani's fixed point theorem [65, Theorem 1] ensures that there exists $x^* \in X$ such that $x^* \in \Gamma(x^*)$. This completes the proof. $\square$

In order to analyze the Nash game and to obtain an equilibrium solution, variational inequalities and complementarity problems are essential mathematical tools. The relationship between Nash games and variational inequalities will be presented later in the next section.

2.3 Variational inequality and complementarity problem

First, we introduce a mathematical definition of variational inequalities and complementarity problems. Then we will describe the relationship between Nash games and variational inequalities.

The definitions of variational inequality and complementarity problems are given as follows.

Definition 2.3.1 (variational inequality)**.** *Given a convex set $S \subset I\!R^n$ and vector-valued function $F \colon I\!R^n \to I\!R^n$, a* variational inequality*, denoted as* VI(S, F)*, is to find $x^* \in S$ such that*

$$\langle F(x^*), x - x^* \rangle \geq 0 \qquad \forall x \in S.$$

By the definition of the normal cone for a convex set, the variational inequality can also be denoted as

$$0 \in F(x^*) + \mathcal{N}_S(x^*).$$

When problem (2.2) is convex, and f is differentiable, the necessary and sufficient condition for optimality corresponds to VI$(S, \nabla f)$.

Definition 2.3.2 (complementarity problem)**.** *For a given vector-valued function $F \colon I\!R^n \to I\!R^n$, a* complementarity problem*, denoted as* CP(F)*, is to find x such that*

$$0 \leq x \perp F(x) \geq 0,$$

where $0 \leq x \perp y \geq 0$ denotes $x \geq 0$, $y \geq 0$, and $x^\top F(x) = 0$.

When $S = I\!R^n_+$, VI(S, F) is reduced to the complementarity problem. More specifically, when $S = I\!R^n$, VI(S, F) is reduced to a nonlinear equation $F(x) = 0$.

Next we describe the reformulation of VI(S, F) and CP(F) into a nonlinear equation (that is not necessarily smooth). Let us define the following natural mapping:

$$F_{\mathrm{nat}}(x) \coloneqq x - \mathbf{proj}_S\,[x - F(x)],$$

where $\mathbf{proj}_S\,[z]$ denotes the projection of $z \in I\!R^n$ onto the convex set $S \subset I\!R^n$. $\mathrm{VI}(S, F)$ is equivalently reformulated as a nonlinear equation.

Theorem 2.3.3 (Facchinei and Pang [30, Proposition 1.5.8]). *A point $x \in I\!R^n$ satisfies $F_{\mathrm{nat}}(x) = 0$ if and only if x is the solution to* $\mathrm{VI}(S, F)$.

Moreover, for a positive constant $\alpha > 0$, let us define

$$H_\alpha(x) := \mathbf{proj}_S\,[x - \alpha^{-1}F(x)].$$

Then the same assertion as Theorem 2.3.3 on $x - H_\alpha(x)$ holds for any $\alpha > 0$ [35].

Next let us introduce the reformulation of $\mathrm{CP}(F)$ into another nonlinear equation.

Definition 2.3.4 (C-function). *A function $\phi\colon I\!R^2 \to I\!R$ satisfying the following properties is referred to as a* C-function[2]*: For any pair $(a, b) \in I\!R^2$,*

$$\phi(a, b) = 0 \iff 0 \le a \perp b \ge 0;$$

equivalently, ϕ is a C-function if the set of its zeros is the two nonnegative semiaxes.

For any given C-functions ϕ, $\mathrm{CP}(F)$ is equivalent to a system of equations:

$$\Phi(x) := \begin{bmatrix} \phi(x_1, F_1(x)) \\ \vdots \\ \phi(x_n, F_n(x)) \end{bmatrix} = 0. \tag{2.6}$$

There are two well-known C-functions: the min-function

$$\phi_{\min}(a, b) := \min(a, b),$$

and the Fischier–Burmeister (FB) function

$$\phi_{\mathrm{FB}}(a, b) := a + b - \sqrt{a^2 + b^2}.$$

It is known that the FB-function is not differentiable at which $(a, b) = (0, 0)$ but the squared FB-function is continuously differentiable everywhere as we will see in Proposition 2.4.3. Meanwhile, $\phi_{\min}(a, b)$ is not differentiable at which $a = b$. Unless otherwise noted, we omit FB for ϕ_{FB} since we mainly use this C-function throughout the thesis.

The generalized Jacobian matrix of Φ is given as follows.

Proposition 2.3.5 (Facchinei and Pang [30, Proposition 9.1.4 (a)]). *Assume that $F\colon \Omega \to I\!R^n$ is continuously differentiable on the open set $\Omega \subset I\!R^n$. The generalized Jacobian matrix of Φ satisfies*

$$\partial\Phi(x) \subset \mathcal{D}_a(x) + \nabla F(x)\mathcal{D}_b(x),$$

where $\mathcal{D}_a(x)$ and $\mathcal{D}_b(x)$ are the sets of $n \times n$ diagonal matrices $\mathrm{diag}(a_1(x), \ldots, a_n(x))$ *and* $\mathrm{diag}(b_1(x), \ldots, b_n(x))$ *respectively, with*

$$(a_i(x), b_i(x)) \begin{cases} = \left(1 - \dfrac{x_i}{\sqrt{x_i^2 + F_i^2(x)}}, 1 - \dfrac{F_i(x)}{\sqrt{x_i^2 + F_i^2(x)}}\right) & \text{if } (x_i, F_i(x)) \neq (0, 0), \\ \in \{(1 - \xi, 1 - \eta) \mid \xi^2 + \eta^2 \le 1\} & \text{if } (x_i, F_i(x)) = (0, 0). \end{cases}$$

[2]'C-' stands for 'complementarity.'

For C^2 functions $G, H\colon \mathbb{R}^n \to \mathbb{R}$, if $(G(z), H(z)) \neq (0,0)$, $\phi(G(z), H(z))$ is twice continuously differentiable with respect to $z \in \mathbb{R}^n$. By direct calculation, its gradient and Hessian are respectively given by

$$\nabla_z \phi(G(z), H(z)) = \left(1 - \frac{G(z)}{\sqrt{G(z)^2 + H(z)^2}}\right) \nabla G(z) + \left(1 - \frac{H(z)}{\sqrt{G(z)^2 + H(z)^2}}\right) \nabla H(z), \tag{2.7}$$

and

$$\begin{aligned}
&\nabla_z^2 \phi(G(z), H(z)) = \\
&\quad \left(1 - \frac{G(z)}{\sqrt{G(z)^2 + H(z)^2}}\right) \nabla^2 G(z) + \left(1 - \frac{H(z)}{\sqrt{G(z)^2 + H(z)^2}}\right) \nabla^2 H(z) \\
&\quad - \frac{1}{\sqrt{G(z)^2 + H(z)^2}^3} [H(z)\nabla G(z) - G(z)\nabla H(z)][H(z)\nabla G(z) - G(z)\nabla H(z)]^\top.
\end{aligned} \tag{2.8}$$

Now we introduce a basic property associated with the existence of a solution to the variational inequality.

Theorem 2.3.6 (Facchinei and Pang [30, Theorem 2.3.3])**.** *Let $S \subset \mathbb{R}^n$ be closed convex and $F\colon S \to \mathbb{R}^n$ be continuous.*

(a) *If F is strictly monotone on S, $\mathrm{VI}(S, F)$ has at most one solution.*

(b) *If F is σ-strongly monotone, $\mathrm{VI}(S, F)$ has a unique solution.*

(c) *If F is defined, Lipschitz continuous, and σ-strongly monotone on a set $\Omega \supset S$, then there exists a constant $c' > 0$ such that for every vector $x \in \Omega$,*

$$\|x - x^*\| \le c' \|F_{\mathrm{nat}}(x)\|,$$

where x^* is the unique solution to $\mathrm{VI}(S, F)$.

For linear constrained variational inequalities, i.e., for $\mathrm{VI}(S, F)$, S is given by a polyhedral set over $\mathbb{R}^n$, it is well-known that if F is strongly monotone and Lipschitz continuous, $\|x - \mathbf{proj}_S\,[x - F(x)]\|$ is a *global error bound* for the solution to $\mathrm{VI}(S, F)$, which plays an important role in analyzing the convergence rate of iterate algorithms [91]. Here, a function $r\colon \mathbb{R}^n \to \mathbb{R}$ is a global error bound for a variational inequality problem $\mathrm{VI}(S, F)$ if the following condition holds: there exists a constant $c > 0$ such that

$$\mathrm{dist}(x, \mathrm{SOL}(S, F)) \le c r(x) \quad \forall x \in \mathbb{R}^n,$$

where $\mathrm{SOL}(S, F) \subset \mathbb{R}^n$ denotes the solution set of $\mathrm{VI}(S, F)$, and $\mathrm{dist}(x, \mathrm{SOL}(S, F))$ denotes the distance between the point $x \in \mathbb{R}^n$ and the set $\mathrm{SOL}(S, F)$.

The rest of this section describes the relationship between Nash games and variational inequalities.

Recall the N-person Nash game in which player $\nu \in \{1, \dots, N\}$ solves (2.5), introduced in Section 2.2. The following assertion is an essential connection between Nash games and variational inequalities.

Proposition 2.3.7 (Facchinei and Pang [30, Proposition 1.4.2]). *For all ν, let each X^ν be closed convex on $\mathbb{R}^{n_\nu}$. Suppose that for each fixed $x^{-\nu}$, the function $\theta_\nu(\cdot, x^{-\nu})$ is convex and continuously differentiable. Then a tuple x^* is a Nash equilibrium if and only if x^* belongs to the solution set of* VI(X, F)*, where*

$$X := X^1 \times \cdots \times X^N, \quad F(x) := \begin{bmatrix} \nabla_{x^1}\theta_1(x^1, x^{-1}) \\ \vdots \\ \nabla_{x^N}\theta_N(x^N, x^{-N}) \end{bmatrix}.$$

Utilizing this property, Nash equilibrium can be computed via algorithms for VI [30] by reformulating the game into VI.

The uniqueness of Nash equilibrium is shown as follows in the context of variational inequalities.

Theorem 2.3.8. *Suppose that the assumption of Proposition* 2.3.7 *holds. Assume that either of the following statements holds:*

(a) *X^ν is compact, and the mapping $F\colon \mathbb{R}^n \to \mathbb{R}^n$ is strictly monotone, i.e., Definition* 2.1.11–(b) *holds.*

(b) *$F\colon \mathbb{R}^n \to \mathbb{R}^n$ is strongly monotone, i.e., Definition* 2.1.11–(c) *holds.*

Then VI(X, F) *has a unique solution, and hence the Nash equilibrium uniquely exists.*

Proof. If (a) holds, then the Nash equilibrium exists by Proposition 2.2.3. It follows from Proposition 2.3.7 that VI(X, F) has a solution. Theorem 2.3.6–(a) ensures that the solution to VI(X, F) is at most one. Therefore, the Nash equilibrium uniquely exists.

If (b) holds, then VI(X, F) has a unique solution by Theorem 2.3.6–(b); hence, the solution coincides with a unique Nash equilibrium. □

When problem (2.5) is nonconvex, the existence of Nash equilibrium is not guaranteed in general, and it is difficult to confirm whether a point obtained via an algorithm for VI is a Nash equilibrium.

Definition 2.3.9 (stationary Nash equilibrium). *Consider the Nash game in which player ν solves* (2.5). *Then the solution to* VI(X, F) *is said to be a* stationary Nash equilibrium, *where X and F are defined in Proposition* 2.3.7.

Obviously, if x^* is a Nash equilibrium, then x^* is a stationary Nash equilibrium, and by Proposition 2.3.7, the converse is true when (2.5) is convex in x^ν for all ν.

Note that even if for every player, $\theta_\nu(x^\nu, x^{-\nu})$ is strongly convex in x^ν for any fixed $x^{-\nu}$, the mapping F defined in Proposition 2.3.7 is not strongly monotone in general. Moreover, if player ν solves a convex optimization problem in x^ν, the Nash game does not necessarily admit any Nash equilibrium as indicated in the following example.

Example 2.3.1 (Lei and Shanbhag [69]). *Consider a two-person Nash game: Player 1 solves*

$$\min_{x^1 \in \mathbb{R}} \quad -x^1 + x^2 \qquad \text{s.t. } x^1 \geq 0,$$

and player 2 solves

$$\min_{x^2 \in \mathbb{R}} \quad \frac{1}{2}(x^2)^2 + x^2 \qquad \text{s.t. } x^2 \geq 0.$$

By Proposition 2.3.7*, this game can be compactly stated as the following complementarity problem:*

$$0 \le \begin{bmatrix} x^1 \\ x^2 \end{bmatrix} \perp \begin{bmatrix} 0 & 1 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} x^1 \\ x^2 \end{bmatrix} + \begin{bmatrix} -1 \\ 1 \end{bmatrix} \ge 0.$$

However, the complementarity problem does not have a solution, and hence the Nash game does not admit any Nash equilibrium.

2.4 Merit functions

Merit functions for VI/CP have also been extensively studied for long years. Roughly speaking, a merit function is a real-valued function to measure the distance between a point on $I\!R^n$ and a solution set to VI.

A function $f\colon I\!R^n \to I\!R$ is the merit function to VI if the following properties hold:

1. $f(x) = 0$ if x is a solution to a VI;

2. $f(x) > 0$ if x is not a solution to a VI.

For example, since the zero of the natural equation $F_{\mathrm{nat}}(x) = 0$ coincides with the solution to VI(S, F) by Theorem 2.3.3, it is obvious that the function $f(x) := \|F_{\mathrm{nat}}(x)\| = \|x - \mathbf{proj}_S\,[x - F(x)]\|$ is the merit function for VI but it is not differentiable even if the mapping F is differentiable; hence, it is not necessarily useful from a numerical perspective.

Let us begin with well-known merit functions for VI(S, F).

Auslender [5] introduced the following *gap function* for VI(S, F):

$$f_\infty(x) := \max_{y \in S} \langle F(x), x - y\rangle.$$

It is easy to see that $f_\infty(x) \ge 0$ for all $x \in S$, and $f_\infty(x) = 0$ if and only if x solves VI(S, F). This function is simple and easy to evaluate but has some numerical drawbacks, such as unboundedness and nondifferentiability.

Fukushima [35] then proposed the following regularized gap function to overcome the issue:

$$f_\alpha(x) := \max_{y \in S} \left\{ \langle F(x), x - y\rangle - \frac{1}{2\alpha}\|y - x\|^2 \right\},$$

where $\alpha > 0$. The function inside the maximization is whenever strongly concave with respect to y for any $\alpha > 0$, and then $f_\alpha(x)$ is differentiable if F is continuously differentiable. It is easy to see that $H_\alpha(x) = \mathbf{proj}_S\,[x - \alpha^{-1}F(x)]$ coincides with the unique solution to the inner maximization of f_α. The derivative of $f_\alpha(x)$ is given by

$$\nabla f_\alpha(x) = F(x) - [\nabla F(x) - \alpha^{-1}I](H_\alpha(x) - x),$$

The function f_α satisfies the merit function for VI(S, F) over S, and a globally optimal solution of the following optimization problem corresponds to the solution to VI(S, F):

$$\min_{x \in S} \quad f_\alpha(x). \tag{2.9}$$

Theorem 2.4.1 (Fukushima [35, Theorem 3.3])**.** *Assume that the mapping* $F\colon \mathbb{R}^n \to \mathbb{R}^n$ *is continuously differentiable. If* x^* *is a stationary point of problem* (2.9)*, i.e.,*

$$\langle \nabla f_\alpha(x^*), x - x^* \rangle \geq 0 \quad \forall x \in S,$$

and the Jacobian matrix $\nabla F(x^*)$ *is positive definite, then* x^* *is a global optimal solution of problem* (2.9)*, and hence it solves* $\mathrm{VI}(S, F)$*.*

Yamashita et al. [128] then found that the following difference of regularized gap functions is the merit function for $\mathrm{VI}(S, F)$ over $\mathbb{R}^n$:

$$f_{\alpha\beta}(x) := f_\alpha(x) - f_\beta(x), \qquad \alpha > \beta > 0.$$

Then $\mathrm{VI}(S, F)$ can be reformulated as the following unconstrained minimization problem:

$$\min_{x \in \mathbb{R}^n} \quad f_{\alpha\beta}(x). \tag{2.10}$$

In addition, Yamashita and Fukushima [127] elucidated that f_α provides an error bound for a certain $\alpha > 0$ under the strong monotonicity of F, and Yamashita et al. [128] investigated conditions under which $f_{\alpha\beta}$ provides a global error bound for the VI.

Theorem 2.4.2 (Yamashita et al. [128, Theorem 3.3])**.** *Assume that* F *is continuously differentiable. If* x^* *is a stationary point of* (2.10)*, i.e.,* $\nabla f_{\alpha\beta}(x^*) = 0$*, and the Jacobian matrix* $\nabla F(x^*)$ *is positive definite, then* x^* *is a global optimal solution of* (2.10)*, and hence it solves* $\mathrm{VI}(S, F)$*.*

Now we introduce a merit function for $\mathrm{CP}(F)$ and its properties.

It is easy to verify that the following function is the merit function for $\mathrm{CP}(F)$:

$$\Psi(x) := \frac{1}{2}\|\Phi(x)\|^2 = \frac{1}{2}\sum_{i=1}^n \phi(x_i, F_i(x))^2.$$

Moreover, Ψ enjoys the following properties.

Proposition 2.4.3 (Facchinei and Soares [31, Proposition 3.4])**.** *If* F *is continuously differentiable,* Ψ *is continuously differentiable and its gradient is* $G\Phi(x)$*, where* $G \in \partial\Phi(x)$ *is the generalized Jacobian matrix of the subdifferential* $\partial\Phi(x)$ *of* Φ *at* x*.*

Definition 2.4.4 (P_0-function)**.** *A function* $F\colon \mathbb{R}^n \to \mathbb{R}^n$ *is a* P_0-function *if, for every* x *and* y *in* $\mathbb{R}^n$ *with* $x \neq y$*, there is an index* i *such that*

$$x_i \neq y_i, \qquad (x_i - y_i)[F_i(x) - F_i(y)] > 0.$$

Theorem 2.4.5 (Facchinei and Soares [31, Theorem 4.1])**.** *Suppose that* F *is a* P_0*-function. Then every stationary point of* Ψ*, i.e.,* $\nabla\Psi(x) = 0$*, is such that* $\Psi(x) = 0$*.*

Next we give the second-order directional derivative of the squared FB-function, which plays an important role to analyze the second-order optimality of $C^{1,1}$ functions in Chapter 3.

Lemma 2.4.6. *Suppose $G, H\colon \mathbb{R}^n \to \mathbb{R}$ are C^2 functions. The second-order directional derivative of the squared FB-function $\phi(G(z), H(z))^2$ at z in the directions $d \in \mathbb{R}^n$ and $v \in \mathbb{R}^n$ is given as follows:*

(a) *If $(G(z), H(z)) \neq (0,0)$, then*

$$\begin{aligned}
&[\phi(G,H)^2]''(z; d, v) = \\
&2\phi(G,H)\left[\left(1-\frac{G}{\sqrt{G^2+H^2}}\right)\nabla G^\top v + \left(1-\frac{H}{\sqrt{G^2+H^2}}\right)\nabla H^\top v\right] \\
&+\phi(G,H)\left[\left(1-\frac{G}{\sqrt{G^2+H^2}}\right)d^\top \nabla^2 G d + \left(1-\frac{H}{\sqrt{G^2+H^2}}\right)d^\top \nabla^2 H d\right] \\
&+\left[\left(1-\frac{G}{\sqrt{G^2+H^2}}\right)\nabla G^\top d + \left(1-\frac{H}{\sqrt{G^2+H^2}}\right)\nabla H^\top d\right]^2 \\
&-\phi(G,H)\frac{\left(H\nabla G^\top d - G\nabla H^\top d\right)^2}{\left(\sqrt{G^2+H^2}\right)^3};
\end{aligned} \tag{2.11}$$

(b) *If $(G(z), H(z)) = (0,0)$, then*

$$[\phi(G,H)^2]''(z; d, v) = \left[\nabla G^\top d + \nabla H^\top d - \sqrt{(\nabla G^\top d)^2 + (\nabla H^\top d)^2}\right]^2. \tag{2.12}$$

Proof. The proof of (a) is obtained from (2.7) and (2.8). The proof of (b) is obtained from the definition of the second-order directional derivatives. □

Chapter 3

Gauss–Seidel method for multi-leader–follower games

3.1 Introduction

When one of the players, called the leader, has the initiative, or it can decide before the other players, called the followers, make decisions, the game is called a Stackelberg game or a single-leader–follower (single-L/F) game. This game was proposed by H.F. von Stackelberg [115] in the 1930s and has been applied in various fields [45]. In the single-L/F game, the leader chooses his/her own strategy taking into account the followers' optimal strategies with the leader's strategy given. The single-L/F game can be regarded as a bilevel optimization. The problem may further be reformulated as a mathematical program with equilibrium constraints (MPEC) by incorporating the optimality conditions for the followers' problems into the constraints of the leader's problem. The MPEC has extensively been studied since the 1990s, see the monographs by Luo et al. [82] and Outrata et al. [90].

In the real world, we may also consider a situation in which two or more leaders decide their strategies first, and then the followers observe the leaders' actions and decide their own strategies. Such a problem can be modeled as the multi-leader–follower (multi-L/F) game. However, it would be more difficult to estimate followers' responses for each leader because the response of each follower depends on all leaders' strategies. Applications of multi-L/F games are found for example in deregulated electricity markets [20, 47, 71, 92].

There are two major approaches for multi-L/F games. The first approach is to regard the followers' responses as functions of leaders' strategies and substitutes those functions for the followers' strategies in each leader's optimization problem. The resultant problem does not include the followers' strategies explicitly. Thus, the multi-L/F game can be regarded as a single level Nash game among the leaders. The works that adopt this approach include [52, 53, 55].

The second approach is to incorporate the optimality conditions of each follower's optimization problem into all leaders' constraints, just like the reformulation of a single-L/F game as an MPEC. The resultant single level Nash game, in which each leader's optimization problem is an MPEC, is called an equilibrium problem with equilibrium constraints (EPEC). In general, the constraints of each leader's problem depend on the other leaders' strategies, and all leaders share decision variables of the followers. The concepts of an equilibrium for EPECs are often defined in terms of a stationarity concept in MPECs. For example, a tuple of solutions to the MPECs of the leaders is called a Bouligand stationary point of an EPEC, if it is composed of Bouligand stationary points of those MPECs. Outrata [89] showed the necessary conditions for an EPEC stationary point. Su [117] proposed a regularization scheme called the sequential NCP method and established convergence to a B-stationary point of an EPEC under some assumptions. For more details about earlier works on multi-L/F games and EPECs, refer to the survey paper by Hu and Fukushima [55].

In this thesis, we adopt the second approach to reformulate a multi-L/F game into an EPEC and propose an algorithm that combines the penalty approach for an MPEC studied by Huang et al. [58] with a nonlinear Gauss–Seidel method. The nonlinear Gauss–Seidel method is one of the diagonalization methods, which solves each leader's MPEC cyclically by fixing the other rival leaders' strategies. In the proposed method, each leader's MPEC is transformed into a differentiable optimization problem by means of a penalty technique, in such a way that the constraints of the problem do not depend on the other rival players' strategies. Hence it can be dealt with as a classical Nash game. Furthermore, the algorithm is easy to implement. We show that a limit point of the sequence generated by the algorithm is an EPEC C-/M-/B-stationarity point under suitable assumptions.

This chapter is organized as follows. In Section 3.2, we provide the basic concepts of the mathematical program with complementarity constraints (MPCC) which is a special class of MPEC. In Section 3.3, we show the convergence of the penalty method for parametrized MPCC. In Section 3.4, we reformulate a multi-L/F game as an EPEC and then propose a Gauss–Seidel penalty method for the EPEC. We also discuss the convergence of the proposed method to a C-/M-/B-stationary point of the multi-L/F game. Moreover, we consider an additional refinement procedure to obtain more accurate solutions. In Section 3.6, we report some results of numerical experiments.

3.2 Preliminaries: Stationarity of parametrized MPCC

In this subsection, we recall some stationarity concepts of the parametrized mathematical program with complementarity constraints (PMPCC). To this end, we define some notions of MPCC.

We consider the following MPCC parametrized by $a \in I\!R^t$:

$$\begin{array}{lll} \mathrm{PMPCC}(a): & \displaystyle\min_{z \in I\!R^n} & f(z,a) \\ & \text{s.t.} & g(z,a) \le 0,\ h(z,a) = 0, \\ & & 0 \le G(z,a) \perp H(z,a) \ge 0, \end{array}$$

where $f(\cdot,a)\colon I\!R^n \to I\!R$, $g(\cdot,a)\colon I\!R^n \to I\!R^r$, $h(\cdot,a)\colon I\!R^n \to I\!R^s$, $G(\cdot,a)\colon I\!R^n \to I\!R^m$ and $H(\cdot,a)\colon I\!R^n \to I\!R^m$ are twice continuously differentiable for any fixed $a \in I\!R^t$, and we assume that $\nabla_z f(\cdot,\cdot)$ and $\nabla_z^2 f(\cdot,\cdot)$ are continuous. Component functions of g, h, G, H will be denoted by g_i, h_i, G_j, H_j, respectively. Let $\mathcal{F}(a)$ denote the feasible set of problem PMPCC(a). For a feasible solution $\bar{z} \in \mathcal{F}(a)$, we define the following sets of indices:

$$\begin{aligned} \mathcal{I}_g(\bar{z},a) &:= \{i : g_i(\bar{z},a) = 0\}, \\ \mathcal{J}_G(\bar{z},a) &:= \{j : G_j(\bar{z},a) = 0\}, \\ \mathcal{J}_H(\bar{z},a) &:= \{j : H_j(\bar{z},a) = 0\}. \end{aligned}$$

It is well known that any feasible solution to MPCC fails to satisfy standard constraint qualifications (CQs) in nonlinear optimization such as linear independence constraint qualification (LICQ) and Mangasarian-Fromovitz constraint qualification (MFCQ). The following CQ is a variant of LICQ tailored to MPCC.

Definition 3.2.1. *The linear independence constraint qualification for MPCC (MPCC-LICQ) is said to hold at* $\bar{z} \in \mathcal{F}(a)$, *if the gradient vectors* $\nabla_z g_i(\bar{z}, a)$ $(i \in \mathcal{I}_g(\bar{z}, a)), \nabla_z h_i(\bar{z}, a)$ $(i = 1, \ldots, q), \nabla_z G_j(\bar{z}, a)$ $(j \in \mathcal{J}_G(\bar{z}, a)), \nabla_z H_j(\bar{z}, a)$ $(j \in \mathcal{J}_H(\bar{z}, a))$ *are linearly independent.*

We give a definition of stationarity for PMPCC(a).

Definition 3.2.2 (Scheel and Scholtes [110])**.** *A feasible point* $\bar{z} \in \mathcal{F}(a)$ *is called a* weak stationary point *of PMPCC*(a) *if there exist* $\bar{\lambda} \in I\!R^r$, $\bar{\mu} \in I\!R^s$, $\bar{\xi} \in I\!R^m$, *and* $\bar{\eta} \in I\!R^m$ *such that*

$$\begin{aligned} \nabla_z f(\bar{z}, a) + \sum_{i \in \mathcal{I}_g(\bar{z}, a)} \bar{\lambda}_i \nabla_z g_i(\bar{z}, a) + \sum_{i=1}^{s} \bar{\mu}_i \nabla_z h_i(\bar{z}, a) & \\ - \sum_{j \in \mathcal{J}_G(\bar{z}, a)} \bar{\xi}_j \nabla_z G_j(\bar{z}, a) - \sum_{j \in \mathcal{J}_H(\bar{z}, a)} \bar{\eta}_j \nabla_z H_j(\bar{z}, a) &= 0, \\ \bar{\lambda}_i \geq 0, \ \bar{\lambda}_i g_i(\bar{z}, a) &= 0, \quad i = 1, \ldots, r \\ G_j(\bar{z}, a) \bar{\xi}_j &= 0, \quad j = 1, \ldots, m, \\ H_j(\bar{z}, a) \bar{\eta}_j &= 0, \quad j = 1, \ldots, m. \end{aligned} \tag{3.1}$$

Definition 3.2.3 (Scheel and Scholtes [110])**.** *A feasible point* $\bar{z} \in \mathcal{F}(a)$ *of* PMPCC(a) *is called*

(a) *a* Clarke (C-) stationary point *at* $\bar{z}$ *if* (3.1) *and* $\bar{\xi}_j \bar{\eta}_j \geq 0$ $(j \in \mathcal{J}_G(\bar{z}, a) \cap \mathcal{J}_H(\bar{z}, a))$ *hold;*

(b) *a* Mordukhovich (M-) stationary point *at* $\bar{z}$ *if* (3.1) *and either* $\bar{\xi}_j > 0$, $\bar{\eta}_j > 0$ *or* $\bar{\xi}_j \bar{\eta}_j = 0$ $(j \in \mathcal{J}_G(\bar{z}, a) \cap \mathcal{J}_H(\bar{z}, a))$ *hold;*

(c) *a* Bouligand (B-) stationary point *at* $\bar{z}$ *if* (3.1) *and* $\bar{\xi}_j \geq 0, \bar{\eta}_j \geq 0$ $(j \in \mathcal{J}_G(\bar{z}, a) \cap \mathcal{J}_H(\bar{z}, a))$ *hold.*

Note that if $\bar{z} \in \mathcal{F}(a)$ is a B-stationary point and satisfies the MPCC-LICQ, $\bar{z}$ is the strong (S-) stationary point. The concepts of C-stationarity, M-stationarity, and B-stationarity are all equivalent if strict complementarity holds at $\bar{z}$, i.e., $\mathcal{J}_G(\bar{z}, a) \cap \mathcal{J}_H(\bar{z}, a) = \emptyset$. However, in general B-stationarity is stronger than M-stationarity which, in turn, is stronger than C-stationarity [110]. Figure 3.1 illustrates the differences among the stationarity concepts related to the Lagrange multipliers $\bar{\xi}_j$ and $\bar{\eta}_j$, where $j \in \mathcal{J}_G(\bar{z}, a) \cap \mathcal{J}_H(\bar{z}, a)$.

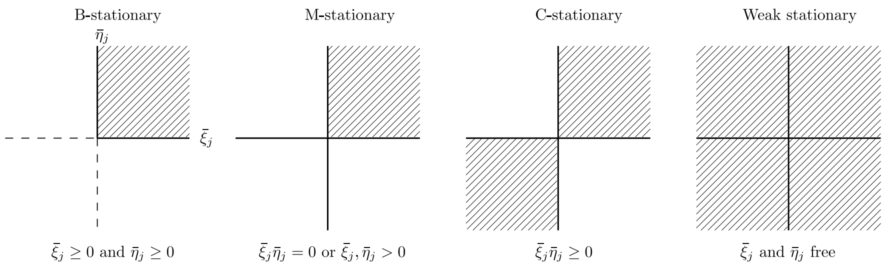


Figure 3.1: MPCC stationarity (B-/M-/C-/weak stationarity); The figure is based on Figure 1 in [3].

Moreover, the following condition plays an important role in discussing convergence of our proposed algorithm to a B-stationary point.

Definition 3.2.4 (Scheel and Scholtes [110]). *The upper-level strict complementarity (ULSC) is said to hold at $\bar{z} \in \mathcal{F}(a)$, if there exist Lagrange multipliers $\bar{\lambda} \in \mathbb{R}^r, \bar{\mu} \in \mathbb{R}^s, \bar{\xi} \in \mathbb{R}^m$, $\bar{\eta} \in \mathbb{R}^m$ satisfying* (3.1)*, and the following condition holds:*

$$\bar{\xi}_j \bar{\eta}_j \neq 0, \ j \in \mathcal{J}_G(\bar{z}, a) \cap \mathcal{J}_H(\bar{z}, a).$$

3.3 Penalty method for parametrized MPCC

In this section, we apply a smooth penalization method for a parametrized family of problems PMPCC(a). The complementarity conditions are often transformed into a system of equations by using the following function $\phi\colon \mathbb{R}^2 \to \mathbb{R}$ called the Fischer-Burmeister (FB) function:

$$\phi(\alpha, \beta) := \alpha + \beta - \sqrt{\alpha^2 + \beta^2}.$$

Since the FB-function has the property that $\phi(\alpha, \beta) = 0$ if and only if $\alpha \geq 0$, $\beta \geq 0, \alpha\beta = 0$, we can rewrite the complementarity conditions as follows:

$$0 \leq G(z, a) \perp H(z, a) \geq 0 \iff \phi(G_j(z, a), H_j(z, a)) = 0, \quad j = 1, \dots, m.$$

The FB-function is not differentiable at $(\alpha, \beta) = (0, 0)$. However, the squared FB-function is continuously differentiable by Proposition 2.4.3, and its gradient is locally Lipschitz [31]. Hence, the second-order directional derivative exists as we stated in Lemma 2.4.6. We utilize this property in the penalty method. By introducing a penalty parameter $\rho > 0$, we define the function $\bar{f}_\rho\colon \mathbb{R}^{n+t} \to \mathbb{R}$ by

$$\bar{f}_\rho(z, a) := f(z, a) + \frac{\rho}{2} \left[\sum_{i=1}^{r} [g_i(z, a)]_+^2 + \sum_{i=1}^{s} |h_i(z, a)|^2 + \sum_{j=1}^{m} |\phi(G_j(z, a), H_j(z, a))|^2 \right],$$

where $[g_i(z, a)]_+ := \max\{g_i(z, a), 0\}$. Observe that $\bar{f}_\rho(\cdot, a)$ is continuously differentiable at any z. We then consider the penalized problem associated with PMPCC(a):

$$\overline{\mathrm{P}}_\rho(a) : \qquad \min_{z \in \mathbb{R}^n} \ \bar{f}_\rho(z, a).$$

Let us define the following index sets:

$$\begin{aligned}
\mathcal{I}_g^+(z, a) &:= \{i : g_i(z, a) > 0\}, \\
\mathcal{I}_h(z, a) &:= \{i : h_i(z, a) = 0\}, \\
\mathcal{I}_h^C(z, a) &:= \{1, \dots, s\} \setminus \mathcal{I}_h(z, a), \\
\mathcal{J}'(z, a) &:= \{j : \phi(G_j(z, a), H_j(z, a)) \neq 0\}, \\
\mathcal{J}_{0+}(z, a) &:= \{j : G_j(z, a) = 0 < H_j(z, a)\}, \\
\mathcal{J}_{+0}(z, a) &:= \{j : G_j(z, a) > 0 = H_j(z, a)\}, \\
\mathcal{J}_{00}(z, a) &:= \{j : G_j(z, a) = 0 = H_j(z, a)\}.
\end{aligned}$$

The feasibility issue underlying this method is verified in the following lemma.

Lemma 3.3.1. *Let $\rho_k \to \infty$, and consider sequences $\{z^k\}$ and $\{a^k\}$ converging to $\bar{z}$ and $\bar{a}$, respectively. If the sequence $\{\bar{f}_{\rho_k}(z^k, a^k)\}$ is bounded above, then $\bar{z}$ is a feasible point of* PMPCC($\bar{a}$).

Proof. By the boundedness assumption on the sequence $\{\bar{f}_{\rho_k}(z^k, a^k)\}$, there exists a real number M satisfying

$$\bar{f}_{\rho_k}(z^k, a^k) = f(z^k, a^k) + \frac{\rho_k}{2}\left[\sum_{i=1}^{r}[g_i(z^k, a^k)]_+^2 + \sum_{i=1}^{s}\left|h_i(z^k, a^k)\right|^2 + \sum_{j=1}^{m}\left|\phi(G_j(z^k, a^k), H_j(z^k, a^k))\right|^2\right] \leq M, \ \forall k > 0.$$

Since $\{f(z^k, a^k)\}$ is bounded, there exists a real number $M' > 0$ such that

$$\sum_{i=1}^{r}[g_i(z^k, a^k)]_+^2 + \sum_{i=1}^{s}\left|h_i(z^k, a^k)\right|^2 + \sum_{j=1}^{m}\left|\phi(G_j(z^k, a^k), H_j(z^k, a^k))\right|^2 \leq \frac{M'}{\rho_k}, \ \forall k > 0.$$

Since $\rho_k \to \infty$, we have

$$[g_i(z^k, a^k)]_+ \to 0, \ h_i(z^k, a^k) \to 0, \ \phi(G_j(z^k, a^k), H_j(z^k, a^k)) \to 0,$$

which implies that $\bar{z}$ is a feasible solution to PMPCC($\bar{a}$). □

To show the first- and second-order condition for optimality of PMPCC($\bar{a}$), we give the following formula: For the C^2 function $g(\cdot, a)$, $[g(z, a)]_+^2$ is a $C^{1,1}$ function, and its second-order directional derivative is given as follows [9, Proposition 3.3]:

$$(g_+^2)''(z, a; d, v) = \begin{cases} 2g(z,a)\nabla g(z,a)^\top v + g(z,a)d^\top \nabla^2 g(z,a)d + (\nabla g(z,a)^\top d)^2, & \text{if } g(z,a) > 0; \\ (\nabla g(z,a)^\top d)^2, & \text{if } g(z,a) = 0; \\ 0, & \text{if } g(z,a) < 0. \end{cases}$$

The next result is a first- and second-order optimality for the penalized problem $\bar{P}_\rho(a)$; its proof is based on Huang et al. [58] for a parameter-free MPCC.

Lemma 3.3.2. *If z is a local optimal solution to $\overline{\mathrm{P}}_\rho(a)$, then it satisfies the first-order necessary condition for optimality*

$$\nabla_z f(z,a) + \rho\left[\sum_{i \in \mathcal{I}_g^+(z,a)} g_i(z,a)\nabla_z g_i(z,a) + \sum_{i \in \mathcal{I}_h^C(z,a)} h_i(z,a)\nabla_z h_i(z,a) + \sum_{j \in \mathcal{J}'(z,a)} \phi(G_j(z,a), H_j(z,a))\nabla_z \phi(G_j(z,a), H_j(z,a))\right] = 0. \tag{3.2}$$

Moreover, it satisfies the weak second-order necessary condition for optimality

$$
\begin{aligned}
&d^\top \nabla_z^2 f(z,a)d + \rho \sum_{i\in\mathcal{I}_g^+(z,a)} \left[g_i(z,a)(d^\top \nabla_z^2 g_i(z,a)d) + (\nabla_z g_i(z,a)^\top d)^2\right] \\
&\quad + \rho \sum_{i\in\mathcal{I}_h^C(z,a)} \left[h_i(z,a)(d^\top \nabla_z^2 h_i(z,a)d) + (\nabla_z h_i(z,a)^\top d)^2\right] \\
&\quad + \rho \sum_{j\in\mathcal{J}'(z,a)} \phi(G_j(z,a),H_j(z,a)) \Bigg[\left(1-\frac{G_j(z,a)}{\sqrt{G_j(z,a)^2+H_j(z,a)^2}}\right) \\
&\quad d^\top \nabla_z^2 G_j(z,a)d + \left(1-\frac{H_j(z,a)}{\sqrt{G_j(z,a)^2+H_j(z,a)^2}}\right) d^\top \nabla_z^2 H_j(z,a)d \\
&\quad - \frac{\left\{H_j(z,a)(\nabla_z G_j(z,a)^\top d) - G_j(z,a)(\nabla_z H_j(z,a)^\top d)\right\}^2}{\left(\sqrt{G_j(z,a)^2+H_j(z,a)^2}\right)^3} \Bigg] \\
&\quad + \rho \sum_{j\in\mathcal{J}'(z,a)} \Bigg[\left(1-\frac{G_j(z,a)}{\sqrt{G_j(z,a)^2+H_j(z,a)^2}}\right) \nabla G_j(z,a)^\top d \\
&\quad + \left(1-\frac{H_j(z,a)}{\sqrt{G_j(z,a)^2+H_j(z,a)^2}}\right) \nabla H_j(z,a)^\top d \Bigg]^2 \geq 0
\end{aligned} \tag{3.3}
$$

for any $d \in I\!R^n$ *such that*

$$
\begin{aligned}
\nabla_z g_i(z,a)^\top d = 0,\ & i \in \mathcal{I}_g(z,a), && (3.4)\\
\nabla_z h_i(z,a)^\top d = 0,\ & i \in \mathcal{I}_h(z,a), && (3.5)\\
\nabla_z G_j(z,a)^\top d = 0,\ & j \in \mathcal{J}_{0+}(z,a) \cup \mathcal{J}_{00}(z,a), && (3.6)\\
\nabla_z H_j(z,a)^\top d = 0,\ & j \in \mathcal{J}_{+0}(z,a) \cup \mathcal{J}_{00}(z,a). && (3.7)
\end{aligned}
$$

Proof. We first show the first-order necessary condition for optimality (3.2). If z is a local optimum of $\overline{\mathrm{P}}_\rho(a)$, then

$$\nabla_z \bar{f}_\rho(z,a) = 0.$$

Since

$$
\begin{aligned}
\nabla_z[g_i(z,a)]_+^2 &= 2g_i(z,a)_+ \nabla_z g_i(z,a),\\
\nabla_z[\phi(G_j(z,a),H_j(z,a))]^2 &= 2\phi(G_j(z,a),H_j(z,a))\nabla_z\phi(G_j(z,a),H_j(z,a)),\\
\nabla_z[h_j(z,a)]^2 &= 2h_j(z,a)\nabla_z h_j(z,a),
\end{aligned}
$$

we obtain

$$\begin{aligned}
\nabla_z \bar{f}_\rho(z,a) =& \nabla_z f(z,a) + \rho \Bigg[\sum_{i=1}^{r} \nabla_z g_i(z,a) g_i(z,a)_+ + \sum_{i=1}^{s} \nabla_z h_i(z,a) h_i(z,a) \\
&+ \sum_{j=1}^{m} \nabla_z \phi(G_j(z,a), H_j(z,a)) \phi(G_j(z,a), H_j(z,a)) \Bigg] \\
=& \nabla_z f(z,a) + \rho \Bigg[\sum_{i \in \mathcal{I}_g^+(z,a)} g_i(z,a) \nabla_z g_i(z,a) + \sum_{i \in \mathcal{I}_h^C(z,a)} h_i(z,a) \nabla_z h_i(z,a) \\
&+ \sum_{j \in \mathcal{J}'(z,a)} \phi(G_j(z,a), H_j(z,a)) \nabla_z \phi(G_j(z,a), H_j(z,a)) \Bigg] = 0.
\end{aligned}$$

This proves the first half.

We next prove the second-order condition (3.3). By the first-order condition (3.2), we have

$$\bar{f}'_\rho(z,a;d) = \nabla_z \bar{f}_\rho(z,a)^\top d = 0 \quad \forall d \in I\!R^n,$$

Then [9, Theorem 6.5] ensures that

$$\bar{f}''_\rho(z,a;d,d) \geq 0 \quad \forall d \in I\!R^n.$$

Calculating the second-order directional derivative of $\bar{f}''_\rho$ leads to

$$\begin{aligned}
&\bar{f}''_\rho(z,a;d,d) = \\
&\quad f''(z,a;d,d) + \frac{\rho}{2} \Bigg[\sum_{i=1}^{r} [(g_i)_+^2]''(z,a;d,d) + \sum_{i=1}^{s} [h_i^2]''(z,a;d,d) + \sum_{j=1}^{m} [\phi(G_j,H_j)^2]''(z,a;d,d) \Bigg] \\
&= \nabla_z f(z,a)^\top d + \rho \Bigg[\sum_{i \in \mathcal{I}_g^+(z,a)} g_i(z,a) \nabla_z g_i(z,a)^\top d + \sum_{i \in \mathcal{I}_h^C(z,a)} h_i(z,a) \nabla_z h_i(z,a)^\top d \qquad (3.8) \\
&\quad + \sum_{j \in \mathcal{J}'(z,a)} \phi(G_j(z,a), H_j(z,a)) \nabla_z \phi(G_j(z,a), H_j(z,a))^\top d \Bigg] \qquad (3.9) \\
&\quad + \frac{1}{2} \Bigg\{ d^\top \nabla_z^2 f(z,a) d + \rho \sum_{i \in \mathcal{I}_g^+(z,a)} g_i(z,a) d^\top \nabla_z^2 g_i(z,a) d \\
&\quad + \rho \sum_{i \in \mathcal{I}_g^+(z,a)} (\nabla_z g_i(z,a)^\top d)^2 + \rho \sum_{i \in \mathcal{I}_g(z,a)} (\nabla_z g_i(z,a)^\top d)_+^2 \\
&\quad + \rho \sum_{i \in \mathcal{I}_h^C(z,a)} \Big\{ (\nabla_z h_i(z,a)^\top d)^2 + h_i(z,a) d^\top \nabla_z^2 h_i(z,a) d \Big\} \\
&\quad + \rho \sum_{i \in \mathcal{I}_h(z,a)} (\nabla_z h_i(z,a)^\top d)^2
\end{aligned}$$

$$
\begin{aligned}
&+\rho \sum_{j\in\mathcal{J}_{00}(z,a)} \left[\nabla G_j(z,a)^\top d + \nabla H_j(z,a)^\top d - \sqrt{(\nabla G_j(z,a)^\top d)^2 + (\nabla H_j(z,a)^\top d)^2}\right]^2 && (3.10)\\
&+\rho \sum_{j\in\{1,\ldots,m\}\setminus\mathcal{J}_{00}(z,a)} \phi(G_j(z,a),H_j(z,a))\left[\left(1-\frac{G_j(z,a)}{\sqrt{G_j(z,a)^2+H_j(z,a)^2}}\right)d^\top\nabla^2 G_j(z,a)d\right. && (3.11)\\
&\left.+\left(1-\frac{H_j(z,a)}{\sqrt{G_j(z,a)^2+H_j(z,a)^2}}\right)d^\top\nabla^2 H_j(z,a)d\right] && (3.12)\\
&+\rho \sum_{j\in\{1,\ldots,m\}\setminus\mathcal{J}_{00}(z,a)} \left[\left(1-\frac{G_j(z,a)}{\sqrt{G_j(z,a)^2+H_j(z,a)^2}}\right)\nabla G_j(z,a)^\top d\right. && (3.13)\\
&\left.+\left(1-\frac{H_j(z,a)}{\sqrt{G_j(z,a)^2+H_j(z,a)^2}}\right)\nabla H_j(z,a)^\top d\right]^2 && (3.14)\\
&\left.-\rho \sum_{j\in\{1,\ldots,m\}\setminus\mathcal{J}_{00}(z,a)} \phi(G_j(z,a),H_j(z,a))\frac{\left(H_j(z,a)\nabla G_j(z,a)^\top d - G_j(z,a)\nabla H_j(z,a)^\top d\right)^2}{\left(\sqrt{G_j(z,a)^2+H_j(z,a)^2}\right)^3}\right\} && (3.15)\\
&\geq 0.
\end{aligned}
$$

The terms (3.8)–(3.9) vanish since they are equal to $\nabla_z \bar{f}_\rho(z,a)^\top d$. The term (3.10) is calculated from (2.12), and terms (3.11) to (3.15) are calculated from (2.11). Then, for any $d \in I\!R^n$ satisfying (3.4)–(3.7), we obtain (3.3) (Note that $\mathcal{J}'(z,a) = \{1,\ldots,m\} \setminus (\mathcal{J}_{0+}(z,a) \cup \mathcal{J}_{+0}(z,a) \cup \mathcal{J}_{00}(z,a)) \subset \{1,\ldots,m\} \setminus \mathcal{J}_{00}(z,a)$). The proof is complete. □

Lemma 3.3.3. *Assume that the assumption of Lemma* 3.3.1 *holds. Suppose that the sequence* $\{z_k\}$ *satisfying first-order condition* (3.2) *for all* k *converges to* $\bar{z}$, *and MPCC-LICQ for PMPCC*(a) *holds at the limit* $\bar{z}$. *Then,* $\bar{z}$ *is a weak stationary point; that is, there exist* $\bar{\lambda} \in I\!R^r$, $\bar{\mu} \in I\!R^s$, $\bar{\xi} \in I\!R^m$, *and* $\bar{\eta} \in I\!R^m$ *such that*

$$
\begin{aligned}
&\nabla_z f(\bar{z},\bar{a}) + \sum_{i\in\mathcal{I}_g(\bar{z},\bar{a})} \bar{\lambda}_i \nabla_z g_i(\bar{z},\bar{a}) + \sum_{i=1}^{s} \bar{\mu}_i \nabla_z h_i(\bar{z},\bar{a})\\
&- \sum_{j\in\mathcal{J}_{0+}(\bar{z},\bar{a})\cup\mathcal{J}_{00}(\bar{z},\bar{a})} \bar{\xi}_j \nabla_z G_j(\bar{z},\bar{a}) - \sum_{j\in\mathcal{J}_{+0}(\bar{z},\bar{a})\cup\mathcal{J}_{00}(\bar{z},\bar{a})} \bar{\eta}_j \nabla_z H_j(\bar{z},\bar{a}) = 0,\\
&\bar{\lambda}_i \geq 0,\ \bar{\lambda}_i g_i(\bar{z},a) = 0, \quad i = 1,\ldots,r\\
&G_j(\bar{z},a)\bar{\xi}_j = 0, \quad j = 1,\ldots,m,\\
&H_j(\bar{z},a)\bar{\eta}_j = 0, \quad j = 1,\ldots,m.
\end{aligned}
\tag{3.16}
$$

Proof. The statement can be proved by appropriate modifications of the arguments in Lemma 4.3 in Huang et al. [58] by the continuity of $\nabla_z f$, $\nabla_z^2 f$, etc.; see Appendix A for the overview of the proof. □

As a result, we establish convergence of the penalty method for the parametrized family of problems PMPCC(a).

Theorem 3.3.4. *Let* $\rho_k \to \infty$, $a^k \to \bar{a}$, *and* $z^k \to \bar{z}$, *where* z^k *is a stationary point of* $\bar{P}_{\rho_k}(a^k)$, *i.e.,* (3.2) *holds, for each* k. *Assume that the sequence* $\{\bar{f}_{\rho_k}(z^k,a^k)\}$ *is bounded above. Suppose that the MPCC-LICQ holds at the limit* $\bar{z}$. *Then,*

(a) $\bar{z}$ *is a C-stationary point of PMPCC*($\bar{a}$);

(b) *if weak second-order condition* (3.3) *holds at* $\bar{z}$, *then* $\bar{z}$ *is an M-stationary point of PMPCC*($\bar{a}$);

(c) *if weak second-order condition* (3.3) *and the ULSC hold at* $\bar{z}$, *then* $\bar{z}$ *is a B-stationary point of PMPCC*($\bar{a}$).

Proof. First, we have $\bar{z} \in \mathcal{F}(\bar{a})$ by Lemma 3.3.1. Moreover, by Lemma 3.3.3, there exist $\bar{\lambda} \in I\!R^r_+$, $\bar{\mu} \in I\!R^s$, $\bar{\xi} \in I\!R^m$, and $\bar{\eta} \in I\!R^m$ such that (3.16) holds at $\bar{z} \in \mathcal{F}(\bar{a})$. The proof of each claim is presented as follows:

(a) It suffices to show that $\bar{\xi}_j \bar{\eta}_j \geq 0$ for $j \in \mathcal{J}_{00}(\bar{z}, \bar{a})$. Indeed, we have

$$\xi^k_j \eta^k_j = \begin{cases} 0 & \text{if } j \in \mathcal{J}_{0+}(z^k, a^k) \cup \mathcal{J}_{+0}(z^k, a^k) \cup \mathcal{J}_{00}(z^k, a^k); \\ (\rho_k)^2 \phi(G_j(z^k, a^k), H_j(z^k, a^k))^2 a^k_j b^k_j & \text{if } j \in \mathcal{J}'(z^k, a^k), \end{cases}$$

where

$$a^k_j := 1 - \frac{G_j(z^k, a^k)}{\sqrt{G_j(z^k, a^k)^2 + H_j(z^k, a^k)^2}}, \quad b^k_j := 1 - \frac{H_j(z^k, a^k)}{\sqrt{G_j(z^k, a^k)^2 + H_j(z^k, a^k)^2}}.$$

Then we have $\xi^k_j \eta^k_j \geq 0$ for all j, and hence $\bar{\xi}_j \bar{\eta}_j \geq 0$ for $j \in \mathcal{J}_{00}(\bar{z}, \bar{a})$.

(b) Assume that $\bar{z}$ is not a M-stationary point; that is, there exists $j^* \in \mathcal{J}_{00}(\bar{z}, \bar{a})$ such that $\bar{\xi}_{j^*} < 0$ and $\bar{\eta}_{j^*} < 0$. Then we can show the rest of the proof as the same technique as Theorem 4.4 in [58][1]; see Appendix A for the sketch of the proof tailored to our model.

(c) Since $\bar{z}$ is an M-stationary point by (b), under the ULSC assumption, it follows from the definition of B-stationary point that $\bar{z}$ is a B-stationary point.

This completes the proof. □

3.4 Method for multi-L/F games

In this section, we propose a numerical method for multi-L/F games by way of EPECs. First we describe the reformulation of multi-L/F games as EPECs. Then we elaborate on a Gauss–Seidel type penalty method for EPECs, and a refinement procedure to obtain more accurate solutions.

3.4.1 Reformulation of the multi-L/F game as EPEC

Recall the multi-L/F game comprised of N leaders and M followers, introduced in Section 1.1. For given $x^{-\nu} := (x^1, \ldots, x^{\nu-1}, x^{\nu+1}, \ldots, x^N) \in I\!R^{n-n_\nu}$ and $y := (y^1, \ldots, y^M) \in I\!R^m$, leader ν solves the following problem:

$$\min_{x^\nu \in I\!R^{n_\nu}} \theta_\nu(x^\nu, x^{-\nu}, y) \quad \text{s.t. } x^\nu \in X^\nu, \tag{3.17}$$

[1]While only B-stationarity is shown in [58], in fact it can be seen that most of the proof is intrinsically devoted to the convergence to the M-stationary point.

where $\theta_\nu \colon I\!R^{n+m} \to I\!R$ is twice continuously differentiable. In this chapter, suppose that the strategy set $X^\nu \subset I\!R^{n_\nu}$ of leader ν in (3.17) is given by

$$X^\nu := \{x^\nu \mid g^\nu(x^\nu) \le 0,\ h^\nu(x^\nu) = 0\},$$

where $g^\nu \colon I\!R^{n_\nu} \to I\!R^{r_\nu}$ and $h^\nu \colon I\!R^{n_\nu} \to I\!R^{s_\nu}$ are twice continuously differentiable.

For a given tuple of leaders' strategies $x := (x^1, \dots, x^N) \in X := X^1 \times \cdots \times X^N \subset I\!R^n$ and the other followers' strategies $y^{-\omega} := (y^1, \dots, y^{\omega-1}, y^{\omega+1}, \dots, y^M) \in I\!R^{m-m_\omega}$, follower ω solves the following problem:

$$\min_{y^\omega \in I\!R^{m_\omega}} \gamma_\omega(x, y^\omega, y^{-\omega}) \quad \text{s.t. } y^\omega \in Y^\omega(x), \tag{3.18}$$

where $\gamma_\omega \colon I\!R^{n+m} \to I\!R$ is thrice continuously differentiable. Suppose that follower ω's strategy set $Y^\omega(x) \subset I\!R^{m_\omega}$ in (3.18) is given by

$$Y^\omega(x) := \{y^\omega \mid u^\omega(x, y^\omega) \le 0,\ v^\omega(x, y^\omega) = 0\},$$

where $u^\omega \colon I\!R^{n+m_\omega} \to I\!R^{p_\omega}$ and $v^\omega \colon I\!R^{n+m_\omega} \to I\!R^{q_\omega}$ are thrice continuously differentiable. We assume that problem (3.18) is convex with respect to y^ω, i.e., for any x and $y^{-\omega}$, $\gamma_\omega(x, \cdot, y^{-\omega})$, and $u^\omega(x, \cdot)$ are convex, and $v^\omega(x, \cdot)$ is affine, and the problem satisfies an appropriate constraint qualification for any fixed x and $y^{-\omega}$.

The convexity assumption for all ω ensures that for each x, the followers' Nash game can be written as the following mixed complementarity system, which is comprised of the KKT conditions for $\omega = 1, \dots, M$:

$$\psi(x, y, z, \lambda, \mu) = 0,\ 0 \le z \perp \lambda \ge 0, \tag{3.19}$$

where

$$\psi(x, y, z, \lambda, \mu) := \begin{bmatrix} \nabla_{y^1}\gamma_1(x, y^1, y^{-1}) + \nabla_{y^1} u^1(x, y^1)\lambda^1 + \nabla_{y^1} v^1(x, y^1)\mu^1 \\ \vdots \\ \nabla_{y^M}\gamma_M(x, y^M, y^{-M}) + \nabla_{y^M} u^M(x, y^M)\lambda^M + \nabla_{y^M} v^M(x, y^M)\mu^M \\ u^1(x, y^1) + z^1 \\ \vdots \\ u^M(x, y^M) + z^M \\ v^1(x, y^1) \\ \vdots \\ v^M(x, y^M) \end{bmatrix} \in I\!R^{m+p+q}.$$

Here, $\lambda^\omega := (\lambda_1^\omega, \dots, \lambda_{p_\omega}^\omega)^\top \in I\!R^{p_\omega}, \mu^\omega := (\mu_1^\omega, \dots, \mu_{q_\omega}^\omega)^\top \in I\!R^{q_\omega}$ are Lagrange multipliers, and $z^\omega \in I\!R^{p_\omega}$ is a vector of slack variables for the inequality constraints $u^\omega(x, y^\omega) \le 0$. Further, $z := (z^1, \dots, z^M) \in I\!R^p$, $\lambda := (\lambda^1, \dots, \lambda^M) \in I\!R^p$, $\mu := (\mu^1, \dots, \mu^M) \in I\!R^q$, where $p := p_1 + \cdots + p_M$ and $q := q_1 + \cdots + q_M$.

By incorporating (3.19) into each leader's optimization problem (3.17), we have the following parametrized MPCC for leader ν:

$$\begin{array}{lll} \mathrm{PMPCC}^\nu(x^{-\nu}) : & \displaystyle\min_{x^\nu, y, z, \lambda, \mu} & \theta_\nu(x^\nu, x^{-\nu}, y) \\ & \text{s.t.} & g^\nu(x^\nu) \le 0,\ h^\nu(x^\nu) = 0, \\ & & \psi(x^\nu, x^{-\nu}, y, z, \lambda, \mu) = 0, \\ & & 0 \le z \perp \lambda \ge 0. \end{array}$$

Thus the multi-L/F game is reduced to the EPEC, which seeks an equilibrium point that simultaneously achieves optimality in $(\mathrm{PMPCC}^\nu(x^{-\nu}))_{\nu=1}^N$. We call $(y, z, \lambda, \mu) \in I\!R^{m+2p+q}$ shared variables, because all leaders have those as decision variables. Now, we define a C-/M-/B-stationary equilibrium point of the EPEC.

Definition 3.4.1. *A tuple* $(x^*, y^*, z^*, \lambda^*, \mu^*) \in I\!R^{n+m+2p+q}$ *is called a* Clarke (C-), Mordukhovich (M-), *and* Bouligand (B-) stationary equilibrium point of the EPEC (or multi-L/F game)*, if for each leader* ν, $(x^{\nu,*}, y^*, z^*, \lambda^*, \mu^*)$ *is a C-stationary, M-stationary, and B-stationary point of* $\mathrm{PMPCC}^\nu(x^{-\nu,*})$*, respectively.*

Next we discuss the uniqueness of a solution to the followers' KKT system. First we show the uniqueness of $y = (y^1, \ldots, y^M)$. Note that the y-part of the solution to (3.19) represents the Nash equilibrium for the followers' Nash game because of the necessity and sufficiency for optimality of the KKT conditions. However, as mentioned in Section 1.1, the equilibrium of the followers' Nash game is not determined uniquely in general. To ensure the existence and uniqueness, we need more assumptions.

To this end, define the mapping $F\colon I\!R^{n+m} \to I\!R^m$ by

$$F(x, y) := \begin{bmatrix} \nabla_{y^1}\gamma_1(x, y^1, y^{-1}) \\ \vdots \\ \nabla_{y^M}\gamma_M(x, y^M, y^{-M}) \end{bmatrix},$$

and the set $Y(x) \subset I\!R^m$ by $Y(x) := Y^1(x) \times \cdots \times Y^M(x)$, which is closed and convex by assumption. It is well known that the Nash game among the followers is equivalent to the following VI problem parametrized by x by the convexity assumption: Find $y^* \in Y(x)$ such that

$$\langle F(x, y^*), y - y^* \rangle \geq 0 \qquad \forall y \in Y(x). \tag{3.20}$$

We denote (3.20) as $\mathrm{VI}(Y(x), F(x, \cdot))$.

Lemma 3.4.2. *If the mapping* $F(x, \cdot)\colon I\!R^m \to I\!R^m$ *is strongly monotone on* $Y(x)$ *for any given* x*, then the solution* $y^* \in Y(x)$ *to* (3.20) *uniquely exists. Moreover, for any given* x*, if the linear independence constraint qualification holds for each* ω*'s optimization problem* (3.18)*, then the pair* (λ, μ) *of Lagrange multipliers uniquely exists.*

Proof. This is directly obtained from Theorem 2.3.8 in the first half statement. The proof of the second half statement is shown by Theorem 2.1.21. □

3.4.2 Gauss–Seidel penalty method

In this and the next subsections, we develop a numerical method for solving multi-L/F games. In this subsection, we propose a Gauss–Seidel penalty method. To this end, we reformulate each leader's MPCC as a nonsmooth optimization problem, and then we transform it into a differentiable unconstrained problem using a penalty technique. Our method is a combination of the smooth penalization method [58] and the nonlinear Gauss–Seidel method.

In leader ν's problem $\mathrm{PMPCC}^\nu(x^{-\nu})$, the complementarity constraints can be replaced by the equality constraints by means of the FB-function $\phi\colon I\!R^2 \to I\!R$ as mentioned in Section 3.3. Specifically, $\mathrm{PMPCC}^\nu(x^{-\nu})$ can be rewritten as

$$\begin{array}{lll} \mathrm{P}^\nu(x^{-\nu}): & \displaystyle\min_{x^\nu, y, z, \lambda, \mu} & \theta_\nu(x^\nu, x^{-\nu}, y) \\ & \text{s.t.} & g^\nu(x^\nu) \leq 0, \ h^\nu(x^\nu) = 0, \\ & & \Psi(x^\nu, x^{-\nu}, y, z, \lambda, \mu) = 0, \end{array}$$

where

$$\Psi(x^\nu, x^{-\nu}, y, z, \lambda, \mu) := \begin{bmatrix} \psi(x^\nu, x^{-\nu}, y, z, \lambda, \mu) \\ \Phi(z^1, \lambda^1) \\ \vdots \\ \Phi(z^M, \lambda^M) \end{bmatrix},$$

$$\Phi(z^\omega, \lambda^\omega) := \begin{bmatrix} \phi(z_1^\omega, \lambda_1^\omega) \\ \vdots \\ \phi(z_{p_\omega}^\omega, \lambda_{p_\omega}^\omega) \end{bmatrix}, \quad \omega = 1, \dots, M.$$

However, $\mathrm{P}^\nu(x^{-\nu})$ is nonsmooth because of the nondifferentiability of the FB-function. To avoid this difficulty, we use the property that the squared FB-function is differentiable everywhere [31]. Define the penalty function associated with problem $\mathrm{P}^\nu(x^{-\nu})$ by

$$\bar{\theta}_\nu^\rho(x^\nu, x^{-\nu}, y, z, \lambda, \mu) := \theta_\nu(x^\nu, x^{-\nu}, y) + \frac{\rho}{2}\left[\sum_{i=1}^{r_\nu} [g_i^\nu(x^\nu)]_+^2 \right. \\ \left. + \sum_{i=1}^{s_\nu} |h_i^\nu(x^\nu)|^2 + \sum_{j=1}^{m+2p+q} |\Psi_j(x^\nu, x^{-\nu}, y, z, \lambda, \mu)|^2 \right],$$

where $\rho > 0$ is a penalty parameter. The penalized problem for leader ν's problem $\mathrm{P}^\nu(x^{-\nu})$ is written as

$$\overline{\mathrm{P}}_\rho^\nu(x^{-\nu}) : \quad \min_{x^\nu, y, z, \lambda, \mu} \quad \bar{\theta}_\nu^\rho(x^\nu, x^{-\nu}, y, z, \lambda, \mu),$$

which is a differentiable unconstrained optimization problem. The proposed algorithm is formally stated in Algorithm 1.

Algorithm 1 Gauss–Seidel Penalty Method

Input: Initial point $x^{(0)} := (x^{1,(0)}, \dots, x^{N,(0)}), y^{(0)}, z^{(0)}, \lambda^{(0)}, \mu^{(0)}$, an increasing positive sequence $\{\rho_k\}$, a tolerance $\varepsilon > 0$, and the maximum number of major iterations $K_{\max}$.

Output: An approximate stationary equilibrium point of EPEC $(\mathrm{PMPCC}^\nu(x^{-\nu}))_{\nu=1}^N$, or the multi-L/F game.

1: Set $k = 0$.

2: For each ν, solve $\overline{\mathrm{P}}_{\rho_k}^\nu(\bar{x}^{-\nu,(k)})$ to obtain the solution

$$\bar{w}^{\nu,(k+1)} := (\bar{x}^{\nu,(k+1)}, \bar{y}^{\nu,(k+1)}, \bar{z}^{\nu,(k+1)}, \bar{\lambda}^{\nu,(k+1)}, \bar{\mu}^{\nu,(k+1)}),$$

where $\bar{x}^{-\nu,(k)} := (\bar{x}^{1,(k+1)}, \dots, \bar{x}^{\nu-1,(k+1)}, \bar{x}^{\nu+1,(k)}, \dots, \bar{x}^{N,(k)})$.

3: Stop if

$$\max\left\{ \max_{1\le i\le r_\nu} [g_i^\nu(\bar{x}^{\nu,(k+1)})]_+, \max_{1\le i\le s_\nu} |h_i^\nu(\bar{x}^{\nu,(k+1)})|, \max_{1\le i\le m+2p+q} |\Psi_i(\bar{w}^{\nu,(k+1)}, \bar{x}^{-\nu,(k)})| \right\} < \varepsilon$$

holds for all ν.

4: If $k < K_{\max}$, set $k := k + 1$ and go to Step 2. If $k = K_{\max}$, terminate.

In Step 2 of Algorithm 1, we use the notation $\bar{y}^\nu, \bar{z}^\nu, \bar{\lambda}^\nu, \bar{\mu}^\nu$ to distinguish among leaders, because all leaders do not necessarily output the same solutions y, z, λ, μ. Note that leader ν

uses the latest rival strategies from leader 1 to $\nu-1$, which originates from the update scheme of a Gauss–Seidel method for a linear equation.

Next we discuss the convergence of the proposed algorithm with $\varepsilon = 0$ and $K_{\max} = \infty$. First, we argue the feasibility issue of Algorithm 1.

Lemma 3.4.3. *Let $\rho_k \to \infty$, and for each ν, $\bar{w}^{\nu,(k)} \to \bar{w}^{\nu,(\infty)}, \bar{x}^{-\nu,(k)} \to \bar{x}^{-\nu,(\infty)}$. Assume that the sequence $\{\bar{\theta}_\nu^{\rho_k}(\bar{w}^{\nu,(k+1)}, \bar{x}^{-\nu,(k)})\}$ is bounded above. Then, $\bar{w}^{\nu,(\infty)}$ is a feasible solution to $\mathrm{P}^\nu(\bar{x}^{-\nu,(\infty)})$, i.e., $\mathrm{PMPCC}^\nu(\bar{x}^{-\nu,(\infty)})$.*

Proof. Setting $a^k := \bar{x}^{-\nu,(k)}$ in Lemma 3.3.1 yields the conclusion of the lemma. □

The next lemma gives conditions under which the sequence of shared variables $\bar{y}^{\nu,(k)}$ and Lagrange multipliers $(\bar{\lambda}^{\nu,(k)}, \bar{\mu}^{\nu,(k)})$ converge to limit points independent of ν.

Lemma 3.4.4. *Assume that the conditions of Lemma 3.4.3 hold. Suppose the sequence $\{(\bar{x}^{\nu,(k)}, \bar{x}^{-\nu,(k)}, \bar{y}^{\nu,(k)}, \bar{z}^{\nu,(k)}, \bar{\lambda}^{\nu,(k)}, \bar{\mu}^{\nu,(k)})\}_{\nu=1}^N$ generated by the algorithm converges to $\{(\bar{x}^{(\infty)}, \bar{y}^{\nu,(\infty)}, \bar{z}^{\nu,(\infty)}, \bar{\lambda}^{\nu,(\infty)}, \bar{\mu}^{\nu,(\infty)})\}_{\nu=1}^N$, and the mapping $F(x,\cdot)\colon \mathbb{R}^m \to \mathbb{R}^m$ is strongly monotone for any fixed x. Then the shared variable $\bar{y}^{\nu,(k)}$ converges to the same limit $\bar{y}^{(\infty)}$ independent of ν. Furthermore, if the LICQ holds at $\bar{y}^{(\infty)}$ in the followers' problems* (3.18)*, then $(\bar{\lambda}^{\nu,(k)}, \bar{\mu}^{\nu,(k)})$ also converges to the same limit $(\bar{\lambda}^{(\infty)}, \bar{\mu}^{(\infty)})$.*

Proof. By Lemma 3.4.3, the limit $(\bar{x}^{\nu,(\infty)}, \bar{y}^{\nu,(\infty)}, \bar{z}^{\nu,(\infty)}, \bar{\lambda}^{\nu,(\infty)}, \bar{\mu}^{\nu,(\infty)})$ is a feasible solution to $\mathrm{P}^\nu(\bar{x}^{-\nu,(\infty)})$, which implies

$$\Psi(\bar{x}^{\nu,(\infty)}, \bar{x}^{-\nu,(\infty)}, \bar{y}^{\nu,(\infty)}, \bar{z}^{\nu,(\infty)}, \bar{\lambda}^{\nu,(\infty)}, \bar{\mu}^{\nu,(\infty)}) = 0.$$

Moreover, by the given assumption, $\{(\bar{x}^{\nu,(k)}, \bar{x}^{-\nu,(k)})\}$ converges to the identical limit $\bar{x}^{(\infty)}$ independent of ν. Consider the system of equations

$$\Psi(\bar{x}^{(\infty)}, y, z, \lambda, \mu) = 0.$$

A solution (y, z, λ, μ) constitutes the Nash equilibrium together with the corresponding Lagrange multipliers in the followers' problems with the leaders' strategies $\bar{x}^{(\infty)}$ given.

Under the given strong monotonicity assumption, the y-part of the solution is unique from the first half of Lemma 3.4.2. Hence, $\bar{y}^{\nu,(\infty)}$ are identical, i.e., $\bar{y}^{\nu,(\infty)} = \bar{y}^{(\infty)}$ for all ν. Under the LICQ assumption, the remaining part of the solution is also unique from the second half of Lemma 3.4.2. Consequently, we have $(\bar{z}^{\nu,(\infty)}, \bar{\lambda}^{\nu,(\infty)}, \bar{\mu}^{\nu,(\infty)}) = (\bar{z}^{(\infty)}, \bar{\lambda}^{(\infty)}, \bar{\mu}^{(\infty)})$ for all ν. □

We are ready to show that if the algorithm converges, then the limit is a C-/M-/B-stationary point of the EPEC under appropriate assumptions.

Theorem 3.4.5. *Let $\rho_k \to \infty$. Suppose, for each $\nu = 1, \dots, M$, $(\bar{w}^{\nu,(k)}, \bar{x}^{-\nu,(k)}) \to (\bar{x}^{(\infty)}, \bar{y}^{\nu,(\infty)}, \bar{z}^{\nu,(\infty)}, \bar{\lambda}^{\nu,(\infty)}, \bar{\mu}^{\nu,(\infty)})$, where $\bar{w}^{\nu,(k+1)}$ is a stationary point of problem $\overline{\mathrm{P}}^\nu_{\rho_k}(\bar{x}^{-\nu,(k)})$ for each k. Assume that the conditions of Lemmas* 3.4.3 *and* 3.4.4 *hold. Moreover, suppose that, for each ν, the MPCC-LICQ for $\mathrm{PMPCC}^\nu(\bar{x}^{-\nu,(\infty)})$ holds at the limit point $(\bar{x}^{\nu,(\infty)}, \bar{y}^{\nu,(\infty)}, \bar{z}^{\nu,(\infty)}, \bar{\lambda}^{\nu,(\infty)}, \bar{\mu}^{\nu,(\infty)})$. Then, the limit point $(\bar{x}^{\nu,(\infty)}, \bar{y}^{\nu,(\infty)}, \bar{z}^{\nu,(\infty)}, \bar{\lambda}^{\nu,(\infty)}, \bar{\mu}^{\nu,(\infty)})$ is*

(a) *a C-stationary point of $\mathrm{PMPCC}^\nu(\bar{x}^{-\nu,(\infty)})$ for each ν;*

(b) *an M-stationary point of $\mathrm{PMPCC}^\nu(\bar{x}^{-\nu,(\infty)})$ for each ν if weak second-order condition* (3.3)*, where z, a, and ρ are replaced by $w^{\nu,(k)}$, $\bar{x}^{-\nu,(k)}$, and ρ_k, respectively, holds;*

(c) *a B-stationary point of* $\mathrm{PMPCC}^\nu(\bar{x}^{-\nu,(\infty)})$ *for each* ν *if the weak second-order condition (see* (b)*) and ULSC hold,*

and those limit points are identical. Consequently, it respectively constitutes a C-/M-/B-stationary equilibrium point of the multi-L/F game.

Proof. Notice that each leader ν solves PMPCC(a) with $a := \bar{x}^{-\nu}$ mentioned in Section 3.3, and $\bar{x}^{-\nu,(k)}$ converges to $\bar{x}^{-\nu,(\infty)}$ by assumption. First observe that, from Lemma 3.4.3, the limit point $(\bar{x}^{\nu,(\infty)}, \bar{y}^{\nu,(\infty)}, \bar{z}^{\nu,(\infty)}, \bar{\lambda}^{\nu,(\infty)}, \bar{\mu}^{\nu,(\infty)})$ is feasible to $\mathrm{PMPCC}^\nu(\bar{x}^{-\nu,(\infty)})$. The claims (a), (b), and (c) can be straightforwardly shown by Theorem 3.3.4. By Lemma 3.4.4, those limit points are identical for all ν. Hence, $(\bar{x}^{(\infty)}, \bar{y}^{(\infty)}, \bar{z}^{(\infty)}, \bar{\lambda}^{(\infty)}, \bar{\mu}^{(\infty)})$ is a C-/M-/B-stationary equilibrium point of the multi-L/F game, respectively. □

From the numerical viewpoint, the squared penalty method has some drawbacks. The main issue is that the penalized problem becomes ill-conditioned as the penalty parameter ρ_k increases, and hence it is difficult to find an accurate solution. In such a case, we may use the algorithm as the identification phase of active sets in the complementarity constraints $0 \le z \perp \lambda \ge 0$, and then transfer to the refinement phase proposed in the next subsection to obtain a more accurate equilibrium of the EPEC.

3.4.3 Refined Gauss–Seidel method

The basic tool used in the previous subsection is the quadratic penalty technique in constrained optimization. However, a computed solution may not be exact for $\mathrm{P}^\nu(x^{-\nu,(k)})$, or $\mathrm{PMPCC}^\nu(x^{-\nu,(k)})$, even for a sufficient large k. Nevertheless, it may provide useful information about active sets in the complementarity constraints. In fact, if the active sets are correctly identified, we may further refine the solution produced by Algorithm 1. To this end, we present another Gauss–Seidel-based method for obtaining a more accurate solution.

Let $\bar{w}^{*,\nu} = (\bar{w}^{*,1}, \dots, \bar{w}^{*,N})$ be a solution obtained by Algorithm 1, and define the index sets:

$$\begin{aligned} \bar{\mathcal{I}}^\nu &:= \{ i : |\bar{z}_i^{*,\nu}| < \delta,\ |\bar{\lambda}_i^{*,\nu}| \ge \delta \ \}, \\ \bar{\mathcal{J}}^\nu &:= \{ i : |\bar{z}_i^{*,\nu}| < \delta,\ |\bar{\lambda}_i^{*,\nu}| < \delta \ \}, \\ \bar{\mathcal{K}}^\nu &:= \{ i : |\bar{z}_i^{*,\nu}| \ge \delta,\ |\bar{\lambda}_i^{*,\nu}| < \delta \ \}, \end{aligned} \tag{3.21}$$

where $\delta > 0$ is a sufficiently small number. We assume that those index sets are independent of ν, i.e., $\bar{\mathcal{I}} := \bar{\mathcal{I}}^\nu, \bar{\mathcal{J}} := \bar{\mathcal{J}}^\nu, \bar{\mathcal{K}} := \bar{\mathcal{K}}^\nu$ for all ν. We define the following optimization problem for each leader ν:

$$\begin{array}{lcl} \widetilde{\mathrm{P}}^\nu(x^{-\nu}) : & \displaystyle\min_{x^\nu, y, z, \lambda, \mu} & \theta_\nu(x^\nu, x^{-\nu}, y) \\ & \text{s.t.} & g^\nu(x^\nu) \le 0,\ h^\nu(x^\nu) = 0, \\ & & \psi(x^\nu, x^{-\nu}, y, z, \lambda, \mu) = 0, \\ & & z_i = 0,\ \lambda_i \ge 0 \ \ (i \in \bar{\mathcal{I}}), \\ & & z_i = 0,\ \lambda_i = 0 \ \ (i \in \bar{\mathcal{J}}), \\ & & z_i \ge 0,\ \lambda_i = 0 \ \ (i \in \bar{\mathcal{K}}). \end{array}$$

Indeed, the above nonlinear optimization may be regarded as an approximation of a *tightened nonlinear program* for $\mathrm{PMPCC}^\nu(x^{-\nu})$ at $\bar{w}^{*,\nu}$ (which may be close to the feasible set of $\mathrm{PMPCC}^\nu(x^{-\nu})$) in the sense of Scheel and Scholtes [110].

Now the algorithm is stated in Algorithm 2.

Algorithm 2 Refined Gauss–Seidel Method

Input: Initial point $\tilde{w}^{(0)} := (\tilde{w}^{1,(0)}, \ldots, \tilde{w}^{N,(0)})$, which is supposed to be the last point obtained by Algorithm 1, the step size tolerance $\varepsilon' > 0$, and the maximum number of iterations $K_{\max}$.

Output: A stationary equilibrium point of the EPEC $(\mathrm{PMPCC}^\nu(x^{-\nu}))_{\nu=1}^N$, or the multi-L/F game.

1: Set $k = 0$.

2: For each ν, solve $\widetilde{\mathrm{P}}^\nu(\tilde{x}^{-\nu,(k)})$ to obtain the solution $\tilde{w}^{\nu,(k+1)}$, where

$$\tilde{x}^{-\nu,(k)} := (\tilde{x}^{1,(k+1)}, \ldots, \tilde{x}^{\nu-1,(k+1)}, \tilde{x}^{\nu+1,(k)}, \ldots, \tilde{x}^{N,(k)}).$$

3: Stop if $||\tilde{w}^{\nu,(k+1)} - \tilde{w}^{\nu,(k)}|| < \varepsilon'$ holds for all ν.

4: If $k < K_{\max}$, set $k := k+1$ and go to Step 2. If $k = K_{\max}$, terminate.

If a tuple of solutions $(\tilde{w}^{*,1}, \ldots, \tilde{w}^{*,N})$ is obtained and, for each leader ν, the KKT conditions of $\widetilde{\mathrm{P}}^\nu(\tilde{x}^{1,(k+1)}, \ldots, \tilde{x}^{\nu-1,(k+1)}, \tilde{x}^{\nu+1,(k+1)}, \ldots, \tilde{x}^{N,(k+1)})$ are sufficiently satisfied, the algorithm successfully terminates. Then by Definition 3.2.2, the point satisfies (at least) the weak stationarity of $\mathrm{PMPCC}^\nu(\tilde{x}^{-\nu})$ for all ν [110].

3.5 Application of multi-L/F games

In this section, we introduce an application of multi-L/F games. In the middle of the 1990s, the deregulation of electricity markets by governments stated mainly in Europe and the United States. Since then, the study of electricity markets has become popular [20, 47, 57, 71]. We introduce a wholesale market of electricity in terms of multi-L/F games or EPECs. The model we discuss is a simple model of competitive bidding under some macroeconomic regulation, which is an extension of [52][2].

In this model, we assume that there are two electricity firms labeled $\nu \in \{\mathrm{I}, \mathrm{II}\}$ and one market maker, called the independent system operator (ISO), who tries to correct the balance of demand and supply of electricity by paying the bid costs under the market clearing mechanism. The ISO also determines the price of electricity and then sells it to consumers. The two firms are competing with each other for market power in an electricity network with M nodes (consumers), and determine the bid price.

Let $x^\nu := (x_1^\nu, \ldots, x_M^\nu) \in I\!R^M$ be the bid parameter of firm ν in which the firm indirectly determines how much it sells the electricity to each node. Let $y := (y_1^{\mathrm{I}}, \ldots, y_M^{\mathrm{I}}, y_1^{\mathrm{II}}, \ldots, y_M^{\mathrm{II}}) \in I\!R^{2M}$ be the quantity of electricity, where y_i^ν means how much quantity of electricity the ISO buys from firm ν and supplies it to consumer i. The bid price function of firm ν is defined by $b^\nu(x^\nu, y) := \sum_{i=1}^M x_i^\nu y_i^\nu$. We assume that two firms produce electricity up to quantities a^{I} and a^{II}, and then send it to all nodes at the price $p_i(y_i^{\mathrm{I}}, y_i^{\mathrm{II}}) := \alpha_i - \beta_i(y_i^{\mathrm{I}} + y_i^{\mathrm{II}})$, where α_i and β_i are positive constants. The revenue for the ISO by selling electricity to node i is given as the cumulative sum from zero to $y_i^{\mathrm{I}} + y_i^{\mathrm{II}}$; see Figure 3.2. Thus, the ISO makes a profit given by $q_i(y_i^{\mathrm{I}}, y_i^{\mathrm{II}}) := \alpha_i(y_i^{\mathrm{I}} + y_i^{\mathrm{II}}) - \frac{\beta_i}{2}(y_i^{\mathrm{I}} + y_i^{\mathrm{II}})^2$.

[2] The author of [52] considered inequality constraints in the ISO's optimization problem. However, they omitted the constraints in the reformulation.

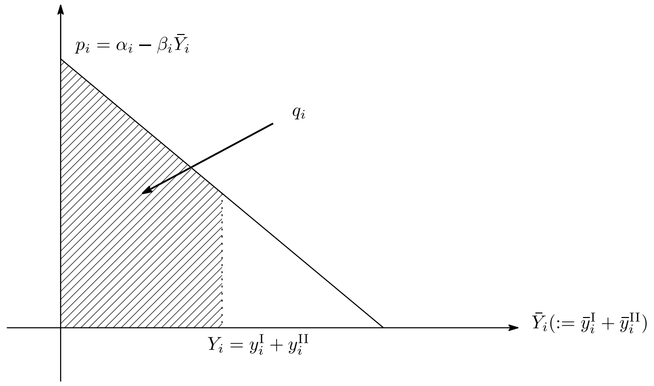


Figure 3.2: Price function p_i and revenue q_i

Firm ν needs to pay the transaction cost according to the bid parameter x_i^ν, which is defined by $t^\nu(x^\nu) := \frac{1}{2}\sum_{i=1}^M \tau_i^\nu (x_i^\nu)^2$ with a constant $\tau_i^\nu > 0$, and tries to maximize its revenue by bidding from the ISO minus transaction costs. Then the optimization problem of firm ν can be written as follows:

$$\begin{array}{ll} \min\limits_{x^\nu \in I\!R^M} & \theta_\nu(x^\nu, x^{-\nu}, y) := t^\nu(x^\nu) - b^\nu(x^\nu, y) \\ \text{s.t.} & x^\nu \in X^\nu, \end{array} \tag{3.22}$$

where X^ν is a nonempty strategy set.

On the other hand, the ISO also tries to maximize its revenue by selling electricity to consumers. Furthermore, we assume that some economic interventionism by governments works in the market to maintain the equilibrium between the quantities of electricity at each node i, or to reflect the ratio of quantities a^{I} and a^{II}, which is denoted by $\frac{\zeta_i}{2}\left(\frac{y_i^{\mathrm{I}}}{a^{\mathrm{I}}} - \frac{y_i^{\mathrm{II}}}{a^{\mathrm{II}}}\right)^2$, where $\zeta_i > 0$ is the interventionism parameter. Hence, the optimization problem of ISO can be written as follows:

$$\begin{array}{ll} \min\limits_{y \in I\!R^{2M}} & \sum\limits_{i=1}^M \left[\frac{\zeta_i}{2}\left(\frac{y_i^{\mathrm{I}}}{a^{\mathrm{I}}} - \frac{y_i^{\mathrm{II}}}{a^{\mathrm{II}}}\right)^2 - q_i(y_i^{\mathrm{I}}, y_i^{\mathrm{II}})\right] + b^{\mathrm{I}}(x^{\mathrm{I}}, y) + b^{\mathrm{II}}(x^{\mathrm{II}}, y) \\ \text{s.t.} & \sum\limits_{i=1}^M y_i^{\mathrm{I}} - a^{\mathrm{I}} \le 0, \\ & \sum\limits_{i=1}^M y_i^{\mathrm{II}} - a^{\mathrm{II}} \le 0, \\ & y \ge 0. \end{array}$$

Note that the ISO's problem is a convex optimization problem for the variable y. In particular, the objective function is strongly convex for any x. Then, the solution is uniquely determined for any given x by Lemma 3.4.2. Furthermore, the response is piecewise linear with respect to the variable x [82].

3.6 Numerical experiments

In this section, we present some numerical results to demonstrate the validity of the proposed method. We coded the algorithm in MATLAB 9.1.0 (2016b).

Examples 3.6.1–3.6.3 below are EPECs taken from [53], and player $\nu \in \{\mathrm{I}, \mathrm{II}\}$ solves the following MPEC with shared linear complementarity constraints:

$$\begin{array}{ll} \min\limits_{x^\nu \in \mathbb{R}^n, y \in \mathbb{R}^m} & \frac{1}{2}(x^\nu)^\top H_\nu x^\nu + (x^\nu)^\top G_\nu x^{-\nu} + (c^\nu)^\top y \\ \text{s.t.} & A_\nu x^\nu \le b^\nu \\ & 0 \le My + N_{\mathrm{I}} x^{\mathrm{I}} + N_{\mathrm{II}} x^{\mathrm{II}} + q \perp y \ge 0. \end{array}$$

In our examples, the matrix M is a P-matrix, which ensures that the solution to the linear complementarity system $0 \le My + N_{\mathrm{I}}x^{\mathrm{I}} + N_{\mathrm{II}}x^{\mathrm{II}} + q \perp y \ge 0$ for the variable y uniquely exists for any $x^{\mathrm{I}}, x^{\mathrm{II}}$. To implement the algorithm, we set $\rho_k := 20(k+1), \varepsilon = 0.01$ in Algorithm 1, and set $(x^{(0)}, y^{(0)}, z^{(0)}, \lambda^{(0)}, \mu^{(0)}) := (0, 0, 0, 0, 0)$. For the identification of active sets, we set $\delta := 0.01$ in (3.21). Then we set ε' =1.0e-07 in Algorithm 2. More detailed numerical results are shown in Appendix B.

Example 3.6.1. *The problem is Example* 6.2 *in* [53]*, and the data are given as follows. We modified the matrix* H_{I} *and* M *in* [53][3]*.*

$N = 2,\ n_\nu = 2\ (\nu = \mathrm{I}, \mathrm{II}),\ m = 2,$

$$H_{\mathrm{I}} = \begin{bmatrix} 3.6 & -1.8 \\ -1.8 & 7.2 \end{bmatrix},\ H_{\mathrm{II}} = \begin{bmatrix} 7.5 & -2.6 \\ -2.6 & 5.7 \end{bmatrix},\ G_{\mathrm{I}} = \begin{bmatrix} 1.1 & -1.3 \\ -2.4 & 1.6 \end{bmatrix},\ G_{\mathrm{II}} = \begin{bmatrix} -1.2 & 2.3 \\ 1.4 & -2.5 \end{bmatrix},$$

$$c^{\mathrm{I}} = \begin{bmatrix} -2.3 \\ -3.2 \end{bmatrix},\ c^{\mathrm{II}} = \begin{bmatrix} -2.5 \\ -2.4 \end{bmatrix},\ A_{\mathrm{I}} = \begin{bmatrix} 3.3 & -2.4 \end{bmatrix},\ A_{\mathrm{II}} = \begin{bmatrix} -2.5 & 2.1 \end{bmatrix},\ b^{\mathrm{I}} = -2.8,\ b^{\mathrm{II}} = -7.5,$$

$$M = \begin{bmatrix} 3.6 & -1.2 \\ -1.5 & 2.8 \end{bmatrix},\ N_{\mathrm{I}} = \begin{bmatrix} 2.1 & -1.3 \\ -3.4 & 2.3 \end{bmatrix},\ N_{\mathrm{II}} = \begin{bmatrix} -5.4 & 1.6 \\ -6.2 & 2.1 \end{bmatrix},\ q = \begin{bmatrix} 1.2 \\ 1.6 \end{bmatrix}.$$

Algorithm 1 produced the following solution after 10 iterations:

$$x^{\mathrm{I}} = \begin{bmatrix} -0.26175 \\ 0.80676 \end{bmatrix},\ x^{\mathrm{II}} = \begin{bmatrix} 2.69422 \\ -0.36402 \end{bmatrix},\ y^{\mathrm{I}} = \begin{bmatrix} 7.15349 \\ 8.51906 \end{bmatrix},\ y^{\mathrm{II}} = \begin{bmatrix} 7.15349 \\ 8.51906 \end{bmatrix},$$

and the distance between y^{I} and y^{II} is 1.0372e-06. We found that the solution already satisfies the KKT conditions for each leader's MPCC and hence an accurate solution was obtained without using Algorithm 2 in this example.

[3] The authors of [53] have confirmed that there were typos in the matrices H_{I} and M shown there. We use the correct data in our numerical experiments and obtained almost the same results as those of Table 1 in [53].

Example 3.6.2. *The problem is Example* 6.3 *in* [53] *and the data are given as follows:*

$$N = 2,\ n_\nu = 3\ (\nu = \mathrm{I}, \mathrm{II}),\ m = 3,$$

$$H_{\mathrm{I}} = \begin{bmatrix} 10.0 & 3.6 & 2.7 \\ 3.6 & 12.0 & -1.9 \\ 2.7 & -1.9 & 15.0 \end{bmatrix},\ H_{\mathrm{II}} = \begin{bmatrix} 12.0 & -1.2 & 3.1 \\ -1.2 & 10.0 & 2.5 \\ 3.1 & 2.5 & 8.0 \end{bmatrix},\ G_{\mathrm{I}} = \begin{bmatrix} 1.2 & 0.0 & -1.6 \\ 1.3 & -2.1 & 0.0 \\ -1.2 & 1.5 & 0.3 \end{bmatrix},$$

$$G_{\mathrm{II}} = \begin{bmatrix} 1.2 & 0.0 & -1.5 \\ 1.5 & 1.4 & 0.0 \\ -1.2 & 1.1 & -1.4 \end{bmatrix},\ c^{\mathrm{I}} = \begin{bmatrix} -3.6 \\ -2.7 \\ -4.8 \end{bmatrix},\ c^{\mathrm{II}} = \begin{bmatrix} -3.2 \\ -2.4 \\ -4.5 \end{bmatrix},\ A_{\mathrm{I}} = \begin{bmatrix} 1.6 & -1.3 & -1.2 \\ 1.2 & -1.7 & 1.3 \end{bmatrix},$$

$$A_{\mathrm{II}} = \begin{bmatrix} 1.3 & -1.5 & -1.2 \\ 1.8 & 1.2 & -1.3 \end{bmatrix},\ b^{\mathrm{I}} = \begin{bmatrix} -2.3 \\ -2.7 \end{bmatrix},\ b^{\mathrm{II}} = \begin{bmatrix} -1.4 \\ -1.6 \end{bmatrix},\ M = \begin{bmatrix} 5.6 & -1.2 & 1.5 \\ 3.2 & 7.2 & -2.4 \\ -1.8 & 2.5 & 6.4 \end{bmatrix},$$

$$N_{\mathrm{I}} = \begin{bmatrix} -1.1 & 0.0 & -1.2 \\ 1.5 & -1.0 & -0.3 \\ -1.4 & 0.0 & 1.3 \end{bmatrix},\ N_{\mathrm{II}} = \begin{bmatrix} -1.3 & 0.9 & -0.6 \\ -1.4 & 1.2 & 0.0 \\ 1.5 & -0.7 & 1.4 \end{bmatrix},\ q = \begin{bmatrix} -3.2 \\ -2.5 \\ -4.8 \end{bmatrix}.$$

Algorithm 1 produced the following solution after 6 iterations:

$$x^{\mathrm{I}} = \begin{bmatrix} -0.71047 \\ 0.99977 \\ -0.11371 \end{bmatrix},\ x^{\mathrm{II}} = \begin{bmatrix} -0.55146 \\ 0.04696 \\ 0.51055 \end{bmatrix},\ y^{\mathrm{I}} = \begin{bmatrix} 0.30697 \\ 0.54867 \\ 0.51239 \end{bmatrix},\ y^{\mathrm{II}} = \begin{bmatrix} 0.30697 \\ 0.54867 \\ 0.51239 \end{bmatrix},$$

and the distance between y^{I} and y^{II} is 9.2085e-07. We found that the solution already satisfies the KKT conditions for each leader's MPCC and hence an accurate solution was obtained without using Algorithm 2 in this example.

Example 3.6.3. *The problem data are the same as in Example* 3.6.2*, except*

$$c^{\mathrm{I}} = \begin{bmatrix} -3.6 \\ 2.7 \\ -4.8 \end{bmatrix},\ c^{\mathrm{II}} = \begin{bmatrix} 3.2 \\ -2.4 \\ 4.5 \end{bmatrix},\ q = \begin{bmatrix} -3.2 \\ 2.5 \\ -4.8 \end{bmatrix}.$$

Algorithm 1 produced the following solution after 16 iterations:

$$x^{\mathrm{I}} = \begin{bmatrix} -0.70535 \\ 1.00460 \\ -0.11212 \end{bmatrix},\ x^{\mathrm{II}} = \begin{bmatrix} -0.53491 \\ 0.04466 \\ 0.53135 \end{bmatrix},\ y^{\mathrm{I}} = \begin{bmatrix} 0.15341 \\ 0.00000 \\ 0.67569 \end{bmatrix},\ y^{\mathrm{II}} = \begin{bmatrix} 0.15345 \\ 0.00000 \\ 0.67566 \end{bmatrix},$$

and the distance between y^{I} and y^{II} is 4.6016e-05. Then we proceed to Algorithm 2 and obtained the following solution after 2 iterations:

$$x^{\mathrm{I},*} = \begin{bmatrix} -0.70535 \\ 1.00460 \\ -0.11212 \end{bmatrix}, x^{\mathrm{II},*} = \begin{bmatrix} -0.53491 \\ 0.04466 \\ 0.53135 \end{bmatrix}, y^{\mathrm{I},*} = \begin{bmatrix} 0.15345 \\ 0.00000 \\ 0.67566 \end{bmatrix}, y^{\mathrm{II},*} = \begin{bmatrix} 0.15345 \\ 0.00000 \\ 0.67566 \end{bmatrix},$$

and the distance between $y^{*,\mathrm{I}}$ and $y^{*,\mathrm{II}}$ is 1.1602e-09. We checked the KKT conditions for each leader's MPCC and confirmed that the solution is a B-stationary equilibrium.

Remark 3.6.4. *In Examples* 3.6.2 *and* 3.6.3*, the solutions reported in* [53] *are inaccurate. In our numerical experiments, we obtained more accurate solutions and confirmed that they are B-stationary equilibrium points of these examples.*

Next we consider a multi-L/F game consisting of N leaders and one follower. Leader ν solves the following optimization problem:

$$\begin{array}{ll} \min\limits_{x^\nu \in I\!R^{n_\nu}} & \frac{1}{2}(x^\nu)^\top H_\nu x^\nu + \sum\limits_{\nu'=1,\nu'\neq\nu}^{N} (x^\nu)^\top G_{\nu,\nu'} x^{\nu'} + (x^\nu)^\top D_\nu y \\ \text{s.t.} & A_\nu x^\nu \leq b^\nu. \end{array} \tag{3.23}$$

On the other hand, the follower solves the following optimization problem:

$$\begin{array}{ll} \min\limits_{y \in I\!R^m} & \gamma(x,y) := \frac{1}{2}y^\top M y + q^\top y - \sum\limits_{\nu=1}^{N} (x^\nu)^\top D_\nu y \\ \text{s.t.} & c^\top y + \sum\limits_{\nu=1}^{N} (d^\nu)^\top x^\nu + a \geq 0. \end{array} \tag{3.24}$$

We reformulate the multi-leader-follower game as the following EPEC.

$$\begin{array}{ll} \min\limits_{x^\nu \in I\!R^{n_\nu}, y \in I\!R^m, \lambda \in I\!R} & \frac{1}{2}(x^\nu)^\top H_\nu x^\nu + \sum\limits_{\nu'=1,\nu'\neq\nu}^{N} (x^\nu)^\top G_{\nu,\nu'} x^{\nu'} + (x^\nu)^\top D_\nu y \\ \text{s.t.} & A_\nu x^\nu \leq b^\nu, \\ & My + q - \sum\limits_{\nu=1}^{N} D_\nu^\top x^\nu - c\lambda = 0, \\ & 0 \leq \lambda \perp c^\top y + \sum\limits_{\nu=1}^{N} (d^\nu)^\top x^\nu + a \geq 0, \end{array}$$

where $\lambda \in I\!R$ is the Lagrange multiplier of the KKT conditions of the follower's optimization problem (3.24).

Example 3.6.5. *In this example, the data in* (3.23) *and* (3.24) *are given by*

$$N = 2,\ n_\nu = 2\ (\nu = \mathrm{I}, \mathrm{II}),\ m = 3,$$

$$H_{\mathrm{I}} = \begin{bmatrix} 2.1 & 1.2 \\ 1.2 & 2.8 \end{bmatrix},\ H_{\mathrm{II}} = \begin{bmatrix} 2.7 & 1.3 \\ 1.3 & 3.6 \end{bmatrix},\ G_{\mathrm{I,II}} = \begin{bmatrix} 1.8 & 1.4 \\ 1.5 & 2.7 \end{bmatrix},\ G_{\mathrm{II,I}} = \begin{bmatrix} 1.3 & 1.7 \\ 2.4 & 0.3 \end{bmatrix},$$

$$D_{\mathrm{I}} = \begin{bmatrix} 2.3 & 1.4 & 2.6 \\ 1.3 & 2.1 & 1.7 \end{bmatrix},\ D_{\mathrm{II}} = \begin{bmatrix} 2.5 & 1.9 & 1.4 \\ 1.3 & 2.4 & 1.6 \end{bmatrix},\ A_{\mathrm{I}} = \begin{bmatrix} 2.0 & 0.2 \\ -1.1 & 1.2 \\ 0.4 & 1.5 \end{bmatrix},\ b^{\mathrm{I}} = \begin{bmatrix} 1.8 \\ -0.6 \\ 1.4 \end{bmatrix},$$

$$A_{\mathrm{II}} = \begin{bmatrix} 1.3 & 0.5 \\ -1.2 & 2.8 \\ -1.4 & -0.3 \end{bmatrix},\ b_{\mathrm{II}} \begin{bmatrix} 0.9 \\ -0.9 \\ 1.8 \end{bmatrix},\ M = \begin{bmatrix} 2.5 & 1.8 & 0.2 \\ 1.8 & 3.6 & 2.1 \\ 0.2 & 2.1 & 4.6 \end{bmatrix},\ q = \begin{bmatrix} 1.4 \\ 2.6 \\ 2.1 \end{bmatrix},\ c = \begin{bmatrix} -0.7 \\ 0.5 \\ 1.5 \end{bmatrix},$$

$$d^{\mathrm{I}} = \begin{bmatrix} 2.5 \\ 2.8 \end{bmatrix},\ d^{\mathrm{II}} = \begin{bmatrix} 0.7 \\ -1.3 \end{bmatrix},\ a = -1.3,$$

where $H_{\mathrm{I}}, H_{\mathrm{II}}, M$ *are positive definite.*

Algorithm 1 produced the following solution after 6 iterations:

$$x^{\mathrm{I}} = \begin{bmatrix} 0.18728 \\ -0.32833 \end{bmatrix},\ x^{\mathrm{II}} = \begin{bmatrix} 0.41732 \\ -0.14258 \end{bmatrix},$$

$$y^{\mathrm{I}} = \begin{bmatrix} -0.71012 \\ -0.38255 \\ 0.64535 \end{bmatrix},\ y^{\mathrm{II}} = \begin{bmatrix} -0.70855 \\ -0.38256 \\ 0.64596 \end{bmatrix},\ \lambda^{\mathrm{I}} = 2.56077,\ \lambda^{\mathrm{II}} = 2.56094,$$

and the distance between $(y^{\mathrm{I}}, \lambda^{\mathrm{I}})$ and $(y^{\mathrm{II}}, \lambda^{\mathrm{II}})$ is 0.0017. Then we proceed to Algorithm 2 and obtained the following solution after 11 iterations:

$$x^{*,\mathrm{I}} = \begin{bmatrix} 0.18615 \\ -0.32936 \end{bmatrix},\ x^{*,\mathrm{II}} = \begin{bmatrix} 0.41828 \\ -0.14216 \end{bmatrix},$$

$$y^{*,\mathrm{I}} = \begin{bmatrix} -0.71278 \\ -0.38023 \\ 0.64692 \end{bmatrix},\ y^{*,\mathrm{II}} = \begin{bmatrix} -0.71278 \\ -0.38023 \\ 0.64692 \end{bmatrix},\ \lambda^{*,\mathrm{I}} = 2.56838,\ \lambda^{*,\mathrm{II}} = 2.56838,$$

$$\text{obj. value of leader I} : 0.25236,\ \gamma(x^{*,\mathrm{I}}, x^{*,\mathrm{II}}, y^{*,\mathrm{I}}) = 1.548565,$$

$$\text{obj. value of leader II} : -0.54157,\ \gamma(x^{*,\mathrm{I}}, x^{*,\mathrm{II}}, y^{*,\mathrm{II}}) = 1.548565,$$

and the distance between $(y^{*,\mathrm{I}}, \lambda^{*,\mathrm{I}})$ and $(y^{*,\mathrm{II}}, \lambda^{*,\mathrm{II}})$ is 8.1342e-08. We checked the KKT conditions for each leader's MPCC and confirmed that the solution is a B-stationary equilibrium of the multi-L/F game.

Example 3.6.6. *In this example, the data in* (3.23) *and* (3.24) *are given by*

$$N = 3,\ n_\nu = 3\ (\nu = \mathrm{I}, \mathrm{II}, \mathrm{III}),\ m = 3,$$

$$H_{\mathrm{I}} = \begin{bmatrix} 2.7 & 1.6 & 1.4 \\ 1.6 & 2.7 & 1.4 \\ 1.4 & 1.4 & 1.9 \end{bmatrix}, H_{\mathrm{II}} = \begin{bmatrix} 3.0 & 1.8 & 1.3 \\ 1.8 & 2.8 & 1.0 \\ 1.3 & 1.0 & 3.5 \end{bmatrix}, H_{\mathrm{III}} = \begin{bmatrix} 3.0 & 1.6 & 1.1 \\ 1.6 & 2.5 & 1.1 \\ 1.1 & 1.1 & 2.8 \end{bmatrix},$$

$$G_{\mathrm{I,II}} \begin{bmatrix} 0.4 & 1.8 & 1.5 \\ 0.8 & 0.6 & 1.2 \\ 0.5 & -0.7 & 0.8 \end{bmatrix}, G_{\mathrm{I,III}} = \begin{bmatrix} 0.7 & -0.9 & -0.8 \\ -0.8 & 1.2 & 0.3 \\ 0.8 & -0.6 & -0.3 \end{bmatrix}, G_{\mathrm{II,I}} = \begin{bmatrix} -0.5 & 1.2 & 1.4 \\ -1.0 & -0.7 & -0.7 \\ -0.2 & 1.4 & -0.9 \end{bmatrix},$$

$$G_{\mathrm{II,III}} = \begin{bmatrix} 0.7 & 0.4 & 1.4 \\ -0.3 & -0.1 & 1.6 \\ 2.0 & 0.7 & 0.1 \end{bmatrix}, G_{\mathrm{III,I}} = \begin{bmatrix} 0.2 & 0.2 & 1.1 \\ -0.4 & 0.7 & -0.9 \\ -0.8 & -0.7 & 1.2 \end{bmatrix},$$

$$G_{\mathrm{III,II}} = \begin{bmatrix} 0.7 & -0.7 & -0.4 \\ 1.2 & 1.2 & -0.3 \\ -0.9 & -0.6 & -0.3 \end{bmatrix}, D_{\mathrm{I}} = \begin{bmatrix} 1.7 & 0.8 & 2.0 \\ 0.2 & 0.3 & 1.3 \\ 0.1 & 2.2 & 0.6 \end{bmatrix}, D_{\mathrm{II}} = \begin{bmatrix} 2.4 & 0.9 & 2.0 \\ 1.1 & 0.5 & 1.1 \\ 2.1 & 2.0 & 2.4 \end{bmatrix},$$

$$D_{\mathrm{III}} = \begin{bmatrix} 1.4 & 2.0 & 1.6 \\ 1.2 & 0.6 & 1.7 \\ 2.3 & 1.5 & 1.4 \end{bmatrix}, A_{\mathrm{I}} = \begin{bmatrix} 1.9 & 0.6 & -0.2 \\ 0.9 & 1.8 & 1.7 \\ 0.1 & 1.6 & -1.9 \end{bmatrix}, b^{\mathrm{I}} = \begin{bmatrix} 2.3 \\ 1.5 \\ 2.2 \end{bmatrix},$$

$$A_{\mathrm{II}} = \begin{bmatrix} 0.8 & -1.8 & -0.1 \\ -0.6 & -0.2 & 1.7 \\ 1.6 & 2.3 & 1.6 \end{bmatrix}, b^{\mathrm{II}} = \begin{bmatrix} 1.0 \\ 0.4 \\ 2.4 \end{bmatrix}, A_{\mathrm{III}} = \begin{bmatrix} 1.5 & 1.1 & -1.3 \\ -1.7 & 1.0 & 1.6 \\ 0.4 & 0.8 & 2.4 \end{bmatrix},$$

$$b^{\mathrm{III}} = \begin{bmatrix} 2.0 \\ 2.8 \\ 1.8 \end{bmatrix}, M = \begin{bmatrix} 3.3 & 1.7 & 1.2 \\ 1.7 & 2.6 & 1.3 \\ 1.2 & 1.3 & 3.4 \end{bmatrix}, l = \begin{bmatrix} 1.7 \\ 2.1 \\ 1.7 \end{bmatrix}, A = \begin{bmatrix} 1.5 & 1.3 & 2.8 \end{bmatrix},$$

$$d^{\mathrm{I}} = \begin{bmatrix} 0.6 \\ 1.2 \\ 1.5 \end{bmatrix}, d^{\mathrm{II}} = \begin{bmatrix} 0.3 \\ 0.7 \\ 1.2 \end{bmatrix}, d^{\mathrm{III}} = \begin{bmatrix} 0.5 \\ 1.3 \\ 0.5 \end{bmatrix}, a = 1.4.$$

Algorithm 1 produced the following solution after 5 iterations:

$$
\begin{aligned}
&x^{\mathrm{I}} = \begin{bmatrix} 0.10548 \\ -0.20951 \\ 0.18411 \end{bmatrix}, \; x^{\mathrm{II}} = \begin{bmatrix} -0.05528 \\ 0.08292 \\ 0.19299 \end{bmatrix}, \; x^{\mathrm{III}} = \begin{bmatrix} 0.10182 \\ 0.16997 \\ -0.01800 \end{bmatrix}, \\
&y^{\mathrm{I}} = \begin{bmatrix} -0.08013 \\ -0.29607 \\ -0.08449 \end{bmatrix}, \; y^{\mathrm{II}} = \begin{bmatrix} -0.08203 \\ -0.29643 \\ -0.08629 \end{bmatrix}, \; y^{\mathrm{III}} = \begin{bmatrix} -0.08191 \\ -0.29343 \\ -0.08845 \end{bmatrix}, \\
&\lambda^{\mathrm{I}} = 0.00000, \; \lambda^{\mathrm{II}} = 0.00000, \; \lambda^{\mathrm{III}} = 0.00000,
\end{aligned}
$$

and the maximum distance between the three points $(y^{\mathrm{I}}, \lambda^{\mathrm{I}}), (y^{\mathrm{II}}, \lambda^{\mathrm{II}})$ and $(y^{\mathrm{III}}, \lambda^{\mathrm{III}})$ is 0.0051. Then we proceed to Algorithm 2 and obtained the following solution after 28 iterations:

$$
\begin{aligned}
&x^{*,\mathrm{I}} = \begin{bmatrix} 0.09671 \\ -0.18608 \\ 0.17755 \end{bmatrix}, \; x^{*,\mathrm{II}} = \begin{bmatrix} -0.05856 \\ 0.08488 \\ 0.18874 \end{bmatrix}, \; x^{*,\mathrm{III}} = \begin{bmatrix} 0.10744 \\ 0.15510 \\ -0.00855 \end{bmatrix}, \\
&y^{*,\mathrm{I}} = \begin{bmatrix} -0.08550 \\ -0.29371 \\ -0.08962 \end{bmatrix}, \; y^{*,\mathrm{II}} = \begin{bmatrix} -0.08550 \\ -0.29371 \\ -0.08962 \end{bmatrix}, \; x^{*,\mathrm{III}} = \begin{bmatrix} -0.08550 \\ -0.29371 \\ -0.08962 \end{bmatrix}, \\
&\lambda^{*,\mathrm{I}} = 0.00000, \; \lambda^{*,\mathrm{II}} = 0.00000, \; \lambda^{*,\mathrm{III}} = 0.00000, \\
&\text{obj. val. of leader I} : -0.11683, \; \gamma(x^{*,\mathrm{I}}, x^{*,-\mathrm{I}}, y^{*,\mathrm{I}}) = -0.22397, \\
&\text{obj. val. of leader II} : -0.14690, \; \gamma(x^{*,\mathrm{II}}, x^{*,-\mathrm{II}}, y^{*,\mathrm{II}}) = -0.22397, \\
&\text{obj. val. of leader III} : -0.13657, \; \gamma(x^{*,\mathrm{III}}, x^{*,-\mathrm{III}}, y^{*,\mathrm{III}}) = -0.22397,
\end{aligned}
$$

and the maximum distance between $(y^{*,\mathrm{I}}, \lambda^{*,\mathrm{I}}), (y^{*,\mathrm{II}}, \lambda^{*,\mathrm{II}})$ and $(y^{*,\mathrm{III}}, \lambda^{*,\mathrm{III}})$ is reduced to 4.4136e-07. We checked the KKT conditions for each leader's MPCC and confirmed that the solution is a B-stationary equilibrium point of the multi-L/F game.

Next we consider the electricity model introduced in Section 3.5.

Example 3.6.7. *The strategy set X^ν in* (3.22) *is given by $X^\nu := \{x^\nu \in \mathbb{R}^M \mid 0 \le x^\nu \le \sigma^\nu\}$. The numerical data are found in the first example in* [52] *as follows:*

$$
\begin{aligned}
&M\,(\textit{number of nodes}) : 2, \\
&\sigma^{\mathrm{I}} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \; \sigma^{\mathrm{II}} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \; \tau^{\mathrm{I}} = \begin{bmatrix} 1.2 \\ 1 \end{bmatrix}, \; \tau^{\mathrm{II}} = \begin{bmatrix} 1.3 \\ 1.5 \end{bmatrix}, \\
&\alpha_1 = 1.5, \; \beta_1 = 0.6, \; \alpha_2 = 1.8, \; \beta_2 = 0.7, \\
&a^{\mathrm{I}} = 1.2, \; a^{\mathrm{II}} = 1.8.
\end{aligned}
$$

When we set the interventionism parameters as $\zeta = (0.05, 0.05)$, we obtained the following solution after 4 iterations:

$$
\begin{aligned}
&x^{\mathrm{I}} = \begin{bmatrix} 0.54065 \\ 0.55121 \end{bmatrix}, \; x^{\mathrm{II}} = \begin{bmatrix} 0.64019 \\ 0.64516 \end{bmatrix}, \\
&y^{\mathrm{I}} = \begin{bmatrix} 0.59275 \\ 0.60725 \\ 0.79670 \\ 1.00330 \end{bmatrix}, \; y^{\mathrm{II}} = \begin{bmatrix} 0.58504 \\ 0.61496 \\ 0.80512 \\ 0.99488 \end{bmatrix}, \; \lambda^{\mathrm{I}} = \begin{bmatrix} 0.12354 \\ 0.03017 \end{bmatrix}, \; \lambda^{\mathrm{II}} = \begin{bmatrix} 0.12358 \\ 0.02683 \end{bmatrix},
\end{aligned}
$$

and the distance between $(y^{\mathrm{I}}, \lambda^{\mathrm{I}})$ and $(y^{\mathrm{II}}, \lambda^{\mathrm{II}})$ is 0.0929. Note that y^ν and λ^ν mean the leader ν's solutions of shared variables (y, λ). Then we proceed to Algorithm 2 and obtained the following solution after 7 iterations:

$$x^{*,\mathrm{I}} = \begin{bmatrix} 0.54027 \\ 0.55168 \end{bmatrix}, \; x^{*,\mathrm{II}} = \begin{bmatrix} 0.63999 \\ 0.64533 \end{bmatrix},$$

$$y^{*,\mathrm{I}} = \begin{bmatrix} 0.58774 \\ 0.61226 \\ 0.80285 \\ 0.99715 \end{bmatrix}, \; y^{*,\mathrm{II}} = \begin{bmatrix} 0.58774 \\ 0.61226 \\ 0.80285 \\ 0.99715 \end{bmatrix}, \; \lambda^{*,\mathrm{I}} = \begin{bmatrix} 0.12356 \\ 0.02687 \end{bmatrix}, \; \lambda^{*,\mathrm{II}} = \begin{bmatrix} 0.12356 \\ 0.02687 \end{bmatrix},$$

$$\text{obj. val. of firm I} : -0.32800, \;\; \gamma(x^{*,\mathrm{I}}, x^{*,-\mathrm{I}}, y^{*,\mathrm{I}}) = -1.683419,$$

$$\text{obj. va. of firm II} : -0.57874, \;\; \gamma(x^{*,\mathrm{II}}, x^{*,-\mathrm{II}}, y^{*,\mathrm{II}}) = -1.683419,$$

and the distance between $(y^{*,\mathrm{I}}, \lambda^{*,\mathrm{I}})$ and $(y^{*,\mathrm{II}}, \lambda^{*,\mathrm{II}})$ is 1.0795e-05. We checked the KKT conditions for each firm's MPCC and confirmed that the solution is a B-stationary point.

Next we observe the ratio of the electric supplies by the two firms $y_i^{\mathrm{I}} : y_i^{\mathrm{II}}$ at each node i. In this example, the ratio of the total electric supplies is $a^{\mathrm{I}} : a^{\mathrm{II}} = 1 : 1.5$. Let y^{I} and y^{II} denote $y^{\mathrm{I}} := (y_1^{\mathrm{I},*}, y_2^{\mathrm{I},*})$, $y^{\mathrm{II}} := (y_3^{\mathrm{I},*}, y_4^{\mathrm{I},*})$. As Table 3.1 shows, we found that the ratios $y_1^{\mathrm{I}} : y_1^{\mathrm{II}}$ and $y_2^{\mathrm{I}} : y_2^{\mathrm{II}}$ are getting closer to $1 : 1.5$ as ζ_i, $i = 1, 2$, increase. The objective function value of each firm was almost unchanged in these three cases. However, that of ISO was is increased as the sum of ζ_i increases.

Table 3.1: Ratio of quantities

ζ	$(0.05, 0.05)$	$(0.05, 0.5)$	$(0.5, 0.05)$	$(0.5, 0.5)$
$y_1^{\mathrm{I}} : y_1^{\mathrm{II}}$	1 : 1.36601	1 : 1.38229	1 : 1.3823	1 : 1.39730
$y_2^{\mathrm{I}} : y_2^{\mathrm{II}}$	1 : 1.62862	1 : 1.61336	1 : 1.6134	1 : 1.59831

3.7 Concluding remarks

In Chapter 3, we proposed a numerical method for solving multi-L/F games based on the penalty method and the nonlinear diagonalized Gauss–Seidel method. The method consists of two phases. The first phase of the method may be regarded as the identification of the active sets in the complementarity constraints, and the second phase is to find more accurate solutions with the active sets identified in the first phase. We discussed the convergence of the Gauss–Seidel penalty method to a C-/M-/B-stationary equilibrium point under respective suitable assumptions. Furthermore, we confirmed the validity of the algorithm through numerical experiments.

Acknowledgments

This chapter is based on A. Hori and M. Fukushima, "Gauss–Seidel method for multi-leader–follower games," *Journal of Optimization Theory and Applications*, **180**, 651–670, 2019. DOI: `https://doi.org/10.1007/s10957-018-1391-5`. Reproduced with permission from Springer Nature.

Chapter 4

Two-stage distributionally robust noncooperative games: Existence of Nash equilibrium and its application to Cournot–Nash competition

4.1 Introduction

Nash games in stochastic situations, which include random variables in the optimization problem of each player, have also been extensively studied for both pure and mixed strategies. It is natural to consider stochastic Nash games in which each player makes multistage decisions in response to changes in the conditions. For multistage stochastic games played with in finite action spaces, many approaches have been extensively studied. However, in continuous cases, because multistage variational inequalities have not been well developed until quite recently, few studies on variational inequality approaches have been conducted.

The notion of multistage stochastic variational inequalities was first explicitly presented by Rockafellar and Wets [107]. Rockafellar and Sun [103] then proposed a progressive hedging algorithm for solving it. The method corresponds to the proximal point algorithm [102] for maximal monotone problems with a linear transformation. As an advantage of progressive hedging, its subproblem can be computed in parallel, which drastically reduces the computational time. Based on their developments, several researchers have extended the idea into cases with continuously underlying random variables [15–17, 61, 62]. In [17, 61, 120, 131, 132], a variational inequality approach for two-stage stochastic Nash games was recently described.

Despite these various efforts on Nash games and variational inequalities under uncertainty, the ambiguity of the probability distributions have been ignored in both single- and multistage cases. Consider certain cases in which the available data may contain noise, and the number of sample data from an observation is small. An empirical distribution may be used and the above mentioned stochastic approaches can be applied. However, is the Nash equilibrium obtained through such approaches reliable? These questions have motivated researchers to consider distributionally robust stochastic Nash games, induced by the recent attention paid

Table 4.1: Recent works on distributionally robust Nash games

	Finite strategies (mixed strategies)[1]	Continuous pure strategies[2]
One-stage	Qu and Goh [96] Loizou [78, 79] Peng et al. [94]	Sun and Xu [119] Liu et al. [77]
Two-stage	————	Li et al. [72] (linear case) Chen et al. [17] (Cournot competition) **this chapter (nonlinear case)**

to distributionally robust optimization (DRO). Here, DRO aims to minimize the worst-case expected value of a measurable objective function from a set of probability distributions, called an *ambiguity set*. This model is supported by decision-making theory, which states that each player makes a decision based on the *maximin criteria*. The recent progress made in distributionally robust Nash games is shown in Table 4.1.

One-stage models have been widely applied in both finite and continuous cases. To our best knowledge, Qu and Goh [96] were the first to consider the distributional robustness in Nash stochastic games. Its extensions were then discussed by Loizou [78, 79] and Peng et al. [94], both of which demonstrated the reformulation into a tractable optimization problem under specific-structured ambiguity sets. Sun and Xu [119] applied the framework to a continuous case. Liu et al. [77] then showed the conditions for the existence of Nash equilibrium in one-stage distributionally robust continuous games. They also demonstrated that some can be reduced to classic stochastic games in special cases with ambiguity sets.

In comparison with one-stage models, studies on two-stage models are still in their infancy. Li et al. [72] considered a linear case in which each player solves a two-stage distributionally robust linear stochastic programming with a Wiesemann–Kuhn–Sim-type ambiguity set [122]. They demonstrated that under the *linear decision rule* [112], an equilibrium of the game can be obtained by solving a deterministic conic variational inequality. Note that the linear decision rule is merely an assumption tailored to numerical tractability; thus, particularly in nonlinear cases, adopting the rule into the games may be more inaccurate than in linear cases because of the complexity of the decision-making. Chen et al. [17] discussed a two-stage distributionally robust Cournot–Nash competition based on an "ex-post" equilibrium concept corresponding to a distribution-free robust Nash equilibrium presented in [2]. However, an ex-post equilibrium may not exist depending on the ambiguity set even when the two-stage DRO of each player is "well-posed" in a certain sense, for example, convexity and compactness; hence, it may be occasionally an ill-posed problem when one considers high uncertainty situations.

Therefore, the motivation of this chapter is to discuss a more general class of two-stage distributionally robust Nash games and give a more certain definition for the games based on the concept of Nash equilibria, which is unlike ex-post equilibrium. The contributions of this chapter are summarized as follows:

[1] We state that a mixed strategy with a finite strategy set is a probability assignment of taking from a set of finite pure strategy sets to a polyhedron.

[2] A case in which each player's pure strategy set is given by a subset of the Euclidean space as well as their decision variable takes continuous.

- We consider a two-stage (nonlinear) distributionally robust Nash game. We propose a definition of an equilibrium concept based on Nash equilibrium under a general setting and show the existence of an equilibrium point under the continuity, compactness, and convexity of each player's optimization.

- As an application, we revisit the two-stage distributionally robust Cournot–Nash competition introduced in [17] and show the existence of an equilibrium based on the definition of Nash equilibrium.

- We conduct a numerical experiment on the Cournot–Nash competition and investigate how distributional robustness affects the two-stage decisions of each player.

This chapter is organized as follows. In Section 4.2, we introduce the model and define a two-stage distributionally robust Nash equilibrium. Then we present the conditions for the existence of the equilibria. In Section 4.3, we consider a reformulation of the game into a variational inequality for analysis and for the construction of solution methods. In Section 4.4, we introduce a two-stage distributionally robust Cournot–Nash competition as an application of the game and provide the conditions for the existence of equilibrium in an economic sense. In Section 4.5, we report the results of some numerical experiments conducted to illustrate how distributional robustness affects the decision-making of each player. In Section 4.6, we provide some concluding remarks.

4.2 Two-stage distributionally robust Nash games and the existence of Nash equilibrium

In Section 1.1 we briefly introduced the two-stage stochastic and distributionally robust Nash games. Again we explain the model in more detail, and then we present sufficient conditions under which the equilibrium point exists.

4.2.1 Model and definition of Nash equilibrium

Let $\xi\colon \Omega \to \Xi \subset \mathbb{R}^t$ be a random vector and $\mathscr{P}(\Xi) := \{P \mid P(\Xi) = 1,\ P(\cdot) \geq 0\}$ be a set of any probability measures supported over Ξ, which is equipped with the measurable space $(\Xi, \mathcal{B}(\Xi))$.

We describe the N-person two-stage stochastic Nash game considered in this chapter. Hereafter, we use the following notations regarding player $\nu \in \{1, \dots, N\}$:

- $x^\nu \in \mathbb{R}^{n_\nu}$, $y^\nu\colon \Xi \to \mathbb{R}^{m_\nu}$: first- (here-and-now) and second-stage (wait-and-see) strategies of player ν, respectively;

- $X^\nu \subset \mathbb{R}^{n_\nu}$, $Y^\nu\colon X^\nu \times \Xi \rightrightarrows \mathbb{R}^{m_\nu}$: first- and second-stage strategy sets of player ν, respectively;

- $\theta_\nu\colon \mathbb{R}^n \to \mathbb{R}$, $\gamma_\nu\colon \mathbb{R}^m \times \mathbb{R}^n \times \Xi \to \mathbb{R}$: first- and second-stage cost functions of player ν, respectively;

- $\mathcal{Y}^\nu := \{y^\nu(\cdot) \mid y^\nu(\cdot) \in Y^\nu(x^\nu, \cdot) \quad \forall x^\nu \in X^\nu\}$: a set of all functions from Ξ to $Y^\nu(x^\nu, \xi) \subset \mathbb{R}^{m_\nu}$ for all $x^\nu \in X^\nu$;

Here, $n := \sum_{\nu=1}^{N} n_\nu$ and $m := \sum_{\nu=1}^{N} m_\nu$ are the sums of the dimensions for all players' strategy vectors at the first and second stages, respectively, and $X := \Pi_{\nu=1}^{N} X^\nu$ and $\mathcal{Y} := \Pi_{\nu=1}^{N} \mathcal{Y}^\nu$ are the Cartesian products of X^ν and $\mathcal{Y}^\nu$, $\nu = 1, \ldots, N$.

We suppose that player ν minimizes θ_ν at the first stage and then minimizes γ_ν at the second stage where $\xi \in \Xi$ is observed.

Player ν solves the following optimization problem at $\xi \in \Xi$ of the second stage:

$$\begin{aligned} Q_\nu(x^\nu, x^{-\nu}, \xi) := \min_{y^\nu(\xi) \in I\!R^{m_\nu}} \quad & \gamma_\nu(y^\nu(\xi), y^{-\nu}(\xi), x^\nu, x^{-\nu}, \xi) \\ \text{s.t.} \quad & y^\nu(\xi) \in Y^\nu(x^\nu, \xi), \end{aligned} \tag{4.1}$$

where $y^{-\nu}(\xi) \in I\!R^{m-m_\nu}$ indicates the other rival players' strategies. We call $Q_\nu(\cdot,\cdot,\xi)\colon I\!R^n \to I\!R$ a recourse function or an optimal value function at $\xi \in \Xi$. We also suppose that player ν does not know which of the scenarios will occur when choosing x^ν. Thus, the player tries to minimize the expected value $\mathbb{E}[Q_\nu(x^\nu, x^{-\nu}, \xi)]$ of the recourse function $Q_\nu(x^\nu, x^{-\nu}, \xi)$ before observing $\xi \in \Xi$ under a probability distribution.

However, our highest interest is for a case in which each player does not have strong confidence in the probability distribution (e.g., because of a lack of sample data to determine the distribution). Hence, we consider that they make their decisions based on the DRO framework, namely, the maximin criterion. That is, player ν minimizes $\sup_{P \in \mathscr{P}^\nu} \mathbb{E}_P[Q_\nu(x^\nu, x^{-\nu}, \xi)]$, where $\mathscr{P}^\nu \subset \mathscr{P}$ denotes an ambiguity set or a collection of probability distributions from the observed data. We assume that the probability distribution of ξ is independent of all players' decisions.

Consequently, the two-stage distributionally robust optimization (DRO) of player ν in the first stage is formulated as

$$\begin{aligned} \min_{x^\nu \in I\!R^{n_\nu}} \quad & \Theta_\nu(x^\nu, x^{-\nu}) := \{\theta_\nu(x^\nu, x^{-\nu}) + \sup_{P \in \mathscr{P}^\nu} \mathbb{E}_P[Q_\nu(x^\nu, x^{-\nu}, \xi)]\} \\ \text{s.t.} \quad & x^\nu \in X^\nu, \end{aligned} \tag{4.2}$$

where $x^{-\nu} \in I\!R^{n-n_\nu}$ is a tuple of the other rival players' strategies.

Note that when $\mathscr{P}^\nu$ is a singleton, (4.2) reduces to a two-stage stochastic programming problem [93]. Moreover, when $\mathscr{P}^\nu = \mathscr{P}(\Xi)$, and Ξ is compact, the model (4.2) coincides with the two-stage (distribution-free) robust optimization since $\mathbb{E}_P[f(\xi)] \le \max_{\xi \in \Xi} f(\xi)$ for any P and measurable functions f. Thus, the distributionally robust framework may also be regarded as a generalization of stochastic/robust approaches. In addition, Li et al. [72] dealt with a case in which θ_ν and γ_ν are linear with respect to player ν's decision variable, and X^ν and $Y^\nu(x^\nu, \xi)$ are the sets of linear constraints, which means our model is a generalization of theirs.

We consider an equilibrium of the game consisting of (4.2), $\nu = 1, \ldots, N$.

Definition 4.2.1. *The point $(x^*, y^*(\cdot)) \in X \times \mathcal{Y}$ is the simultaneous strategy of the first and second stages, respectively, and is called a* two-stage distributionally robust Nash equilibrium *(TSDRNE) if and only if the following conditions hold for all $\nu \in \{1, \ldots, N\}$:*

$$x^{*,\nu} \in \arg \min_{x^\nu \in X^\nu} \Theta_\nu(x^\nu, x^{*,-\nu}), \tag{4.3}$$

$$y^{*,\nu}(\xi) \in \arg \min_{y^\nu(\xi) \in Y^\nu(x^{*,\nu},\xi)} \gamma_\nu(y^\nu(\xi), y^{*,-\nu}(\xi), x^{*,\nu}, x^{*,-\nu}, \xi) \quad \forall \xi \in \Xi. \tag{4.4}$$

Note that a similar definition of equilibria is found in Zhang et al. [130, Definition 2.7]. However, under a first-stage condition (4.3), we consider the distributional robustness in Θ_ν.

When two-stage DRO (4.2) is linear for all ν, Li et al. [72] reformulate the two-stage distributionally robust Nash games into a deterministic variational inequality. The authors showed that the solution to the variational inequality satisfies (4.3) and (4.4), although they did not explicitly introduce the above definition.

4.2.2 Existence of two-stage Nash equilibrium

We provide some assumptions for the existence of a TSDRNE.

Assumption 4.2.2. *The following assertions hold for all $\nu \in \{1, \dots, N\}$:*

(a) *The function θ_ν is continuous, and $\theta_\nu(\cdot, x^{-\nu})$ is convex for each fixed $x^{-\nu}$;*

(b) *The feasible set $X^\nu \subset I\!R^{n_\nu}$ is compact and convex;*

(c) *The ambiguity set $\mathscr{P}^\nu$ is weakly compact*[3]*;*

(d) *The function γ_ν is continuous, and $\gamma_\nu(\cdot, y^{-\nu}(\xi), \cdot, x^{-\nu}, \xi)$ is jointly convex, i.e., γ_ν is convex with respect to $(x^\nu, y^\nu(\xi))$, for each fixed $x^{-\nu}$, $y^{-\nu}(\xi)$ and $\xi \in \Xi$;*

(e) *$Y^\nu \colon X^\nu \times \Xi \rightrightarrows I\!R^{m_\nu}$ is continuous, and $Y^\nu(x^\nu, \xi)$ is nonempty (namely, is a relatively complete recourse), compact, and convex for every $(x^\nu, \xi) \in X^\nu \times \Xi$.*

Remark 4.2.1. *The convexity of $Q_\nu(\cdot, x^{-\nu}, \xi)$ is stated from Theorem 34 by Birge and Louveaux* [11] *when the following feasible set $Y^\nu(x^\nu, \xi)$ is convex for any $x^\nu \in X^\nu$ and $\xi \in \Xi$:*

$$Y^\nu(x^\nu, \xi) := \left\{ y^\nu(\xi) \in I\!R^{m_\nu} \;\middle|\; \begin{array}{l} g_i^\nu(y^\nu(\xi), x^\nu, \xi) \le 0, \;\; i = 1, \dots, r'_\nu, \\ g_i^\nu(y^\nu(\xi), x^\nu, \xi) = 0, \;\; i = r'_\nu + 1, \dots, r_\nu. \end{array} \right\}.$$

where $g_i^\nu(\cdot, \cdot, \xi)$, $i = 1, \dots, r_\nu$, are continuous. The convexity of $Y^\nu(x^\nu, \xi)$ is guaranteed if the functions $g_i^\nu(\cdot, \cdot, \xi)$, $i = 1, \dots, r'_\nu$, are jointly convex and $g_i^\nu(\cdot, \cdot, \xi)$, $\nu = r'_\nu + 1, \dots, r_\nu$, are affine with respect to $(y^\nu(\xi), x^\nu)$.

For one-stage distributionally robust games, Liu et al. [77] do not assume the convexity of X^ν but only its compactness. However, as we can see in the following lemma, the two-stage model also requires the convexity of X^ν to ensure the convexity of the recourse function $Q_\nu(\cdot, x^{-\nu}, \xi)$.

Lemma 4.2.3. *Suppose that Assumption* 4.2.2 *holds. Then the recourse function $Q_\nu(\cdot, \cdot, \xi)$ is continuous, and $Q_\nu(\cdot, x^{-\nu}, \xi)$ is convex with respect to x^ν for every fixed $x^{-\nu}$ and $\xi \in \Xi$.*

Proof. It suffices to show the above assertion for a specific case in which player ν's two-stage DRO is independent of the other rival players' strategies; hence, we omit label ν and a tuple of rival players' strategies $x^{-\nu}$ and $y^{-\nu}(\xi)$.

By Assumptions 4.2.2-(d) and (e), the continuity of the recourse function Q holds.

Next, we show the convexity of $Q(\cdot, \xi)$ for each fixed $\xi \in \Xi$. Let us define $\mathcal{S}(\xi) := Y(\xi) \times X$, where $Y(\xi) := \{y(\xi) \mid y(\xi) \in Y(x, \xi) \;\; \forall x \in X\}$, and $\mathcal{S}(\xi)$ is convex for any $\xi \in \Xi$ by

[3] We state a space $\mathscr{A}$ is weakly compact if and only if every sequence $\{P_i\} \subset \mathscr{A}$ contains a subsequence $\{P_{i'}\}$ and $P^* \in \mathscr{A}$ such that $P_{i'}$ weakly converges to P^*.

Assumptions 4.2.2-(b) and (e). Suppose that $y^1(\xi) \in Y(x^1,\xi)$ and $y^2(\xi) \in Y(x^2,\xi)$ are optimal solutions to the second stage problem for fixed $x^1 \in X$ and $x^2 \in X$, respectively. For any $\alpha \in (0,1)$, let $(y'(\xi),x') = \alpha(y^1(\xi),x^1) + (1-\alpha)(y^2(\xi),x^2)$, and thus $(y'(\xi),x') \in \mathcal{S}(\xi)$ by the convexity of $\mathcal{S}(\xi)$. It follows from the joint convexity of γ in Assumption 4.2.2-(d) that

$$Q(x',\xi) \leq \gamma(y'(\xi),x',\xi) \leq \alpha\gamma(y^1(\xi),x^1,\xi) + (1-\alpha)\gamma(y^2(\xi),x^2,\xi) = \alpha Q(x^1,\xi) + (1-\alpha)Q(x^2,\xi).$$

Therefore, we have completed the proof. □

Combining the continuity of Q_ν by Lemma 4.2.3 and the relatively complete recourse (Assumption 4.2.2-(e)), $\mathbb{E}_P[Q_\nu(x^\nu,x^{-\nu},\xi)]$ is also continuous and bounded for any $x \in X$ and $P \in \mathscr{P}^\nu$. Thus, by the weak compactness of $\mathscr{P}^\nu$ (Assumption 4.2.2-(c)), there exists $P \in \mathscr{P}^\nu$ that achieves the maximum value of $\mathbb{E}_P[Q_\nu(x^\nu,x^{-\nu},\xi)]$ for any $x \in X$. Moreover, by the convexity of $Q_\nu(\cdot,x^{-\nu},\xi)$ from Lemma 4.2.3 and $\theta_\nu(\cdot,x^{-\nu})$ from Assumption 4.2.2-(a), we can easily show the convexity of $\Theta_\nu(\cdot,x^{-\nu})$, which is stated as follows.

Lemma 4.2.4. *Suppose that Assumption* 4.2.2 *holds. Then the objective function* Θ_ν *in* (4.2) *is continuous and convex with respect to* $x^\nu \in X^\nu$ *for any* $x^{-\nu}$.

We now show the existence of TSDRNE points under Assumption 4.2.2.

Theorem 4.2.5. *Suppose that Assumption* 4.2.2 *holds. Then the two-stage distributionally robust Nash equilibrium* $(x^*,y^*(\cdot)) \in X \times \mathcal{Y}$ *exists.*

Proof. Lemma 4.2.4 states the continuity of Θ_ν and convexity of $\Theta_\nu(\cdot,x^{-\nu})$ for each fixed $x^{-\nu}$. By Assumption 4.2.2-(b) and using Proposition 2.2.3, a Nash equilibrium $x^* \in X$ that satisfies (4.3) exists.

The existence of the second stage Nash equilibrium $y^*(\xi)$ can be likewise shown. By Assumptions 4.2.2-(d), (e), and using Proposition 2.2.3, there exists a Nash equilibrium $y^*(\xi)$, and the assertion holds for each fixed $\xi \in \Xi$, thus implying (4.4). Therefore, the proof is complete. □

4.3 Two-stage distributionally robust variational inequality under discrete probability distributions

This section presents a variational inequality reformulation for the condition of the TSDRNE stated in Definition 4.2.1.

Hereafter, we consider discrete probability cases, where the support set is given by $\Xi := \{\xi_1,\ldots,\xi_K\}$, and the probability of ξ_k for player ν is represented by $P^\nu(\xi_k)$. Let $P^\nu := (P^\nu(\xi_1),\ldots,\ P^\nu(\xi_K)) \in I\!R^K$ and $P := (P^1,\ldots,P^N) \in I\!R^{NK}$. We also denote $\Delta := \{p \in I\!R^K \mid \sum_{k=1}^K p_k = 1,\ p \geq 0\}$ as the polyhedron of the probabilities supported on Ξ, and $\mathscr{P} := \Pi_{\nu=1}^N \mathscr{P}^\nu$ is the Cartesian product of $\mathscr{P}^\nu$, $\nu = 1,\ldots,N$.

First, we give the definition of a two-stage distributionally robust variational inequality, which is inspired by the one-stage version by Sun et al. [118].

Definition 4.3.1. *Let* $\mathscr{P}^\nu \subset \Delta$, $\nu = 1,\ldots,N$ *be convex ambiguity sets supported on* Ξ. *Suppose that* $X \subset I\!R^n$ *is a nonempty closed convex set and that* $Y(\xi) \subset I\!R^m$ *is also a nonempty* *closed convex set for each fixed* $\xi \in \Xi$. *A* two-stage distributionally robust variational inequality *(TSDRVI) is to find a pair* $(x^*, y^*(\cdot)) \in X \times \mathcal{Y}$ *and* $P^* \in \mathscr{P}$ *satisfying the following inclusions:*

$$0 \in \mathbb{E}_{P^*}[F(x^*, y^*(\xi), \xi)] + \mathcal{N}_X(x^*), \tag{4.5}$$

$$0 \in G(x^*, y^*(\xi), \xi) + \mathcal{N}_{Y(\xi)}(y^*(\xi)) \quad \forall \xi \in \Xi, \tag{4.6}$$

$$P^{*,\nu} \in \arg\max_{P \in \mathscr{P}^\nu} \mathbb{E}_P[f_\nu(x^{*,\nu}, x^{*,-\nu}, \xi)], \quad \nu = 1, \ldots, N, \tag{4.7}$$

where $F\colon I\!R^n \times I\!R^m \times \Xi \to I\!R^n$, $G\colon I\!R^n \times I\!R^m \times \Xi \to I\!R^m$, $f_\nu\colon I\!R^{n_\nu} \times \Xi \to I\!R$, *and* $\mathcal{N}_X(x) := \{z \in I\!R^n \mid \langle z, w - x\rangle \le 0\ \forall w \in X\}$ *denotes a normal cone of* X *at* $x \in X$. *The expectation operator* $\mathbb{E}_P[F(x, y(\xi), \xi)]$ *is defined as*

$$\mathbb{E}_P[F(x, y(\xi), \xi)] := (\mathbb{E}_{P^1}[F_1(x, y(\xi), \xi)], \ldots, \mathbb{E}_{P^N}[F_N(x, y(\xi), \xi)]),$$

where $\mathbb{E}_{P^\nu}[F_\nu(x, y(\xi), \xi)] := \sum_k P^\nu(\xi_k) \cdot F_\nu(x, y(\xi_k), \xi_k)$ *and* $F_\nu := I\!R^n \times I\!R^m \times \Xi \to I\!R^{n_\nu}$.

When the ambiguity set $\mathscr{P}^\nu$ is a singleton for all ν, the above TSDRVI reduces to two-stage stochastic variational inequalities [107, 120].

Note that for a variational inequality $0 \in \hat{F}(x) + \mathcal{N}_X(x)$, when X is given by the nonnegative orthant $I\!R^n_+$, the inclusion reduces to the complementarity $0 \le x \perp \hat{F}(x) \ge 0$, which suggests that the class of variational inequalities includes complementarity problems.

Remark 4.3.1. *Chen et al.* [17] *considered a two-stage distributionally robust linear complementarity problem in the following form:*

$$0 \le x \perp Ax + \mathbb{E}_P[B(\xi)y(\xi)] + q_1 \ge 0 \quad \forall P \in \mathscr{P}, \tag{4.8}$$

$$0 \le y(\xi) \perp M(\xi)y(\xi) + N(\xi)x + q_2(\xi) \ge 0 \quad \text{for } P\text{–almost every } \xi \in \Xi, \tag{4.9}$$

where $A \in I\!R^{n\times n}, q_1 \in I\!R^n, B\colon I\!R^t \to I\!R^{n\times m}, M\colon I\!R^t \to I\!R^{m\times m}, N\colon I\!R^t \to I\!R^{n\times m}$ *and* $q_2\colon I\!R^t \to I\!R^m$ *are continuous matrix/vector-valued mappings. When* $X = I\!R^n_+$ *and* $Y(\xi) \equiv I\!R^m_+$ *for all* $\xi \in \Xi$ *in* (4.5) *and* (4.6)*, the difference between TSDRVI* (4.5)–(4.7) *and* (4.8)–(4.9) *is based on whether the solution* $x \in I\!R^n_+$ *to the first stage variational inequality (or the linear complementarity problem) depends on the probability distributions. In addition, we should emphasize that the solution to* (4.8) *and* (4.9)*, called an* 'ex-post' equilibrium, *is also the solution to TSDRVI* (4.5)–(4.6)*. However, the converse does not hold in general. This suggests that the notion of TSDRVI* (4.5)–(4.7) *is weaker than that of the ex-post equilibrium formulation* (4.8)-(4.9)*, which is shown in Appendix* C.

We now show the main assertion of this section.

Theorem 4.3.2. *Suppose that Assumption* 4.2.2 *holds and that* $\mathscr{P}^\nu \subset \Delta$ *is convex and compact for all* ν. *The tuple* $(x^*, y^*(\cdot))$ *is a TSDRNE if and only if there exists* $P^* \in \mathscr{P}$ *such that* $(x^*, y^*(\cdot))$ *satisfies the following TSDRVI:*

$$0 \in F_\theta(x^*) + \mathbb{E}_{P^*}[v(x^*, \xi)] + \mathcal{N}_X(x^*), \tag{4.10}$$

$$0 \in G(x^*, y^*(\xi), \xi) + \mathcal{N}_{Y(x^*,\xi)}(y^*(\xi)) \quad \forall \xi \in \Xi, \tag{4.11}$$

$$P^{*,\nu} \in \arg\max_{P \in \mathscr{P}^\nu} \mathbb{E}_P[Q_\nu(x^{*,\nu}, x^{*,-\nu}, \xi)], \quad \nu = 1, \ldots, N, \tag{4.12}$$

where

$$F_\theta(x) := [\nabla_{x^\nu}\theta_\nu(x^\nu, x^{-\nu})]_{\nu=1}^N,$$
$$v(x,\xi) \in \partial_{x^1}Q_1(x^1, x^{-1}, \xi) \times \cdots \times \partial_{x^N}Q_N(x^N, x^{-N}, \xi) \subset I\!R^n,$$
$$G(x, y(\xi), \xi) := [\nabla_{y^\nu(\xi)}\gamma_\nu(y^\nu(\xi), y^{-\nu}(\xi), x^\nu, x^{-\nu}, \xi)]_{\nu=1}^N,$$
$$Y(x,\xi) := \Pi_{\nu=1}^N Y^\nu(x^\nu, \xi).$$

Proof. We first show the 'only if' part. Let $(x^*, y^*(\cdot))$ be a TSDRNE. By the compactness of $\mathscr{P}^\nu$, for all $\nu \in \{1, \ldots, N\}$, there exists a probability vector $P^{*,\nu} \in \mathscr{P}^\nu$ that achieves the maximum of $\mathbb{E}_P[Q_\nu(x^{*,\nu}, x^{*,-\nu}, \xi)]$.

Now, we focus on the νth two-stage DRO, that is, problem (4.2) and the right-hand side of (4.1). Since $(x^{*,\nu}, y^{*,\nu}(\cdot))$ is the global optimal solution to (4.2), we have

$$0 \in \nabla_{x^\nu}\theta_\nu(x^{*,\nu}, x^{*,-\nu}) + \partial_{x^\nu}\mathbb{E}_{P^{*,\nu}}[Q_\nu(x^{*,\nu}, x^{*,-\nu}, \xi)] + \mathcal{N}_{X^\nu}(x^{*,\nu}), \tag{4.13}$$
$$0 \in \nabla_{y^\nu(\xi)}\gamma_\nu(y^{*,\nu}(\xi), y^{*,-\nu}(\xi), x^{*,\nu}, x^{*,-\nu}, \xi) + \mathcal{N}_{Y^\nu(x^{*,\nu},\xi)}(y^{*,\nu}(\xi)) \quad \forall \xi \in \Xi, \tag{4.14}$$
$$P^{*,\nu} \in \arg\max_{P\in\mathscr{P}^\nu} \mathbb{E}_P[Q_\nu(x^{*,\nu}, x^{*,-\nu}, \xi)]. \tag{4.15}$$

Since $Q_\nu(\cdot, x^{-\nu}, \xi)$ is convex from Lemma 4.2.3, it is Clarke regular [21, Definition 2.3.4]. It then follows from [21, Corollary 3 (p.40)] that, for all $P^\nu \in \mathscr{P}^\nu$,

$$\partial_{x^\nu}\mathbb{E}_{P^\nu}[Q_\nu(x^{*,\nu}, x^{*,-\nu}, \xi)] = \mathbb{E}_{P^\nu}[\partial_{x^\nu}Q_\nu(x^{*,\nu}, x^{*,-\nu}, \xi)].$$

Hence, (4.13) implies that there exists $v^\nu(x^{*,\nu}, x^{*,-\nu}, \xi) \in \partial_{x^\nu}Q_\nu(x^{*,\nu}, x^{*,-\nu}, \xi)$ such that

$$0 \in \nabla_{x^\nu}\theta_\nu(x^{*,\nu}, x^{*,-\nu}) + \mathbb{E}_{P^{*,\nu}}[v^\nu(x^{*,\nu}, x^{*,-\nu}, \xi)] + \mathcal{N}_{X^\nu}(x^{*,\nu}). \tag{4.16}$$

It then follows from (4.16), (4.14), and (4.15) that (4.10)–(4.12) holds. This completes the proof of the 'only if' part.

Next, we show the 'if' part. Suppose that there exists $(x^*, y^*(\cdot))$ and $P^* \in \mathscr{P}$ that satisfy TSDRVI (4.10)–(4.12). In the first-stage variational inequality (4.10), by the definition of the normal cone, (4.10) implies (4.16) for all ν. Since first-stage problem (4.2) is convex by Assumption 4.2.2-(b), (4.16) is the necessary and sufficient condition for the optimality of player ν's first-stage optimization (4.2). Then, $x^* \in X$ satisfies condition (4.3) of the first-stage Nash equilibrium. Likewise, we can show that $y^*(\xi)$ satisfies condition (4.4) of the second-stage Nash equilibrium for every $\xi \in \Xi$. Therefore, $(x^*, y^*(\cdot))$ is a TSDRNE. The proof is completed. □

The calculation of ∂Q_ν depends on the properties of the second-stage optimization problem in practice, which is beyond the scope of this thesis to mention a general approach to obtain an explicit form for ∂Q_ν; for example, see Ralph and Xu [99], and Bonnans and Shapiro [12] for advanced discussions. Meanwhile, as we will see in Section 4.4, an explicit form for ∂Q_ν is given under specific assumptions.

Unfortunately, solution methods for solving two-stage stochastic variational inequalities under a distributional ambiguity have yet to be established. However, a TSDRNE can be obtained using recent results on two-stage stochastic variational inequalities. For such an example, we will present the progressive hedging algorithm [103] in Section 4.5.1 for the numerical experiments.

4.4 Application to Cournot–Nash competition

In this section, we consider a two-stage distributionally robust Cournot–Nash competition in an oligopoly market and investigate the conditions for the existence of market equilibria in an economic sense.

First, let us distinguish conventional works from our study relating to two-stage Cournot–Nash competitions under uncertainty. The most similar two-stage distributionally robust Cournot–Nash competition was analyzed by Chen et al. [17] from the viewpoint of an ex-post equilibrium, i.e., the solution to (4.8) and (4.9) as shown in Remark 4.3.1. However, they do not provide the sufficient conditions for the existence of a solution satisfying (4.8) and (4.9). Also note that another similar two-stage stochastic Cournot–Nash competition under which the probability distribution is exactly known was developed by Zhang et al. [130] and Xu et al. [123]. The authors introduced a class of stochastic equilibrium problems with equilibrium constraints (SEPEC) to analyze the competition and find market equilibria. However, solving SEPECs has some numerical difficulties, such as nonconvexity and nonmonotonicity, and hence it is difficult to find a global Nash equilibrium in general. In addition, an SEPEC approach requires the second-stage problem to have a unique solution for any given first-stage variables and $\xi \in \Xi$ in practice. As seen later, we do not need such uniqueness of the second-stage equilibrium for our model.

Next we provide an example of two-stage stochastic Cournot–Nash competitions in real-world applications. Consider an oligopoly market of N firms who compete in investing and supplying homogeneous products, e.g., crude oil, steel, and some other resources whose qualities are independent of producers. For example, in a crude oil market, the majority of the world's crude oil is supplied by a few large oil exporting countries and they are viewed collectively as a finite number of large agents from which price-taking consumers purchase the product at the same price [61]. Jiang et al. [61] investigated the crude oil market through a two-stage stochastic Cournot–Nash competition model. In the model, each major oil-producing country takes action in every term (e.g., daily, weekly, and monthly) on how much oil to produce and supply. Finally, they found that the stochastic Cournot–Nash competition model is suitable to reproduce, predict, and potentially capable to explain stable market shares of crude oil through numerical experiments.

We now move on to our model. In the first stage, firm ν determines the upper capacity $x^\nu \geq 0$ of the product without certain information regarding the market demand in the future, and the investment cost of firm ν for the capacity is defined by $\theta_\nu(x^\nu)$. In the second stage, firm ν decides how much the products to supply to the market, not to exceed x^ν, which denotes $y^\nu(\xi) \in [0, x^\nu]$. Here, suppose that the future demand is only characterized by an inverse demand function $p(q(\xi), \xi)$, where $q(\xi) := \sum_{\nu=1}^{N} y^\nu(\xi)$ is the aggregate quantity of products in the market under scenario $\xi \in \Xi$. Firm $\nu \in \{1, \dots, N\}$ tries to maximize the following worst-case expected profit from the ambiguity set $\mathscr{P}^\nu$:

$$\begin{aligned} \max_{x^\nu \in \mathbb{R}} \quad & \Theta_\nu(x^\nu) := \{ \min_{P \in \mathscr{P}^\nu} \mathbb{E}_P[Q_\nu(x^\nu, \xi)] - \theta_\nu(x^\nu) \} \\ \text{s.t.} \quad & x^\nu \geq 0, \end{aligned} \tag{4.17}$$

where $Q_\nu(x^\nu, \xi)$ is the optimal value of the profit maximization under scenario $\xi \in \Xi$:

$$\begin{aligned} \max_{y^\nu(\xi) \in \mathbb{R}} \quad & p(q(\xi), \xi) y^\nu(\xi) - H_\nu(y^\nu(\xi), \xi) \\ \text{s.t.} \quad & 0 \leq y^\nu(\xi) \leq x^\nu, \end{aligned} \tag{4.18}$$

where $H_\nu(y^\nu(\xi),\xi)$ is the cost of supplying the product to the market.

Following Xu [123], we now make certain assumptions to ensure the convexity of the problems.

Assumption 4.4.1. *For all ν, $\theta_\nu(x^\nu)$ and $H_\nu(y^\nu(\xi),\xi)$ are twice continuously differentiable with respect to x^ν and $y^\nu(\xi)$, respectively, and*

$$\begin{aligned}
&\theta'_\nu(x^\nu) \geq 0,\ \theta''_\nu(x^\nu) \geq 0 \quad \text{for} \quad x^\nu \geq 0,\\
&H'_\nu(y^\nu(\xi),\xi) \geq 0,\ H''_\nu(y^\nu(\xi),\xi) \geq 0 \quad \text{for} \quad y^\nu(\xi) \geq 0 \text{ and } \xi \in \Xi.
\end{aligned}$$

Assumption 4.4.2. *The inverse demand function $p(q(\xi),\xi)$ satisfies the following conditions:*

(a) $p(q(\xi),\xi)$ is twice continuously differentiable in $q(\xi)$ and $p'_q(q(\xi),\xi) < 0$ for $q(\xi) \geq 0$ and $\xi \in \Xi$;

(b) $p'_q(q(\xi),\xi) + q(\xi)p''_{qq}(q(\xi),\xi) \leq 0$ for $q(\xi) \geq 0$ and $\xi \in \Xi$.

Assumptions 4.4.1 and 4.4.2–(a) indicate the monotonicity of the investment/supply cost functions and inverse demand functions, respectively. To explain the meanings of Assumption 4.4.2–(b), consider a monopoly market with an extraneous supply $\bar{q} \geq 0$. If the monopoly's output is $q(\xi)$, then its revenue at demand scenario ξ is $q(\xi)p(q(\xi)+\bar{q},\xi)$. The marginal revenue is $p(q(\xi)+\bar{q},\xi) + q(\xi)p'_q(q(\xi)+\bar{q},\xi)$. The rate of change of this marginal revenue with respect to the increase in the extraneous supply $\bar{q}$ is $p'_q(q(\xi)+\bar{q},\xi)+q(\xi)p''_{qq}(q(\xi)+\bar{q},\xi)$. Assumption 4.4.2–(b) implies that this rate is not positive when $\bar{q} = 0$ for any $\xi \in \Xi$. In other words, any extraneous supply will potentially reduce the monopoly's marginal revenue in any demand scenario; see Sherali et al. [113] for a similar explanation for a deterministic leader–followers' market.

Under the above assumptions, Xu [123] established the following result.

Proposition 4.4.3 (Xu [123, Proposition 2.4])**.** *Suppose that Assumption 4.4.2 holds. Then the following assertions hold: For a fixed $\bar{q} \geq 0$,*

i. *$p'_q(q(\xi)+\bar{q},\xi) + qp''_{qq}(q(\xi)+\bar{q},\xi) \leq 0$ for $q(\xi) \geq 0$ and $\xi \in \Xi$;*

ii. *$q(\xi)p(q(\xi)+\bar{q},\xi)$ is strictly concave in $q(\xi)$ for $q(\xi) \geq 0$ and $\xi \in \Xi$.*

Using the above results, it is easy to see that the first- and second-stage optimization problems of firm ν are convex with respect to x^ν and $y^\nu(\xi)$ for $\xi \in \Xi$, respectively (i.e., the objective functions and strategy sets at each stage are concave and convex, respectively).

Under convexity Assumptions 4.4.1 and 4.4.2, the necessary and sufficient condition for the optimality of problem (4.17) of firm ν is written as

$$0 \in \theta'_\nu(x^\nu) - \mathbb{E}_{P^\nu}[\partial_{x^\nu} Q_\nu(x^\nu,\xi)] + \mathcal{N}_{[0,\infty)}(x^\nu), \tag{4.19}$$

$$\begin{aligned}
0 \leq y^\nu(\xi) \perp H'_\nu(y^\nu(\xi),\xi) &- p(q(\xi),\xi)\\
&- p'_q(q(\xi),\xi))y^\nu(\xi) + \lambda^\nu(\xi) \geq 0 \quad \forall \xi \in \Xi,
\end{aligned} \tag{4.20}$$

$$0 \leq \lambda^\nu(\xi) \perp x^\nu - y^\nu(\xi) \geq 0 \quad \forall \xi \in \Xi, \tag{4.21}$$

$$P^\nu \in \arg\min_{P \in \mathscr{P}^\nu} \mathbb{E}_P[Q_\nu(x^\nu,\xi)], \tag{4.22}$$

where $\lambda^\nu(\xi)$ is the Lagrange multiplier for $y^\nu(\xi) \le x^\nu$. By the result of [17], $\partial_{x^\nu} Q_\nu(x^\nu, \xi)$ is calculated as

$$\partial_{x^\nu} Q_\nu(x^\nu, \xi) = \begin{cases} \lambda^\nu(\xi) & \text{if } x^\nu > 0, \\ \{\lambda^\nu(\xi) \mid \lambda^\nu(\xi) \ge [p(q(\xi), \xi) - H'_\nu(0, \xi)]_+\} & \text{if } x^\nu = 0. \end{cases}$$

Note that when $x^\nu > 0$, $\lambda^\nu(\xi)$ uniquely exists from the linear independence constraint qualification of problem (4.18).

Summarizing both cases finally yields the following two-stage distributionally robust variational inequality:

$$0 \le x^\nu \perp \theta'_\nu(x^\nu) - \mathbb{E}_P[\lambda^\nu(\xi)] \ge 0, \tag{4.23}$$

$$\begin{aligned} 0 \le y^\nu(\xi) \perp H'_\nu(y^\nu(\xi), \xi) - p(q(\xi), \xi) - \\ p'_{y^\nu}(q(\xi), \xi)) y^\nu(\xi) + \lambda^\nu(\xi) \ge 0 \quad \forall \xi \in \Xi, \end{aligned} \tag{4.24}$$

$$0 \le \lambda^\nu(\xi) \perp x^\nu - y^\nu(\xi) \ge 0 \quad \forall \xi \in \Xi, \tag{4.25}$$

$$P^\nu \in \arg\min_{P \in \mathscr{P}^\nu} \mathbb{E}_P[Q_\nu(x^\nu, \xi)]. \tag{4.26}$$

Using the variational inequality reformulation above, we can state the following condition for the nontriviality of x^ν in the competition.

Proposition 4.4.4. *For firm ν, if the following inequality holds, then $x^\nu > 0$.*

$$\theta'_\nu(0) < \mathbb{E}_P[p(q^{-\nu}(\xi), \xi) - H'_\nu(0, \xi)] \quad \forall P \in \mathscr{P}^\nu, \tag{4.27}$$

where $q^{-\nu}(\xi) := \sum_{\nu' \ne \nu}^N y^{\nu'}(\xi)$.

Proof. We show by contradiction: Assume that $x^\nu = 0$. The optimality condition for two-stage DRO of firm ν is written as (4.23)–(4.26). By the assumption of $x^\nu = 0$, we have $y^\nu(\xi) \equiv 0$ for all $\xi \in \Xi$, and then we can reduce (4.23)–(4.25) to

$$\begin{aligned} \theta'_\nu(0) - \mathbb{E}_P[\lambda^\nu(\xi)] \ge 0, \\ \lambda^\nu(\xi) \ge [p(q^{-\nu}(\xi), \xi) - H'_\nu(0, \xi)]_+ \quad \forall \xi \in \Xi. \end{aligned} \tag{4.28}$$

It follows from the second inequality that

$$\mathbb{E}_P[\lambda^\nu(\xi)] \ge \mathbb{E}_P[p(q^{-\nu}(\xi), \xi) - H'_\nu(0, \xi)] \quad \forall P \in \mathscr{P}^\nu.$$

Hence, the above inequality, (4.28), and (4.27) yield

$$0 \le \theta'_\nu(0) - \mathbb{E}_P[\lambda^\nu(\xi)] \le \theta'_\nu(0) - \mathbb{E}_P[p(q^{-\nu}(\xi), \xi) - H'_\nu(0, \xi)] < 0 \quad \forall P \in \mathscr{P}^\nu.$$

This is a contradiction, and hence the proof is complete. □

Proposition 4.4.4 indicates that if the worst-case expected marginal profit of firm ν at $x^\nu = 0$ ($y^\nu(\xi) \equiv 0$ for all $\xi \in \Xi$) is greater than the first-stage marginal cost, the firm has an incentive to invest at least a small number of products.

To ensure the existence of a TSDRNE in the Cournot–Nash competition, we need the following extra assumption.

Assumption 4.4.5. *For all ν, there exists $\bar{x}^\nu \geq 0$ and $P^\nu \in \mathscr{P}^\nu$ such that*

$$\mathbb{E}_{P^\nu}[p(y^\nu(\xi), \xi)] < \theta'_\nu(x^\nu) \quad \text{for } x^\nu \geq \bar{x}^\nu \ (0 \leq y^\nu(\xi) \leq x^\nu \ \forall \xi \in \Xi), \tag{4.29}$$

where $P^\nu \in \arg\min_{P \in \mathscr{P}^\nu} \mathbb{E}_P[Q_\nu(x^\nu, \xi)]$.

The above assumption suggests that (even if firm ν monopolized the market) the marginal cost exceeds the expected market price under the worst-case probability for a certain $x^\nu \geq \bar{x}^\nu$. In practice, Assumption 4.4.5 is not special when the law of increasing marginal costs holds in the market.

We now show the existence of the market equilibrium.

Theorem 4.4.6. *Suppose that Assumptions 4.4.1, 4.4.2, and 4.4.5 hold. Then a TSDRNE of the Cournot–Nash competition exists.*

Proof. Note that this game satisfies Assumption 4.2.2 except for the compactness of the first-stage constraint set of each firm. If the set is compact, then a TSDRNE of the competition exists from Theorem 4.2.5. Thus, it suffices to show that there exists a finite number M_ν such that

$$\sup_{x^\nu \geq 0} \Theta_\nu(x^\nu) = \max_{0 \leq x^\nu \leq M_\nu} \Theta_\nu(x^\nu).$$

Consider optimality condition (4.23)–(4.26). By Assumptions 4.4.1 and 4.4.2, we have

$$\begin{aligned}[p'_q(q(\xi), \xi) y^\nu(\xi) + p(q(\xi), \xi) - H'_\nu(y^\nu(\xi), \xi)]_+ &\leq p(q(\xi), \xi) \\ &\leq p(y^\nu(\xi), \xi) \quad \forall \xi \in \Xi,\end{aligned}$$

where the last inequality holds from the monotonicity of $p(\cdot, \xi)$ for all $\xi \in \Xi$. Then the above inequalities also hold regarding their expected values under the probability distribution $P^\nu \in \arg\min_{P \in \mathscr{P}^\nu} \mathbb{E}_P[Q_\nu(x^\nu, \xi)]$.

Now we can take $\lambda^\nu(\xi) = [p'_q(q(\xi), \xi) y^\nu(\xi) + p(q(\xi), \xi) - H'_\nu(y^\nu(\xi), \xi)]_+$ for all $\xi \in \Xi$ because the solution of second-stage TSDRVI (4.24)–(4.25) is $\lambda^\nu(\xi) \geq [p'_q(q(\xi), \xi) y^\nu(\xi) + p(q(\xi), \xi) - H'_\nu(y^\nu(\xi), \xi)]_+ \geq 0$, and $\mathbb{E}_P[\lambda^\nu(\xi)] \leq \theta'_\nu(x^\nu)$ in (4.23). Then it follows from Assumption 4.4.5 that

$$\mathbb{E}_{P^\nu}[\lambda^\nu(\xi)] \leq \mathbb{E}_{P^\nu}[p(y^\nu(\xi), \xi)] < \theta'_\nu(x^\nu) \quad \text{for } x^\nu \geq \bar{x}^\nu.$$

This implies that there exists $x^\nu \leq \bar{x}^\nu$ such that the first stage optimality (4.23) holds; that is,

$$\sup_{x^\nu \geq 0} \Theta_\nu(x^\nu) = \max_{0 \leq x^\nu \leq \bar{x}^\nu} \Theta_\nu(x^\nu).$$

We have thus completed the proof. □

Roughly speaking, Theorem 4.4.6 means that under the law of increasing marginal costs and Assumption 4.4.5 in the market, the firm has no incentive to invest more than $\bar{x}^\nu$; hence, an equilibrium point exists in the market.

In fact, the result of the existence of a TSDRNE can be easily obtained by assuming $x^\nu \leq M_\nu$ for a large number $M_\nu > 0$ in the first stage constraint of firm ν. As a benefit of the absence of such a capacity limit, the first-stage complementarity condition does not require an additional Lagrange multiplier for the upper constraint $x^\nu \leq M_\nu$. In addition, Assumption 4.4.5 is meaningful for analyzing the detailed economic behavior of each firm.

4.5 Numerical experiments

In this section, we report some results of numerical experiments conducted to investigate the characteristics of the TSDRNE in the two-stage distributionally robust Cournot–Nash competition presented in Section 4.4. First, we consider a more specific case and provide a TSDRVI reformulation of the competition. We then provide a solution method for solving the TSDRVI. Finally, we report the results of the numerical experiments conducted.

Consider a duopoly market, i.e., $N = 2$. Recall that each firm competes with each other to maximize the worst-case expected profit: In the first stage, firm $\nu \in \{1, 2\}$ solves

$$\max_{x^\nu \geq 0} \quad \big\{ \min_{P \in \mathscr{P}^\nu} \mathbb{E}_P[Q_\nu(x^\nu, \xi)] - \theta_\nu(x^\nu) \big\}, \tag{4.30}$$

and the second-stage optimization is defined as

$$Q_\nu(x^\nu, \xi) \coloneqq \max_{0 \leq y^\nu(\xi) \leq x^\nu} \{ p(q(\xi), \xi) y^\nu(\xi) - H_\nu(y^\nu(\xi), \xi) \} . \tag{4.31}$$

In this experiment, the cost functions and inverse demand are given as follows:

$$\begin{aligned}
\theta_\nu(x^\nu) &\coloneqq \frac{1}{2} a_\nu (x^\nu)^2 + b_\nu x^\nu + c_\nu \quad (a_\nu, b_\nu, c_\nu > 0), \\
H_\nu(y^\nu(\xi), \xi) &\coloneqq \frac{1}{2} \eta_\nu(\xi)(y^\nu(\xi))^2 + \zeta_\nu(\xi) y^\nu(\xi) + s_\nu(\xi) \\
&\qquad\qquad\qquad (\eta_\nu(\xi), \zeta_\nu(\xi), s_\nu(\xi) > 0) \quad \forall \xi \in \Xi, \\
p(q(\xi), \xi) &\coloneqq \alpha(\xi) - \beta(\xi) q(\xi) \quad (\alpha(\xi) > \beta(\xi) > 0) \quad \forall \xi \in \Xi.
\end{aligned}$$

Note that Assumptions 4.4.1, 4.4.2, and 4.4.5 hold for this case; hence, a TSDRNE of the Cournot–Nash competition exists in this competition.

In this thesis, we use a Kullback–Leibler (KL) divergence-based ambiguity set because this has been widely used in the literature on distributionally robust optimization [56, 63, 76, 77] as well as the numerical tractability[4]. Suppose that Ξ consists of K scenarios, i.e., $\Xi = \{\xi_1, \ldots, \xi_K\}$. Let the ambiguity set $\mathscr{P}^\nu$ of firm ν be defined as follows:

$$\mathscr{P}^\nu \coloneqq \{ P \in \Delta \mid \mathbb{D}_{\mathrm{KL}}(P \| P_0^\nu) \leq \rho_\nu \}, \tag{4.32}$$

where $\rho_\nu \geq 0$, P_0^ν is a nominal (empirical) probability distribution of firm ν, and $\mathbb{D}_{\mathrm{KL}}(P \| P_0^\nu)$ denotes the KL divergence by

$$\mathbb{D}_{\mathrm{KL}}(P \| P_0^\nu) \coloneqq \sum_{k=1}^{K} P(\xi_k) \cdot \log \left(\frac{P(\xi_k)}{P_0^\nu(\xi_k)} \right),$$

where $P(\xi_k)$ and $P_0^\nu(\xi_k)$ are the probabilities when $\xi = \xi_k$ under the distributions P and P_0^ν, respectively, and also $\mathscr{P}^\nu$ is a convex set.

[4] Besides KL divergence, a distributionally robust optimization with a class of measures called ϕ-divergence is known to be recast as a convex conic optimization; see Table 3 of Rahimian and Mehrotra [97] for more detail.

Since problems (4.30) and (4.31) are convex with respect to x^ν and $y^\nu(\xi)$, respectively, the necessary and sufficient optimality conditions of firm ν can be written as follows:

$$0 \le x^{*,\nu} \perp a_\nu x^{*,\nu} + b_\nu - \mathbb{E}_{P^{*,\nu}}[\lambda^{*,\nu}(\xi)] \ge 0, \tag{4.33}$$

$$\begin{aligned}0 \le y^{*,\nu}(\xi) \perp (\eta_\nu(\xi) + 2\beta(\xi))y^{*,\nu}(\xi) + \lambda^{*,\nu}(\xi) + \beta(\xi) \sum_{\nu' \neq \nu} y^{*,\nu'}(\xi) - \\ \alpha(\xi) + \zeta_\nu(\xi) \ge 0 \quad \forall \xi \in \Xi,\end{aligned} \tag{4.34}$$

$$0 \le \lambda^{*,\nu}(\xi) \perp x^{*,\nu} - y^{*,\nu}(\xi) \ge 0 \quad \forall \xi \in \Xi, \tag{4.35}$$

$$P^{*,\nu} \in \arg \min_{P \in \mathscr{P}^\nu} \mathbb{E}_P[Q_\nu(x^{*,\nu}, \xi)]. \tag{4.36}$$

Gathering the systems (4.33)–(4.35) and (4.36) of all firms, the condition for the TSDRNE can be written as the following TSDRVI:

$$\begin{aligned}0 \le x^* \perp Ax^* - \mathbb{E}_{P^*}[\lambda^*(\xi)] + b \ge 0, \\ 0 \le y^*(\xi) \perp \Pi(\xi)y^*(\xi) + \lambda^*(\xi) + r(\xi) \ge 0 \quad \forall \xi \in \Xi, \\ 0 \le \lambda^*(\xi) \perp x^* - y^*(\xi) \ge 0 \quad \forall \xi \in \Xi,\end{aligned} \tag{4.37}$$

$$P^{*,\nu} \in \arg \min_{P \in \mathscr{P}^\nu} \mathbb{E}_P[Q_\nu(x^{*,\nu}, \xi)], \quad \nu = 1, \dots, N. \tag{4.38}$$

where

$$\begin{aligned}A &:= \mathrm{diag}(a_1, a_2),\ b := (b_1, b_2)^\top,\ \Pi(\xi) := \mathrm{diag}(\eta_1(\xi), \eta_2(\xi)) + \beta(\xi)(I + \mathbf{1}\mathbf{1}^\top) \\ \mathbf{1} &:= (1,1)^\top \in I\!R^2,\ r(\xi) := (r_1(\xi), r_2(\xi)),\ r_\nu(\xi) := \zeta_\nu(\xi) - \alpha(\xi),\ \nu = 1, 2.\end{aligned}$$

Note that since each firm's optimization problem and the ambiguity set $\mathscr{P}^\nu$ are convex, a solution of TSDRVI (4.37) and (4.38) is a TSDRNE of the Cournot–Nash competition by Theorem 4.3.2.

4.5.1 Solution method using progressive hedging

Here, an algorithm based on progressive hedging is presented to solve TSDRVI (4.37) and (4.38).

The progressive hedging algorithm (PHA) for multistage stochastic variational inequalities was recently developed by Rockafellar and Sun [103] as an extension of [105] for solving multistage stochastic programming. The benefit of using the PHA is to reduce the computational complexity by solving the variational inequalities for each scenario $\xi \in \Xi$ in parallel.

Preserving the computational efficiency, our idea is to alternately solve (4.37) and (4.38) because if we fix a probability distribution, the two-stage stochastic variational inequality (4.37) can be solved by the PHA. The main loop of the proposed method is given in Algorithm 3, and its inner loop is shown in Algorithm 4, where the regularization parameter $\sigma > 0$ determines the performance of the algorithm and depends on numerical instances. In line 3 of Algorithm 4, it is well known that the subproblem always has a unique solution regarding $(x, y(\cdot), \lambda(\cdot))$ by Theorem 2.3.6 since the mapping of the variational inequality is strongly monotone by proximal terms. Note that Algorithm 3 has no guarantee of convergence in general, while Algorithm 4 ensures the convergence of sequence generated under the monotonicity of SVI (4.37) for a fixed probability distribution [103].

In practice, the sample data ξ used in the algorithm needs to follow a *reference probability distribution* which is an empirical distribution in a certain sense; hence, we assume that the ambiguity set consists of probability distributions that include the reference probability distribution. We may also suppose that the true probability distribution is included in $\mathscr{P}^\nu$.

Algorithm 3 Main loop: Solve TSDRVI

Input: $(x^0, y^0(\xi_1), \dots, y^0(\xi_K), \lambda^0(\xi_1), \dots, \lambda^0(\xi_K))$.
Output: $(x^*, y^*(\xi_1), \dots, y^*(\xi_K), \lambda^*(\xi_1), \dots, \lambda^*(\xi_K))$

1: Set $j = 0$, $x^{(j)} = x^0$, $y^{(j)}(\xi_k) = y^0(\xi_k)$, and $\lambda^{(j)}(\xi_k) = \lambda^0(\xi_k)$ for all k.
2: For each ν, solve

$$P^{\nu,(j+1)} \in \arg\min_{P \in \mathscr{P}^\nu} \mathbb{E}_P[p(q^{(j)}(\xi), \xi) y^{\nu,(j)}(\xi) - H_\nu(y^{\nu,(j)}(\xi), \xi)],$$

where $q^{(j)}(\xi) \coloneqq \sum_{\nu=1}^N y^{\nu,(j)}(\xi)$.
3: Stop if $(x^{(j)}, y^{(j)}(\xi_1), \dots, y^{(j)}(\xi_K), \lambda^{(j)}(\xi_1), \dots, \lambda^{(j)}(\xi_K))$ and $P^{(j+1)}$ approximately satisfies (4.37) and (4.38).
4: Solve two-stage SVI (4.37) for a fixed $P^{(j+1)}$ using Algorithm 4 and obtain a solution $(x^{(j+1)}, y^{(j+1)}(\xi_1), \dots, y^{(j+1)}(\xi_K), \lambda^{(j+1)}(\xi_1), \dots, \lambda^{(j+1)}(\xi_K))$.
5: Set $j \coloneqq j + 1$ and go to line 2.

Since the $\mathscr{P}^\nu$ is given in the KL ambiguity set, by utilizing the result [56], the worst-case probability distribution in line 2 of Algorithm 3 is efficiently computed as follows:

$$P^{\nu,(j+1)}(\xi_k) \coloneqq P_0^\nu(\xi_k) \cdot \frac{g(y^{\nu,(j)}(\xi_k), y^{-\nu,(j)}(\xi_k), \alpha^{\nu,(j+1)})}{\mathbb{E}_{P_0}\left[g(y^{\nu,(j)}(\xi), y^{-\nu,(j)}(\xi), \alpha^{\nu,(j+1)})\right]} \quad \forall k,\ \nu = 1, \dots, N,$$

where

$$\alpha^{\nu,(j+1)} \coloneqq \arg\min_{\alpha \ge 0} \left\{\alpha \log \mathbb{E}_{P_0}[g(y^{\nu,(j)}(\xi), y^{-\nu,(j)}(\xi), \alpha)] + \alpha\rho_\nu\right\},$$

$$g(y^\nu(\xi), y^{-\nu}(\xi), \alpha) \coloneqq \exp\left(\frac{p(q(\xi), \xi) y^\nu(\xi) - H_\nu(y^\nu(\xi), \xi)}{\alpha}\right).$$

Recall that $P_0^\nu(\xi_k)$ denotes the probability when ξ takes ξ_k under the nominal probability distribution of firm ν, which is empirically estimated in practice. We adopt the following stopping criterion for Algorithm 3:

$$\|\min(z, Mz + h)\|_2 \le \epsilon = 1.0 \times 10^{-6},$$

where

$$M \coloneqq \begin{bmatrix} A & B_1 & \dots & B_K \\ E & D_1 & & \\ \vdots & & \ddots & \\ E & & & D_K \end{bmatrix},\ h \coloneqq [b^\top, \hat{h}_1^\top, \dots, \hat{h}_K^\top]^\top$$

$$B_k \coloneqq [0,\ -\mathrm{diag}(P^1(\xi_k), P^2(\xi_k))],\ E \coloneqq [0,\ I]^\top,\ D_k \coloneqq \begin{bmatrix} \Pi(\xi_k) & I \\ -I & 0 \end{bmatrix},$$

$$\hat{h}_k \coloneqq (r(\xi_k)^\top, 0^\top)^\top,\ z \coloneqq (x, y(\xi_1), \dots, y(\xi_K), \lambda(\xi_1), \dots, \lambda(\xi_K)).$$

The tolerance of the inner iteration in Algorithm 4 is 10^{-8}, and the regularization parameter is set as $\sigma = 0.8$. Note that $\min(z, Mz + h) = 0$ is equivalent to two-stage stochastic variational inequality (4.37) for a fixed probability distribution.

Algorithm 4 Progressive Hedging Algorithm

Input: $(x^{(j),0}, y^{(j),0}(\xi_1), \ldots, y^{(j),0}(\xi_K), \lambda^{(j),0}(\xi_1), \ldots, \lambda^{(j),0}(\xi_K))$, $P^{(j+1)}$, and $w^{(0)}(\xi_1), \ldots, w^{(0)}(\xi_K)$ such that $\sum_k w^{(0)}(\xi_k) = 0$.

Output: $(x^{(j+1)}, y^{(j+1)}(\xi_1), \ldots, y^{(j+1)}(\xi_K), \lambda^{(j+1)}(\xi_1), \ldots, \lambda^{(j+1)}(\xi_K))$

1: Set $\ell = 0$, $x^{(\ell)}(\xi_k) = x^{(j),0}$, $y^{(\ell)}(\xi_k) = y^{(j),0}(\xi_k)$ and $\lambda^{(\ell)}(\xi_k) = \lambda^{(j),0}(\xi_k)$ for all k. Let $x^{(\ell)} := x^{(j),0}$.

2: Stop if $(x^{(\ell)}, y^{(\ell)}(\xi_1), \ldots, y^{(\ell)}(\xi_K), \lambda^{(\ell)}(\xi_1), \ldots, \lambda^{(\ell)}(\xi_K))$ satisfies a stopping criterion.

3: For each scenario $\xi \in \Xi$, obtain a unique solution $(\hat{x}^{(\ell)}(\xi), \hat{y}^{(\ell)}(\xi), \hat{\lambda}^{(\ell)}(\xi))$ to

$$\begin{aligned}
&0 \le x(\xi) \perp Ax(\xi) - \lambda(\xi) + b + w^{(\ell)}(\xi) + \sigma(x(\xi) - x^{(\ell)}(\xi)) \ge 0\\
&0 \le y(\xi) \perp \Pi(\xi)y(\xi) + \lambda(\xi) + r(\xi) + \sigma(y(\xi) - y^{(\ell)}(\xi)) \ge 0\\
&0 \le \lambda(\xi) \perp x(\xi) - y(\xi) + \sigma(\lambda(\xi) - \lambda^{(\ell)}(\xi)) \ge 0
\end{aligned}$$

4: Let $\bar{x}^{\nu,(\ell+1)} := \sum_{k=1}^{K} P^{\nu,(j+1)}(\xi_k) \cdot \hat{x}^{\nu,(\ell)}(\xi_k)$ for $\nu = 1, 2$, and for all k, let

$$\begin{gathered}
x^{(\ell+1)}(\xi_k) := x^{(\ell+1)} := \bar{x}^{(\ell+1)},\ y^{(\ell+1)}(\xi_k) := \hat{y}^{(\ell)}(\xi_k),\ \lambda^{(\ell+1)}(\xi_k) := \hat{\lambda}^{(\ell)}(\xi_k),\\
w^{(\ell+1)}(\xi_k) := w^{(\ell)}(\xi_k) + \sigma(\hat{x}^{(\ell)}(\xi_k) - \bar{x}^{(\ell+1)})
\end{gathered}$$

5: Set $\ell := \ell + 1$ and go to line 2.

4.5.2 Numerical results

We use the PATH5 solver [27] for MATLAB to obtain the solution to the variational inequalities in line 3 in Algorithm 4, the tolerance of which is 10^{-9}. Throughout this chapter, we carry out the experiments on a computer with an Intel Xeon 2.10 GHz CPU, 128 GB of RAM, and 64-bit Windows 10 OS.

We use the following data for the model:

$$\begin{gathered}
a_1 = 0.0874, a_2 = 0.1767, b_1 = 1.7162, b_2 = 1.9021, c_1 = 0.5212, c_2 = 0.8314,\\
\eta_1(\xi) = 0.2065 + 0.1\xi_1, \eta_2(\xi) = 0.2700 + 0.1\xi_1,\\
\zeta_1(\xi) = 0.5598 + 0.1\xi_1, \zeta_2(\xi) = 0.9748 + 0.1\xi_1,\\
s_1(\xi) = 0.1602 + 0.1\xi_1, s_2(\xi) = 0.1932 + 0.1\xi_1, \alpha(\xi) = 20 + 5\xi_1, \beta(\xi) = 2 + \xi_2,
\end{gathered}$$

where $\xi \in \Xi = [-1, 1]^3$, $\xi_i := -1 + 2\xi_i^0$, $i = 1, 2, 3$, and the reference probability distribution of the random variable $\xi^0 \in [0, 1]^3$ is uniform. The constant of each data except $\alpha(\xi)$ and $\beta(\xi)$ is randomly generated so that Assumptions 4.4.1 and 4.4.2 hold, and we verified that the sequence $\{(x^{(j)}, y^{(\nu)}(\cdot), \lambda^{(\nu)}(\cdot), P^{(\nu)})\}_j$ generated by Algorithm 3 converges in this numerical case.

We set the sample size $K = 60$, and ρ_ν, $\nu = 1, 2$, in ambiguity set (4.32) are set as $\rho_1 = \rho \ge 0$, $\rho_2 = 2 - \rho \ge 0$. To avoid a sample dependence of ξ, we conducted 30 trials for each $\rho \in \{0, 0.1, 0.2, \ldots, 2.0\}$ by changing the sample data ξ during each trial and plotting the average results.

The numerical results are shown in Figures 4.1a–4.1d. Figures 4.1a and 4.1b represent the profit and strategy $(x^{*,\nu}, \bar{y}^{*,\nu})$ for each player, respectively, where $\bar{y}^{*,\nu} := (y^{*,\nu}(\xi_1) + \cdots + y^{*,\nu}(\xi_K))/K$ denotes the average value of $y^{*,\nu}(\xi_k)$, $k = 1, \ldots, K$. These values decrease as

ρ_ν increases. It is noteworthy that the difference $x^{*,\nu} - \bar{y}^{*,\nu}$ also decreases as ρ_ν increases; that is, when ρ_ν is large, the difference $x^{*,\nu} - y^{*,\nu}(\xi)$ is zero for any scenario $\xi \in \Xi$. In addition, as Figure 4.1c indicates, the decrease of $x^{*,\nu} - \bar{y}^{*,\nu}$ also affects the slope of the curve for the average shadow price $\bar{\lambda}^{*,\nu} := (\lambda^{*,\nu}(\xi_1) + \cdots + \lambda^{*,\nu}(\xi_K))/K$, i.e, a mean of marginal revenues when a corresponding firm unilaterally increases a unit of production (with fixed rival production and supply). When the difference $x^{*,\nu} - \bar{y}^{*,\nu}$ is zero, and the difference $x^{*,-\nu} - \bar{y}^{*,-\nu}$ of the rival firm is positive, the curve of $\bar{\lambda}^{*,\nu}$ is decreasing, which means that firm ν has a passive involvement in the market because the rival firm has much more market information than firm ν. This also implies a monopolization by the rival firm. Note that under a catastrophe (e.g., economic crisis) in the market, the rival firm may incur significant losses, whereas firm ν does not lose much in comparison.

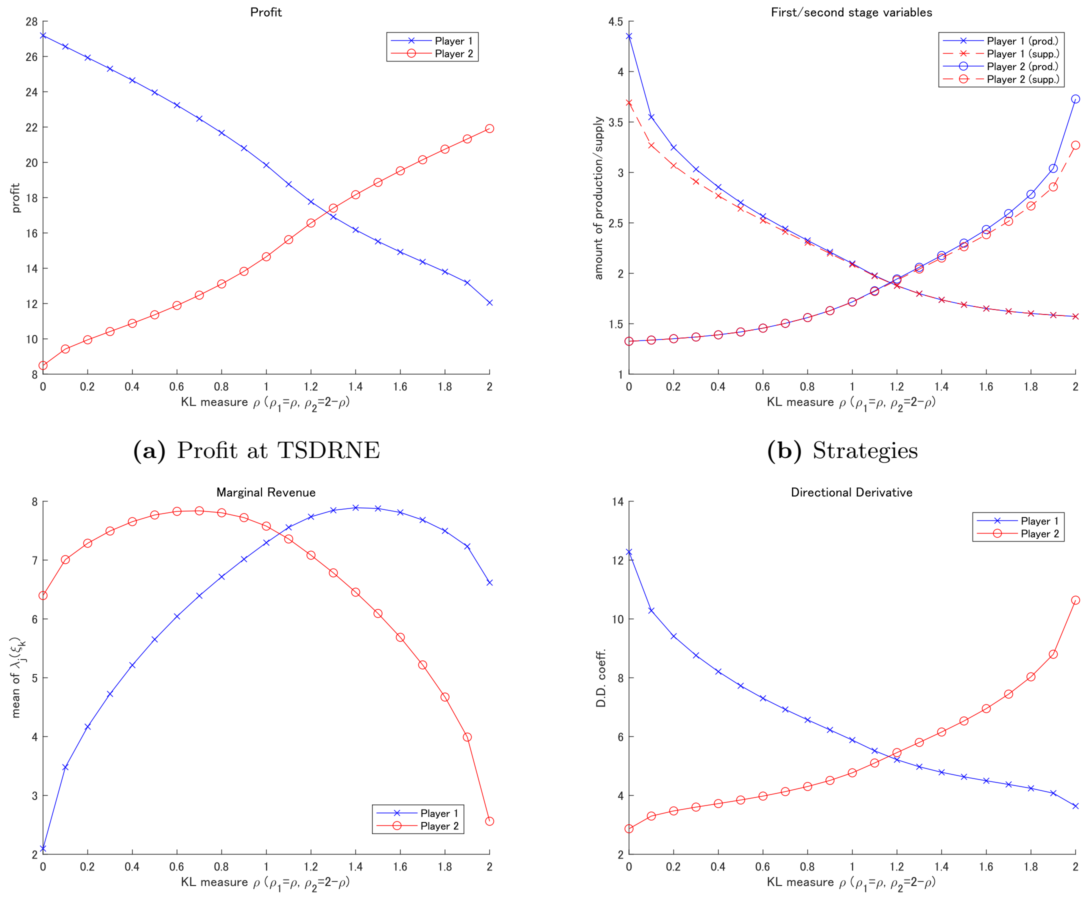
Figure 4.1d represents the curve of the maximum of the absolute directional derivative $|\delta^\top Q_\nu(x^{*,\nu}, \cdot)|$ for the expected value $\mathbb{E}_{P^{*,\nu}}[Q_\nu(x^{*,\nu}, \xi)]$ with respect to the perturbed probability $\delta \in I\!R^K$ subject to $\sum \delta_k = 0$ (because $P^{*,\nu} + \delta$ must be included in Δ under $\sum P^{*,\nu}(\xi_k) = 1$). As the absolute value $|\delta^\top Q_\nu(x^{*,\nu}, \cdot)|$ decreases, it suggests that the solution $x^{*,\nu}$ is robust regarding the perturbation of the probability distribution. Eventually, a small directional derivative indicates that the performance of the out-of-sample validation is less sensitive.

4.6 Concluding remarks

In this chapter we discussed a class of nonlinear two-stage distributionally robust Nash games and demonstrated the existence of TSDRNE points under convexity, compactness, and continuity. We introduced a two-stage distributionally robust variational inequality to construct a solution method for finding an equilibrium point. We also considered a two-stage distributionally robust Cournot–Nash competition as an application and showed the existence of equilibria.

We have only provided the existence of TSDRNE, and some readers may be interested in the uniqueness of TSDRNE. However, establishing the uniqueness of TSDRNE may be more difficult than its existence. A general technique to show the uniqueness of equilibrium in such a noncooperative game is to identify the strict/strong monotonicity of a mapping for reformulated variational inequalities, as we have shown in Theorem 4.3.2. We have tried this approach and investigated such properties of the TSDRVI. However, as Sun et al. [118] have already pointed out the same thing in one-stage DRVI, the class of DRVI is no longer strictly monotone since DRVI intrinsically contains the probability maximization such as (4.12) in Section 4.3, which can also be said to the TSDRVI; please refer to [118, Section 3.3] for more detailed mathematical reasons. For the above reasons, the two-stage distributionally robust Nash equilibrium (TSDRNE) may not be unique in general.

Two challenges still exist: 1) More efficient algorithms should be established to find a solution to TSDRVI, which also guarantees global convergence. 2) The idea should be extended to a case in which the random variable ξ follows a continuous probability distribution, and the convergence of its discrete approximation methods should be analyzed.



(a) Profit at TSDRNE

(b) Strategies

(c) Shadow price $\bar{\lambda}^{*,\nu}$

(d) Directional derivative

Figure 4.1: Results of numerical experiments in the Cournot–Nash competition ($N = 2$)

Acknowledgments

This chapter and Appendix C are based on A. Hori and N. Yamashita, "Two-stage distributionally robust noncooperative games: Existence of Nash equilibrium and its application to Cournot–Nash competition," *Journal of Industrial and Management Optimization*, 2022. DOI: `https://doi.org/10.3934/jimo.2022221`. Reproduced with permission from American Institute of Mathematical Sciences.

Chapter 5

Distributionally robust expected residual minimization for stochastic variational inequality problems

5.1 Introduction

In this chapter, we consider the following stochastic variational inequality (SVI) in the almost sure formulation: Find $x^* \in S$ such that

$$\langle F(x^*, \xi), x - x^* \rangle \geq 0 \quad \forall x \in S, \qquad (5.1)$$
$$\text{for } \xi \in \Xi, \text{ almost surely,}$$

where $F\colon I\!R^n \times \Xi \to I\!R^n$, and $S \subset I\!R^n$ is closed and convex set. Hereafter, we consider the case where the probability distribution of the random vector ξ may be unknown and provided only partial information, and let $\Xi \subset I\!R^m$ be a closed convex set referred to as the support of distributions of ξ. SVI (5.1) is applied in several fields such as economics or engineering to design a market or traffic model, respectively. In particular, when the set S is given as the nonnegative orthant $I\!R^n_+ := \{x \in I\!R^n \mid x \geq 0\}$, SVI (5.1) can be deduced as the stochastic nonlinear complementarity problem (SNCP): Find x^* such that

$$x^* \geq 0, \; F(x^*, \xi) \geq 0, \; \langle F(x^*, \xi), x^* \rangle = 0, \qquad (5.2)$$

and it has also been studied for a long time. If the mapping F is linear, SNCP (5.2) is referred to as the stochastic linear complementarity problem (SLCP).

In general, there may be no solution that satisfies (5.1) or (5.2) for almost every $\xi \in \Xi$; thus, the important goal is to find a reasonable solution that minimizes the violation of (5.1). To obtain such solutions, several models have been considered such as the expected value (EV) formulation [41], expected residual minimization (ERM) model [14, 80], and distributionally robust model [118, 133].

The EV model considers the following deterministic variational inequality:

$$\langle \hat{F}(x^*), x - x^* \rangle \geq 0 \quad \forall x \in S, \qquad (5.3)$$

where $\hat{F}(x) := \mathbb{E}\,[F(x, \xi)]$. Note that an alternative way can also be considered for the expected value of F, such as $F(x, \mathbb{E}\,[\xi])$; however, this is not equivalent to the mapping $\hat{F}$ in general.

On the other hand, the ERM was proposed by Chen and Fukushima [14] for the SLCP. The primary purpose was to reformulate (5.2) as a stochastic optimization problem by using a merit function for the LCP, e.g., the squared Fischer–Burmeister function. They verified that the ERM tends to output more conservative solutions compared with the EV because the ERM is designed to minimize the mean distance to the solution set of VI for each $\xi \in \Xi$, while the EV only considers the mean $\hat{F}$ of the mapping $F(\cdot, \xi)$.

As the natural extension, the ERM for SVI (5.1) can be considered as follows by using a merit function $f(\cdot, \xi)\colon \mathbb{R}^n \to \mathbb{R}_+$ for variational inequalities:

$$\begin{array}{lll} \text{(ERM)} & \min & \mathbb{E}\left[f(x,\xi)\right] \\ & \text{s.t.} & x \in S, \end{array}$$

where the function $f(\cdot, \xi)$ satisfies the following properties for any fixed $\xi \in \Xi$:

(i) $f(x, \xi) \geq 0$ for every $x \in S$;

(ii) $x^* \in S$ is a solution of the VIP if and only if $f(x^*, \xi) = 0$.

To date, several ERM models have been proposed corresponding to each merit function [18, 80, 81].

However, the ERM has two drawbacks. First, its distribution of the random vector ξ is assumed to be known in spite of the fact that it may not be observed in various real situations. Even if one can estimate a distribution from observations, the reliability and robustness of solutions for SVI (5.1) or SNCP (5.2) are not guaranteed unless the estimation is sufficiently close to the true distribution, which is referred to as 'black swans' in risk theory. Second, the ERM requires a numerical integration such as the (quasi-)Monte Carlo method to evaluate the expected residual value. However, numerical integration is computationally expensive in general; it is advisable to avoid such a sample-based approach.

To tackle these issues, Zhu et al. [133] proposed the following conservative approximation model for SNCP (5.2):

$$\begin{array}{lll} \min\limits_{x \in \mathbb{R}^n} & \sup\limits_{P \in \mathscr{P}} & \left\{\mathbb{E}_P\left[\Psi(x,\xi)\right] \mid P(\{F(x,\xi) \geq 0\} \cap \Xi) \geq 1-\varepsilon\right\} \\ \text{s.t.} & x \geq 0, & \end{array} \tag{5.4}$$

where $0 < \varepsilon < 1$ is a tolerance parameter, and $\Psi\colon \mathbb{R}^n \times \Xi \to \mathbb{R}$ is a complementarity measure, e.g., $\Psi(x, \xi) = \|x \circ F(x, \xi)\|_2^2$, where $\circ$ denotes the Hadamard product defined by $x \circ y = (x_1y_1, x_2y_2, \ldots, x_ny_n)$ for the vectors $x \in \mathbb{R}^n$ and $y \in \mathbb{R}^n$. Here, $\mathbb{E}_P[\cdot]$ is the expected value with respect to a distribution function $P(\cdot) \in \mathscr{P}$, where $\mathscr{P}$ is an uncertainty set of the distribution functions supported over Ξ called an ambiguity set. They considered $\mathscr{P}$ as the following moment ambiguity set:

$$\mathscr{P} = \left\{P \in \mathscr{M}_\Xi \;\middle|\; \mathbb{E}_P[\xi] = \mu_0, \mathbb{E}_P[\xi\xi^\top] = \Sigma_0 + \mu_0\mu_0^\top\right\}, \tag{5.5}$$

where $\mathscr{M}_\Xi$ denotes a set of all probability measures supported over Ξ, and μ_0 and Σ_0 respectively denote the (estimated) mean and variance of ξ from observation. Then they reformulated (5.4) into a nonlinear semidefinite programming problem (NSDP). In the definition of (5.5), however, it is implicitly assumed that an observer knows the exact mean μ_0 and variance Σ_0. In the absence of this assumption, the model may not perform properly because observation errors are not considered. In terms of the distributionally robust optimization (DRO), it is often considered that μ_0 and Σ_0 cannot be estimated exactly, e.g., the lack of sample data, which motivates us to adopt a more general moment ambiguity set.

In this study, we propose a distributionally robust model of SVI (5.1) under uncertainty of distribution, where the ambiguity set is based on Delage and Ye [23] (eq. (5.10) in Assumption 5.2.1). Note that our methodology differs from an analysis of the (qualitative or quantitative) statistical robustness [40, 60, 67] of a solution obtained from a sample average approximation approach, whose data may contain noise; this is one of the key concepts to study a stochastic model under the uncertainty distribution. This thesis rather focuses on distributional robustness by constructing the ambiguity set with the data. We propose the following distributionally robust ERM (DRERM) model:

$$\text{(DRERM)}\quad \begin{array}{ll} \min & \sup\limits_{P\in\mathscr{P}} \mathbb{E}_P\left[f(x,\xi)\right] \\ \text{s.t.} & x\in S. \end{array}$$

This model can be regarded as an extension of the ERM and utilizes some remarkable aspects as stated below: We illustrate a reformulation of (DRERM) into an NSDP under certain suitable assumptions. Consequently, it is not required to compute numerical integrals to evaluate the expected value of the stochastic gap functions.

In this thesis, we mainly focus on the following regularized gap function [35] as a merit function f in (DRERM):

$$f(x,\xi) = f_\alpha(x,\xi) := \max_{y\in S}\left\{\langle F(x,\xi), x-y\rangle - \frac{1}{2\alpha}\|y-x\|^2\right\}. \tag{5.6}$$

where $\alpha > 0$ is a regularization parameter. When $S = \mathbb{R}^n$, the regularized gap function is reduced to $(\alpha/2)\|F(x,\xi)\|^2$. Therefore, the ERM with f_α is regarded as an extension of the least square problem, and hence it is popular [1, 15, 80, 81]. Moreover, as we will see in Section 5.2.2, (DRERM) with f_α can be reformulated into a convex NSDP for certain SVIs. Note that the NSDP approximation proposed in [133] is not convex in general.

The remainder of this chapter is organized as follows. In Section 5.2, we propose an NSDP model that conservatively approximates (DRERM). In addition, we show the convexity of the NSDP under certain assumptions. In Section 5.3, we conduct two types of numerical experiments to illustrate the behavior of our reformulation model. In Section 5.4, we conclude this study.

5.2 Reformulation and convexity of DRERM

First, we introduce several approaches to solving (DRERM). Second, we reformulate (DRERM) into a deterministic NSDP to find its solution efficiently. Finally, we provide a sufficient condition for the convexity of the NSDP when the mapping F is affine with respect to x.

A general technique for solving (DRERM), regardless of the definition of $\mathscr{P}$, is to reformulate it into the following semi-infinite programming and apply the cutting-surface method [83]:

$$\begin{array}{ll} \min\limits_{x,\theta} & \theta \\ \text{s.t.} & \theta \ge \mathbb{E}_P\left[f_\alpha(x,\xi)\right] \quad \forall P\in\mathscr{P}, \\ & x\in S. \end{array} \tag{5.7}$$

Moreover, when Ξ is a finite sample space, i.e., $\Xi := \{\xi^1, \xi^2, \ldots, \xi^L\}$, problem (5.7) is consequently reduced to the following robust optimization problem because $\mathscr{P}$ can be regarded as a subset of $\mathbb{R}^L$:

$$\begin{aligned} \min_{x,\theta} \quad & \theta \\ \text{s.t.} \quad & \theta \geq \frac{1}{L}\sum_{k=1}^{L} f_\alpha(x,\xi^k)P_k \quad \forall P \in \mathscr{P} \subset \{P \in \mathbb{R}^L_+ \mid \textstyle\sum_{k=1}^L P_k = 1\}, \\ & x \in S. \end{aligned} \tag{5.8}$$

Thus, nonlinear robust optimization frameworks can be directly applied to (5.8). For more details, see [7, 8, 10].

Another strategy to solve (DRERM) is to consider the duality of the inner supremum part:

$$\sup_{P\in\mathscr{P}} \mathbb{E}_P\left[f_\alpha(x,\xi)\right], \tag{5.9}$$

and solve the dual problem. We adopt this approach and demonstrate that (DRERM) can be reformulated as a deterministic NSDP under certain assumptions. For more detailed techniques to deal with general DRO, see [97].

In the remainder of this study, we assume that ξ is a continuous random variable, and the ambiguity set $\mathscr{P}$ is assumed to be given as the following moment set [23], which has been widely applied in the existing literature on DRO.

Assumption 5.2.1 (Delage and Ye [23])**.** *The ambiguity set $\mathscr{P}$ is given by*

$$\mathscr{P} := \left\{ P \in \mathscr{M}_\Xi \;\middle|\; \begin{array}{l} \left(\mathbb{E}_P[\xi] - \mu_0\right)^\top \Sigma_0^{-1} \left(\mathbb{E}_P[\xi] - \mu_0\right) \leq \gamma_1 \\ \mathbb{E}_P\left[(\xi-\mu_0)(\xi-\mu_0)^\top\right] \preceq \gamma_2\Sigma_0 \end{array} \right\} \tag{5.10}$$

where $\gamma_1 \geq 0$, $\gamma_2 \geq 1$, $\mu_0 \in \Xi$, and $\Sigma_0 \in \mathbb{S}^m_{++}$.

The first condition of (5.10), i.e., $\left(\mathbb{E}_P[\xi] - \mu_0\right)^\top \Sigma_0^{-1} \left(\mathbb{E}_P[\xi] - \mu_0\right) \leq \gamma_1$, represents the uncertainty of the true mean $\mathbb{E}_P[\xi]$ given by an ellipsoid centered on the estimated mean μ_0. In addition, if $\gamma_1 = 0$, then $\mathbb{E}_P[\xi] = \mu_0$. The second condition $\mathbb{E}_P[(\xi-\mu_0)(\xi-\mu_0)^\top] \preceq \gamma_2\Sigma_0$ refers to the uncertainty of the true variance-covariance $\mathbb{E}_P[(\xi-\mu_0)(\xi-\mu_0)^\top]$. The parameters γ_1 and γ_2 determine the strength of the confidence of estimations μ_0 and Σ_0, respectively; hence, they are referred to as confidence parameters. A method for determining suitable γ_1 and γ_2 from observed samples is introduced in Section 3.4 in [23].

Remark 5.2.1. *When $\gamma_1 = 0$, $\gamma_2 = 1$, and the equality holds in the variance-covariance condition in* (5.10)*, the set $\mathscr{P}$ is reduced to* (5.5) *considered by Zhu et al* [133].

Under Assumption 5.2.1, we obtain the following property.

Theorem 5.2.2. *Suppose that Assumption* 5.2.1 *holds. Then* (DRERM) *is equivalently reformulated as the following semi-infinite programming with second-order cone constraints:*

$$\begin{aligned} \text{(SIP)} \quad \min_{(x,y_0,y,Y,z_0)\in\mathcal{V}} \quad & y_0 + z_0 + \mu_0^\top y + \langle \gamma_2\Sigma_0 + \mu_0\mu_0^\top, Y\rangle \\ \text{s.t.} \quad & z_0 \geq \sqrt{\gamma_1}\left\|\Sigma_0^{1/2}(y + 2Y\mu_0)\right\|, \\ & \xi^\top Y\xi + \xi^\top y + y_0 \geq f_\alpha(x,\xi) \quad \forall \xi \in \Xi, \\ & x \in S,\ Y \in \mathbb{S}^m_+, \end{aligned}$$

where $\mathcal{V} := \mathbb{R}^n \times \mathbb{R} \times \mathbb{R}^m \times \mathbb{S}^m \times \mathbb{R}$.

Proof. From Assumption 5.2.1 and Lemma 1 of [23], for any fixed x, the optimal value of (5.9), which is denoted by $\Psi(x;\gamma_1,\gamma_2)$, is equal to that of the following dual problem of (5.9):

$$\begin{array}{ll} \min\limits_{y_0,y,Y,z_0} & y_0+z_0+\mu_0^\top y+\left\langle \gamma_2\Sigma_0+\mu_0\mu_0^\top, Y\right\rangle \\ \text{s.t.} & z_0 \geq \sqrt{\gamma_1}\left\|\Sigma_0^{1/2}(y+2Y\mu_0)\right\|, \\ & \xi^\top Y\xi+\xi^\top y+y_0 \geq f_\alpha(x,\xi) \quad \forall\xi\in\Xi, \\ & Y\in\mathbb{S}_+^m. \end{array} \tag{5.11}$$

Thus, we obtain the equivalent reformulation of (DRERM) by considering $\min\{\Psi(x;\gamma_1,\gamma_2) \mid x\in S\}$. Since optimal values of (SIP) and (DRERM) are equal, the assertion is shown. □

5.2.1 Reformulation of SIP into NSDP

The goal of this section is to prove that the semi-infinite constraint

$$\xi^\top Y\xi+\xi^\top y+y_0 \geq f_\alpha(x,\xi) \quad \forall\xi\in\Xi, \tag{5.12}$$

can be reformulated as a semidefinite constraint by using the duality for the inner maximization of (5.6).

In the remainder of this chapter, we assume that the closed convex set S is given as a polyhedron:

$$S := \{x\in I\!R^n \mid Ax=b, x\geq 0\},$$

where $A\in I\!R^{l\times n}$ and $b\in I\!R^l$.

First, we provide an equivalent form of (5.12) by using the strong duality of the maximization problem in (5.6).

Lemma 5.2.3. *The point (x,y_0,y,Y) satisfies (5.12) if and only if there exists $(\lambda,\mu)\in I\!R^l\times I\!R_+^n$ such that*

$$\xi^\top Y\xi+\xi^\top y+y_0 \geq \omega_\alpha(x,\lambda,\mu;\xi) \quad \forall\xi\in\Xi. \tag{5.13}$$

Here,

$$\omega_\alpha(x,\lambda,\mu;\xi) := \frac{\alpha}{2}\|F(x,\xi)+A^\top\lambda-\mu\|^2+\langle b-Ax,\lambda\rangle+\langle\mu,x\rangle. \tag{5.14}$$

Proof. First, we prove (5.13) implies (5.12). We have the following minimization problem by considering the duality of the maximization problem included in (5.6).

$$\begin{array}{ll} \min\limits_{(\lambda,\mu)\in I\!R^l\times I\!R^n} & \omega_\alpha(x,\lambda,\mu;\xi) \\ \text{s.t.} & \mu\in I\!R_+^n \end{array} \tag{5.15}$$

From the weak duality, we have $\omega_\alpha(x,\lambda,\mu;\xi)\geq f_\alpha(x,\xi)$ for each $(x,\xi)\in S\times\Xi$. Thus, if there exists $(\lambda,\mu)\in I\!R^l\times I\!R_+^n$ such that (x,λ,μ,y_0,y,Y) satisfies (5.13), then the point (x,y_0,y,Y) satisfies (5.12).

Next, we prove the converse, i.e., (5.12) implies (5.13). The inner maximization in the function f_α is a convex optimization problem whose optimal value is finite for any $x\in S$. Moreover, owing to the strong duality, there exists $(\lambda,\mu)\in I\!R^l\times I\!R_+^n$ such that $f_\alpha(x,\xi)=\omega_\alpha(x,\lambda,\mu;\xi)$ for each $(x,\xi)\in S\times\Xi$. Therefore, if (x,y_0,y,Y) satisfies the condition (5.12), there exists $(\lambda,\mu)\in I\!R^l\times I\!R_+^n$ such that (5.13) holds. □

Now, we make assumptions on the mapping F and the support Ξ in SVI (5.1). Similar assumptions on F and Ξ have already been considered by Zhu et al. [133] for SNCP (5.2)[1]. For certain examples that satisfy the following assumptions on SVI (5.1), see [1].

Assumption 5.2.4.

(i) *The i-th element of the mapping F is affine with respect to ξ:*

$$F_i(x,\xi) := \left(c^i(x)\right)^\top \xi + c_0^i(x), \ \ i = 1,2,\dots,n.$$

(ii) *The support Ξ is given as*

$$\Xi := \{\xi \in I\!R^m \mid g_i(\xi) \le 0, i = 1,2,\dots,p\}. \tag{5.16}$$

Here, $g_i \colon I\!R^m \to I\!R$ is defined by

$$g_i(\xi) := \xi^\top \tilde{A}_i \xi + 2\tilde{b}_i^\top \xi + \tilde{c}_i, \quad i = 1,2,\dots,p, \tag{5.17}$$

where $\tilde{A}_i \in \mathbb{S}^m$, $\tilde{b}_i \in I\!R^m$, and $\tilde{c}_i \in I\!R$.

As preliminaries, let us introduce the S-procedure and its special case.

Lemma 5.2.5 (S-procedure Derinkuyu and Pınar [25])**.** *Let Ξ be given as* (5.16) *and*

$$g_0(\xi) := \xi^\top \bar{A}_0 \xi + 2\xi^\top \bar{b}_0 + \bar{c}_0, \tag{5.18}$$

where $\bar{A}_0 \in \mathbb{S}^m$, $\bar{b}_0 \in I\!R^m$, and $\bar{c}_0 \in I\!R$. Assume that there exists $s \in I\!R^p_+$ such that

$$g_0(\xi) + \sum_{i=1}^p s_i g_i(\xi) \ge 0 \quad \forall \xi \in I\!R^m. \tag{5.19}$$

Then, $g_0(\xi) \ge 0$ for all $\xi \in \Xi$.

The following lemma indicates that the converse also holds when $p = 1$ in Lemma 5.2.5.

Lemma 5.2.6 (Pólik and Terlaky [95])**.** *Suppose that Ξ is given by* (5.16) *with $p = 1$ and let $g_0(\xi)$ be defined as* (5.18)*. Assume that there exists $\hat{\xi}_0$ such that $g_1(\hat{\xi}_0) < 0$. Then, the statements* (i) *and* (ii) *are equivalent:*

(i) *For all $\xi \in I\!R^m$, $g_1(\xi) \le 0$ implies $g_0(\xi) \ge 0$;*

(ii) *there exists some nonnegative number $s \ge 0$ such that*

$$g_0(\xi) + s g_1(\xi) \ge 0 \quad \forall \xi \in I\!R^m.$$

We further introduce an equivalence between nonnegative quadratic functions on $I\!R^m$ and semidefiniteness.

[1] Only when the complementarity measure is evaluated by $\|x \circ F(x,\xi)\|_\infty$, the mapping F of (5.4) is allowed up to second-order with respect to ξ.

Lemma 5.2.7 (Proposition 2 in Sturm and Zhang [116])**.** *Let $\tilde{A} \in \mathbb{S}^m$, $\tilde{b} \in I\!R^m$, and $\tilde{c} \in I\!R$ be given. Then, the following two conditions* (i) *and* (ii) *are equivalent:*

$$\text{(i)} \quad [1,\ \xi^\top] \begin{bmatrix} \tilde{c} & \tilde{b}^\top \\ \tilde{b} & \tilde{A} \end{bmatrix} \begin{bmatrix} 1 \\ \xi \end{bmatrix} \geq 0 \quad \forall \xi \in I\!R^m;$$

$$\text{(ii)} \quad \begin{bmatrix} \tilde{c} & \tilde{b}^\top \\ \tilde{b} & \tilde{A} \end{bmatrix} \succeq O.$$

Zhu et al. [133] proposed a certain NSDP that conservatively approximates DRO (5.4), where the conservative approximation denotes that the optimal value of the NSDP is not less than that of DRO (5.4). In this chapter, we also provide the following conservative approximation of (DRERM) based on their technique.

$$\begin{array}{lll} \text{(NSDP)} & \displaystyle\min_{(w,z_0,s)\in\mathcal{W}\times I\!R\times I\!R^p} & z_0 + y_0 + \mu_0^\top y + \left\langle \gamma_2\Sigma_0 + \mu_0\mu_0^\top, Y \right\rangle \\ & \text{s.t.} & z_0 \geq \sqrt{\gamma_1} \left\| \Sigma_0^{1/2}(y + 2Y\mu_0) \right\|, \\ & & \mathcal{D}_\alpha(w) + \displaystyle\sum_{i=1}^p s_i \tilde{\mathcal{A}}_i \succeq O, \\ & & x \in S,\ \mu \in I\!R^n_+,\ Y \in \mathbb{S}^m_+,\ s \in I\!R^p_+, \end{array}$$

where $w := (x, \lambda, \mu, y_0, y, Y) \in \mathcal{W} := I\!R^n \times I\!R^l \times I\!R^n \times I\!R \times I\!R^m \times \mathbb{S}^m$, and $\mathcal{D}_\alpha \colon \mathcal{W} \to \mathbb{S}^{m+1}$ is a symmetric-matrix-valued function defined as follows:

$$\mathcal{D}_\alpha(w) := \begin{bmatrix} y_0 & 1/2y^\top \\ 1/2y & Y \end{bmatrix} - \left\{ G(x,\lambda,\mu) + \frac{\alpha}{2} \sum_{i=1}^n H^i(x,\lambda,\mu) \right\}, \tag{5.20}$$

where

$$G(x,\lambda,\mu) := \begin{bmatrix} \langle b - Ax, \lambda \rangle + \langle \mu, x \rangle & 0^\top \\ 0 & O_{m\times m} \end{bmatrix},$$

$$H^i(x,\lambda,\mu) := \begin{bmatrix} p_0^i(x,\lambda,\mu)^2 & p_0^i(x,\lambda,\mu)c^i(x)^\top \\ p_0^i(x,\lambda,\mu)c^i(x) & c^i(x)c^i(x)^\top \end{bmatrix},\ i = 1,2,\ldots,n,$$

$$p_0^i(x,\lambda,\mu) := c_0^i(x) + \sum_{j=1}^l a^{ji}\lambda_j - \mu_i,\ i = 1,2,\ldots,n,$$

and $\tilde{\mathcal{A}}_i$ is defined as

$$\tilde{\mathcal{A}}_i := \begin{bmatrix} \tilde{c}_i & \tilde{b}_i^\top \\ \tilde{b}_i & \tilde{A}_i \end{bmatrix} \quad i = 1,2,\ldots,p.$$

Next, we provide several definitions and lemmas to prove that (NSDP) gives a conservative approximation of (DRERM). Now, we define

$$\tilde{A}_0 := Y - \frac{\alpha}{2}\sum_{i=1}^n c^i(x)c^i(x)^\top,\ \tilde{b}_0 := \frac{y}{2} - \frac{\alpha}{2}\sum_{i=1}^n p_0^i(x,\lambda,\mu)c^i(x),$$

$$\tilde{c}_0 := y_0 - \langle b - Ax, \lambda \rangle - \langle \mu, x \rangle - \frac{\alpha}{2}\sum_{i=1}^n p_0^i(x,\lambda,\mu)^2,$$

and

$$h(\xi) := \xi^\top \tilde{A}_0 \xi + 2\xi^\top \tilde{b}_0 + \tilde{c}_0. \tag{5.21}$$

Under Assumption 5.2.4–(i), (5.14) is written as

$$\omega_\alpha(x,\lambda,\mu;\xi) = \begin{bmatrix}1, & \xi^\top\end{bmatrix} \left\{ G(x,\lambda,\mu) + \frac{\alpha}{2}\sum_{i=1}^n H^i(x,\lambda,\mu) \right\} \begin{bmatrix} 1 \\ \xi \end{bmatrix}. \tag{5.22}$$

Through the straightforward calculation, we obtain the following equalities:

$$\begin{bmatrix}1, & \xi^\top\end{bmatrix} \mathcal{D}_\alpha(w) \begin{bmatrix} 1 \\ \xi \end{bmatrix} = \xi^\top Y\xi + \xi^\top y + y_0 - \omega_\alpha(x,\lambda,\mu;\xi) = h(\xi). \tag{5.23}$$

Lemma 5.2.8. *The nonlinear semidefinite constraint included in* (NSDP)*, i.e.,*

$$\mathcal{D}_\alpha(w) + \sum_{i=1}^p s_i\tilde{\mathcal{A}}_i \succeq O, \tag{5.24}$$

is equivalent to

$$h(\xi) + \sum_{i=1}^p s_i g_i(\xi) \geq 0 \quad \forall \xi \in I\!R^m. \tag{5.25}$$

Proof. By Lemma 5.2.7, (5.24) is equivalent to

$$\begin{bmatrix}1, & \xi^\top\end{bmatrix} \left(\mathcal{D}_\alpha(w) + \sum_{i=1}^p s_i\tilde{\mathcal{A}}_i \right) \begin{bmatrix} 1 \\ \xi \end{bmatrix} \geq 0 \quad \forall \xi \in I\!R^m. \tag{5.26}$$

Since $[1,\ \xi^\top]\mathcal{D}_\alpha(w)\begin{bmatrix} 1 \\ \xi \end{bmatrix} = h(\xi)$ from (5.23) and $[1,\ \xi^\top]\tilde{\mathcal{A}}_i\begin{bmatrix} 1 \\ \xi \end{bmatrix} = g_i(\xi)$, (5.26) can be equivalently represented as (5.25). □

The next lemma provides a sufficient condition for semi-infinite constraint (5.12).

Lemma 5.2.9. *Suppose that Assumption* 5.2.4 *holds. Whenever $p \geq 1$, if there exists $(w,s) \in \mathcal{W} \times I\!R^p$ such that $\mu \in I\!R^n_+$, $s \in I\!R^p_+$, and* (5.24)*, i.e.,*

$$\mathcal{D}_\alpha(w) + \sum_{i=1}^p s_i\tilde{\mathcal{A}}_i \succeq O,$$

then the subvector (x, y_0, y, Y) satisfies the semi-infinite constraint (5.12)*, i.e.,*

$$\xi^\top Y\xi + \xi^\top y + y_0 \geq f_\alpha(x,\xi) \quad \forall \xi \in \Xi.$$

Furthermore, when $p = 1$ and the assumption of Lemma 5.2.6 *holds, the converse is also true, i.e., if (x, y_0, y, Y) satisfies* (5.12)*, then there exists $(\lambda,\mu,s) \in I\!R^l \times I\!R^n_+ \times I\!R^p_+$ such that* (5.24) *satisfies.*

Proof. First, we prove the general case where $p \geq 1$. Assume that there exist $w \in \mathcal{W}$ and $s \in I\!R^p_+$ such that semidefinite constraint (5.24) holds. Then, by Lemma 5.2.8, we have (5.25), i.e.,

$$h(\xi) + \sum_{i=1}^p s_i g_i(\xi) \geq 0 \quad \forall \xi \in I\!R^m.$$

By regarding $h(\xi)$ as $g_0(\xi)$ in Lemma 5.2.5, (5.25) implies $h(\xi) \geq 0$ for all $\xi \in \Xi$, and it then follows from (5.23) that (5.13) holds, i.e.,

$$\xi^\top Y \xi + \xi^\top y + y_0 \geq \omega_\alpha(x, \lambda, \mu; \xi) \quad \forall \xi \in \Xi.$$

Finally, Lemma 5.2.3 states that w satisfies (5.13) if and only if its subvector (x, y_0, y, Y) satisfies semi-infinite constraint (5.12). The first part of the proof is completed.

Next, we prove that the converse when $p = 1$ and the assumption of Lemma 5.2.6 holds. Suppose that (x, y_0, y, Y) satisfies (5.12). By Lemma 5.2.3, (5.12) holds if and only if there exists $(\lambda, \mu) \in I\!R^l \times I\!R^n_+$ such that (5.13) holds. Note that under Assumption 5.2.4–(i), the function $\omega_\alpha(x, \lambda, \mu; \xi)$ is given as (5.22). Then, $\xi^\top Y \xi + \xi^\top y + y_0 - \omega_\alpha(x, \lambda, \mu; \xi) \geq 0$ and (5.23) yield $h(\xi) \geq 0$. Note that $h(\xi) \geq 0$ for all $\xi \in \Xi$ if and only if for all $\xi \in I\!R^m$, $g_1(\xi) \leq 0$ implies $h(\xi) \geq 0$. It then follows from Assumption 5.2.4–(ii) with $p = 1$ and Lemma 5.2.6 that there exists $s \in I\!R_+$ such that $h(\xi) + s g_1(\xi) \geq 0$ for all $\xi \in I\!R^m$. By Lemma 5.2.8, this condition is equivalent to semidefinite constraint (5.24) in (NSDP). Thus, we have proved the converse. □

The following result shows the feasibility between the constraints of (SIP) and (NSDP).

Proposition 5.2.10. *Suppose that Assumption* 5.2.4 *holds. Whenever* $p \geq 1$*, if* $(w, z_0, s) \in \mathcal{W} \times I\!R \times I\!R^p$ *is feasible to* (NSDP)*, then its subvector* $(x, y_0, y, Y, z_0) \in \mathcal{V}$ *is also feasible to* (SIP)*. Moreover, when* $p = 1$ *and the assumption of Lemma* 5.2.6 *holds, if* $(x, y_0, y, Y, z_0) \in \mathcal{V}$ *is feasible to* (SIP)*, then there exists* $(\lambda, \mu, s) \in I\!R^l \times I\!R^n_+ \times I\!R^p_+$ *such that* $(w, z_0, s) \in \mathcal{W} \times I\!R \times I\!R_+$ *is also feasible to* (NSDP)*.*

Proof. Note that all constraints in (NSDP) except (5.24) coincide with those in (SIP) excluding semi-infinite constraint (5.12). This statement and Lemma 5.2.9 ensure that if $p \geq 1$, and (w, s, z_0) is the feasible solution of (NSDP), then its subvector (x, y_0, y, Y, z_0) is the feasible solution to (SIP). Thus, we showed the general case where $p \geq 1$.

Suppose that $p = 1$, and $(x, y_0, y, Y, z_0) \in \mathcal{V}$ is a feasible solution to (SIP). As mentioned above, (w, z_0, s) satisfies the constraints of (NSDP) except (5.24). Moreover, Lemma 5.2.9 guarantees that there exists $(\lambda, \mu, s) \in I\!R^l \times I\!R^n_+ \times I\!R^p_+$ such that (w, z_0, s) satisfies constraint (5.24). We have completed the proof. □

By using the above lemmas, we show one of the main results.

Theorem 5.2.11. *Suppose that Assumptions* 5.2.4 *holds. Then,* (SIP) *can be conservatively approximated as* (NSDP)*.*

Proof. Suppose that $(w, z_0, s) \in \mathcal{W} \times I\!R \times I\!R^p_+$ is a feasible point of (NSDP). It then follows from Proposition 5.2.10 that the subvector (x, y_0, y, Y, z_0) satisfies the constraints of (SIP). From the above facts, the optimal value of (NSDP) can never be less than that of (SIP). Therefore, (NSDP) is a conservative approximation of (SIP). The proof is completed. □

Here, we provide some examples of Ξ that can be expressed as the intersection of nonnegative quadratic functions.

Example 5.2.2 (Box set)**.** *Consider* Ξ *given by the following box set:*

$$\Xi := \left\{ \xi \in I\!R^m \mid \xi_i^l \leq \xi_i \leq \xi_i^u, \; i = 1, 2, \ldots, m \right\}.$$

By using a quadratic function, $\xi_i^l \le \xi \le \xi_i^u$ can be rewritten as follows:

$$g_i(\xi) = \xi_i(\xi_i^u + \xi_i^l) - \xi_i^u\xi_i^l - \xi_i^2 = \begin{bmatrix} 1, & \xi^\top \end{bmatrix} T_i \begin{bmatrix} 1 \\ \xi \end{bmatrix} \ge 0,$$

where

$$T_i := \begin{bmatrix} -\xi_i^u\xi_i^l & -\frac{1}{2}\xi_i(\xi_i^u + \xi_i^l)(e^i)^\top \\ -\frac{1}{2}\xi_i(\xi_i^u + \xi_i^l)e^i & -\tilde{I}_i \end{bmatrix}.$$

Here, $e^i \in I\!R^m$ is the i-th column vector of the identity matrix, and $\tilde{I}_i \in I\!R^{m\times m}$ is a matrix whose elements are all zero except the (i,i) entry which is 1.

This example corresponds to the case where $\tilde{A}_i = \tilde{I}_i, \tilde{b}_i = \frac{1}{2}\xi_i(\xi_i^u + \xi_i^l)e^i$, and $\tilde{c}_i = \xi_i^u\xi_i^l$ in (NSDP). □

Example 5.2.3 (Ellipsoids)**.** *Consider Ξ given by the following ellipsoids:*

$$\Xi := \left\{\xi \in I\!R^m \mid (\xi - \hat{\xi}^i)^\top P_i^{-1}(\xi - \hat{\xi}^i) \le 1, i = 1, 2, \ldots, p\right\}, \tag{5.27}$$

where the vector $\hat{\xi}^i \in I\!R^m$ is the center of the i-th ellipsoid, and the matrix P_i is supposed to be positive definite. This example corresponds to the case where $\tilde{A}_i = P^{-1}, \tilde{b}_i = -P_i^{-1}\hat{\xi}^i$, and $\tilde{c}_i = (\hat{\xi}^i)^\top P_i^{-1}\hat{\xi}^i - 1$ in (NSDP). □

Next, we illustrate the special case of Theorem 5.2.11, which ensures that a solution of (NSDP) solves (DRERM).

Corollary 5.2.12. *Suppose that $p = 1$ in* (5.27)*, and that the assumption of Lemma 5.2.6 holds. Then, if $(w, z_0, s) \in \mathcal{W} \times I\!R \times I\!R$ is a global optimum of* (NSDP)*, then (x, y_0, y, Y, z_0) and x are also global optima to* (SIP) *and* (DRERM)*, respectively. In addition, the optimal value of* (NSDP) *is equal to those of* (SIP) *and* (DRERM)*.*

Proof. Let (w, z_0, s) be a global optimum to (NSDP). Assume that its subvector (x, y_0, y, Y, z_0) is not a global optimum of (SIP). Note that (x, y_0, y, Y, z_0) is the feasible solution to (SIP) from Proposition 5.2.10. By the assumption, there exists a feasible solution (x', y_0', y', Y', z_0') in (SIP) such that

$$z_0' + y_0' + \mu_0^\top y' + \left\langle \gamma_2\Sigma_0 + \mu_0\mu_0^\top, Y' \right\rangle < z_0 + y_0 + \mu_0^\top y + \left\langle \gamma_2\Sigma_0 + \mu_0\mu_0^\top, Y \right\rangle. \tag{5.28}$$

Proposition 5.2.10 guarantees that if the solution (x', y_0', y', Y', z_0') is the feasible point to (SIP), then there exists $(\lambda', \mu', s') \in I\!R^l \times I\!R_+^n \times I\!R_+$ such that $(w', z_0', s') \in \mathcal{W} \times I\!R \times I\!R_+$ is the feasible solution to (NSDP), where $w' := (x', \lambda', \mu', y_0', y', Y') \in \mathcal{W}$. Because the objective functions of (SIP) and (NSDP) coincide, the solution (w', z_0', s') of (NSDP) also satisfies the inequality (5.28). Hence, it contradicts that (w, z_0, s) is a global optimum to (NSDP). We have that (x, y_0, y, Y, z_0), which is the subvector of the global optimum (w, z_0, s) of (NSDP), is also the global optimum in (SIP), and optimal values are equal because their objective functions coincide. Moreover, since (DRERM) is equivalent to (SIP) from Theorem 5.2.2, x is also a global optimum to (DRERM), and their optimal values are equal. □

In addition, when $\Xi = I\!R^m$, then we can show an equivalence between (SIP) (or (DRERM)) and the following NSDP:

$$
\begin{array}{lll}
(\mathrm{NSDP}') & \min\limits_{(w,z_0)\in\mathcal{W}\times I\!R} & z_0 + y_0 + \mu_0^\top y + \left\langle \gamma_2\Sigma_0 + \mu_0\mu_0^\top, Y\right\rangle \\
& \text{s.t.} & z_0 \geq \sqrt{\gamma_1}\left\|\Sigma_0^{1/2}(y + 2Y\mu_0)\right\|, \\
& & \mathcal{D}_\alpha(w) \succeq O, \\
& & x \in S,\ \mu \in I\!R^n_+.
\end{array}
$$

To show this property, we prepare a lemma below.

Lemma 5.2.13. *Let (x, y_0, y, Y) be given. Then, the following two statements are equivalent:*

(i) *There exists $(\lambda, \mu) \in I\!R^l \times I\!R^n_+$ such that $\mathcal{D}_\alpha(w) \succeq O$;*

(ii) *$Y \in \mathbb{S}^m_+$ and*

$$\xi^\top Y\xi + \xi^\top y + y_0 \geq f_\alpha(x, \xi) \quad \forall \xi \in \Xi = I\!R^m. \tag{5.29}$$

Proof. First, we show that (i) implies (ii). By Lemma 5.2.7 and the first equality of (5.23), $\mathcal{D}_\alpha(w) \succeq O$ if and only if

$$\xi^\top Y\xi + \xi^\top y + y_0 \geq \omega_\alpha(x, \lambda, \mu; \xi) \quad \forall \xi \in I\!R^m. \tag{5.30}$$

As we mentioned in Lemma 5.2.3, $\omega_\alpha(x, \lambda, \mu; \xi)$ is the dual function of the maximization problem in f_α. This implies that for any $x \in S$ and $\xi \in I\!R^m$, $\omega_\alpha(x, \lambda, \mu; \xi) \geq f_\alpha(x, \xi) \geq 0$. Then, (5.30) implies $\xi^\top Y\xi + \xi^\top y + y_0 \geq 0$, and by Lemma 5.2.7, we have

$$\begin{bmatrix} y_0 & 1/2y^\top \\ 1/2y & Y \end{bmatrix} \succeq O.$$

By the Schur complement, this ensures the positive semidefiniteness of Y. Furthermore, (5.30) implies (5.29) by Lemma 5.2.3. We have proved the former part of the proof.

Next, we prove that (ii) implies (i). Suppose that $Y \in \mathbb{S}^m_+$ and (5.29) holds. Then, by Lemma 5.2.3, there exists $(\lambda, \mu) \in I\!R^l \times I\!R^n_+$ such that (5.30) holds, and it is immediately observed that $\mathcal{D}_\alpha(w) \succeq O$. Hence, the proof is completed. □

We obtain the relation regarding the feasibility between (SIP) and (NSDP′) by using Lemma 5.2.13.

Proposition 5.2.14. *The point $(x, y_0, y, Y, z_0) \in \mathcal{V}$ is a feasible solution to* (SIP) *if and only if there exists $(\lambda, \mu) \in I\!R^l \times I\!R^n_+$ such that $(w, z_0) \in \mathcal{W} \times I\!R$ is also a feasible solution to* (NSDP′).

Proof. Similar to the proof of Proposition 5.2.10, all the constraints in (NSDP′) except the semidefinite constraint $\mathcal{D}_\alpha(w) \succeq O$ coincide with those in (SIP) excluding semi-infinite constraint (5.12). This statement and Lemma 5.2.13 ensure that for given (x, y_0, y, Y), the point (x, y_0, y, Y, z_0) is feasible to (SIP) if and only if there exists $(\lambda, \mu) \in I\!R^l \times I\!R^n_+$ such that (w, z_0) is feasible to (NSDP′). □

Finally, the optimality between (NSDP′) and (SIP) is obtained as follows.

Theorem 5.2.15. *Suppose that Assumption* 5.2.4–(i) *holds, and that* $\Xi = I\!R^m$. *If* (w, z_0) *is a global optimum to* (NSDP′), *then its subvector* (x, y_0, y, Y, z_0) *and* x *are also global optima for* (SIP) *and* (DRERM), *respectively.*

Proof. Let (w, z_0) be a global optimum to (NSDP′). Assume that its subvector (x, y_0, y, Y, z_0) is not a global optimum of (SIP). Note that (x, y_0, y, Y, z_0) is a feasible solution to (SIP) by Proposition 5.2.14. Since (x, y_0, y, Y, z_0) is not a global optimum of (SIP), there exists a feasible solution (x', y_0', y', Y', z_0') such that

$$z_0' + y_0' + \mu_0^\top y' + \left\langle \gamma_2 \Sigma_0 + \mu_0 \mu_0^\top, Y' \right\rangle < z_0 + y_0 + \mu_0^\top y + \left\langle \gamma_2 \Sigma_0 + \mu_0 \mu_0^\top, Y \right\rangle . \tag{5.31}$$

Moreover, by Proposition 5.2.14, the feasible solution (x', y_0', y', Y', z_0') of (SIP) is also feasible to (NSDP′) for some $(\lambda', \mu') \in I\!R^l \times I\!R^n_+$. This statement and (5.31) contradict each other; thus, (x, y_0, y, Y, z_0) is a global optimum of (SIP).

Since (DRERM) is equivalent to (SIP), x, which is the subvector of the global optimum (x, y_0, y, Y, z_0) of (SIP), is also a global optimum to (DRERM). Thus, the optimal value of (NSDP′) coincides with those of (SIP) and (DRERM), respectively. □

Remark 5.2.4. *Zhu et al.* [133] *have only shown a conservative NSDP approximation for problem* (5.4)*; that is, the subvector* $x \geq 0$ *of a global optimal solution of the conservative approximated NSDP may not globally solve* (5.4) *in general. However, as Theorem* 5.2.15 *and Corollary* 5.2.12 *state, if* Ξ *is* $I\!R^m$ *or a single ellipsoid, the variable* $x \in S$ *of a global optimal point obtained from* (NSDP) *or* (NSDP′) *solves* (DRERM).

5.2.2 Convexity of NSDP

First, the sufficient condition is presented under which (NSDP) and (NSDP′) are convex.

Assumption 5.2.16. *The mapping* F *is affine with respect to* x, *i.e.,*

$$F(x, \xi) := M(\xi)x + q(\xi),$$

where $M\colon \Xi \to I\!R^{n \times n}$ *and* $q\colon \Xi \to I\!R^n$. *Here, the* (i, j)*-entry of* $M(\xi)$ *is denoted by* $(M(\xi))_{ij} := \left(m^{ij}\right)^\top \xi + m_0^{ij}$, *and the* i*-th element of* $q(\xi)$ *is* $(q(\xi))_i := \left(q^i\right)^\top \xi + q_0^i$, *where* $m^{ij}, q^i \in I\!R^m$, $m_0^{ij}, q_0^i \in I\!R$. *Hence,* $c^i(x)$ *and* $c_0^i(x)$ *defined in Assumption* 5.2.4 *can be rewritten as follows:*

$$c^i(x) := q^i + \bar{M}_i x \in I\!R^m, \; i = 1, 2, \ldots, n,$$
$$c_0^i(x) := q_0^i + (\bar{m}_0^i)^\top x \in I\!R, \; i = 1, 2, \ldots, n,$$

where

$$\bar{M}_i := [m^{i,1}, m^{i,2}, \ldots, m^{i,n}] \in I\!R^{m \times n}, \quad i = 1, 2, \ldots, n,$$
$$\bar{m}_0^i := [m_0^{i,1}, m_0^{i,2}, \ldots, m_0^{i,n}]^\top \in I\!R^n, \quad i = 1, 2, \ldots, n.$$

Remark 5.2.5. *In Assumption* 5.2.16, *suppose that*

$$M(\xi) = M \cdot \mathrm{repvec}(\xi; n) + M_0, \quad q(\xi) = Q\xi + q_0,$$

where

$$M := \begin{bmatrix} (m^{1,1})^\top & (m^{1,2})^\top & \dots & (m^{1,n})^\top \\ (m^{2,1})^\top & (m^{2,2})^\top & \dots & (m^{2,n})^\top \\ \vdots & \vdots & \ddots & \vdots \\ (m^{n,1})^\top & \dots & \dots & (m^{n,n})^\top \end{bmatrix} \in I\!R^{n\times mn},$$

$$\mathrm{repvec}(\xi;n) := \begin{bmatrix} \xi & & & \\ & \xi & & \\ & & \ddots & \\ & & & \xi \end{bmatrix} \in I\!R^{mn\times n}, \quad M_0 := [\bar{m}_0^1, \bar{m}_0^2, \dots, \bar{m}_0^n]^\top \in I\!R^{n\times n},$$

$$Q := [q^1, q^2, \dots, q^n]^\top \in I\!R^{n\times m}, \;\; q_0 := [q_0^1, q_0^2, \dots, q_0^n] \in I\!R^n.$$

Then, $F(x,\xi)$ can also be written as

$$F(x,\xi) = (M \cdot \mathrm{repvec}(\xi;n) + M_0)\, x + (Q\xi + q_0).$$

Let us introduce the convexity of nonlinear matrix-valued functions and its related property.

Definition 5.2.17 (Shapiro [111])**.** *A nonlinear matrix-valued function $X\colon I\!R^m \to \mathbb{S}^n$ is said to be positive semidefinite (psd-) convex if*

$$X(\gamma x + (1-\gamma)y) - \gamma X(x) - (1-\gamma)X(y) \preceq O \tag{5.32}$$

for all $x, y \in I\!R^m$ and $\gamma \in [0,1]$.

Proposition 5.2.18. *The mapping X is psd-convex if and only if for any $v \in I\!R^n$ with $v_1 = 1$, the function $\phi(\cdot; v)\colon I\!R^m \to I\!R$ defined by*

$$\phi(x;v) := \begin{bmatrix} 1, & v^\top \end{bmatrix} X(x) \begin{bmatrix} 1 \\ v \end{bmatrix}$$

is convex with respect to $x \in I\!R^m$.

Proof. Lemma 5.2.7 ensures that matrix inequality (5.32) is equivalent to

$$\begin{bmatrix} 1, & v'^\top \end{bmatrix} (X(\gamma x + (1-\gamma)y) - \gamma X(x) - (1-\gamma)X(y)) \begin{bmatrix} 1 \\ v' \end{bmatrix} \le 0,$$

for any $v' \in I\!R^{n-1}$. Hence, we have

$$\phi(\gamma x + (1-\gamma)y; v) \le \gamma\phi(x;v) + (1-\gamma)\phi(y;v)$$

for any $v \in I\!R^n$ with $v_1 = 1$. Therefore, X is psd-convex if and only if $\phi(\cdot, v)$ is convex with respect to $x \in I\!R^m$ for every $v \in I\!R^n$. □

We show the convexity of (NSDP) and (NSDP′).

Theorem 5.2.19. *Suppose that Assumption* 5.2.16 *holds and that the matrix $M(\xi)$ defined in Assumption* 5.2.16 *satisfies the following condition: There exists $\beta_0 > 0$ such that*

$$\inf_{\xi \in \Xi, \|v\|=1} v^\top M(\xi) v \geq \beta_0. \tag{5.33}$$

Then the matrix-valued function $-\mathcal{D}_\alpha$ is psd-convex for all $\alpha \geq 1/(2\beta_0)$; thus, (NSDP) *and* (NSDP$'$) *are convex.*

Proof. Note that if the matrix-valued function $-\mathcal{D}_\alpha$ is psd-convex, (NSDP) and (NSDP$'$) are convex optimization problems. Therefore, we verify that $-\mathcal{D}_\alpha$ is psd-convex for all $\alpha \geq 1/(2\beta_0)$.

Suppose that $\alpha \geq 1/(2\beta_0)$. Proposition 5.2.18 states that $-\mathcal{D}_\alpha$ is psd-convex if and only if for all $\xi \in I\!R^m$, the following function $\phi_\alpha(\cdot, \xi) \colon \mathcal{W} \to I\!R$ is convex with respect to w:

$$\phi_\alpha(w;\xi) := \begin{bmatrix} 1, \xi^\top \end{bmatrix} (-\mathcal{D}_\alpha(w)) \begin{bmatrix} 1 \\ \xi \end{bmatrix} = -y_0 - \xi^\top y - \xi^\top Y \xi + \omega_\alpha(x, \lambda, \mu; \xi),$$

where the last equality follows from (5.23).

Now, since the function $\phi_\alpha(\cdot, \xi)$ is linear with respect to (y_0, y, Y), it suffices to show that ω_α is convex with respect to (x, λ, μ) for all $\alpha \geq 1/(2\beta_0)$. The Hessian of ω_α in regard to (x, λ, μ) is given by

$$\nabla^2_{(x,\lambda,\mu)} \omega_\alpha(x, \lambda, \mu; \xi) = \alpha \begin{bmatrix} M(\xi)^\top M(\xi) & (M(\xi) - \frac{1}{\alpha} I)^\top A^\top & -M(\xi)^\top + \frac{1}{\alpha} I \\ A(M(\xi) - \frac{1}{\alpha} I) & AA^\top & -A \\ -M(\xi) + \frac{1}{\alpha} I & -A^\top & I \end{bmatrix}.$$

By considering the Schur complement of the above matrix,

$$\begin{aligned} &\begin{bmatrix} M(\xi)^\top M(\xi) & (M(\xi) - \frac{1}{\alpha} I)^\top A^\top \\ A(M(\xi) - \frac{1}{\alpha} I) & AA^\top \end{bmatrix} - \begin{bmatrix} -M(\xi)^\top + \frac{1}{\alpha} I \\ -A \end{bmatrix} \begin{bmatrix} -M(\xi) + \frac{1}{\alpha} I & -A^\top \end{bmatrix} \\ &= \frac{1}{\alpha} \begin{bmatrix} (M(\xi)^\top + M(\xi)) - \frac{1}{\alpha} I & O \\ O & O \end{bmatrix} \succeq O \end{aligned}$$

if and only if $\nabla^2_{(x,\lambda,\mu)} \omega_\alpha(x, \lambda, \mu; \xi) \succeq O$. Since $(M(\xi)^\top + M(\xi)) - 1/\alpha\, I \succeq O$ from $\alpha \geq 1/(2\beta_0)$, it can be easily seen that $\nabla^2_{(x,\lambda,\mu)} \omega_\alpha(x, \lambda, \mu; \xi) \succeq O$, i.e., $-\mathcal{D}_\alpha$ is psd-convex for all $\alpha \geq 1/(2\beta_0)$. Hence, (NSDP) and (NSDP$'$) are convex optimization problems. □

Remark 5.2.6. *Condition* (5.33) *is rather restrictive for some applications. One remedy is to add a proximal term $\epsilon(x - x^k)$ to the mapping F, where $\epsilon > 0$ is a sufficiently small constant.*

Remark 5.2.7. *When $S = I\!R^n_+$, problem* (5.4) *for the SLCP proposed by Zhu et al.* [133] *may not be reformulated as a convex NSDP because the objective function $\Psi(x, \xi) = \|x \circ F(x, \xi)\|_2^2$ is not convex with respect to x in general.*

Although we adopt the regularized gap function for the NSDP approximation, similar results may also be obtained by utilizing another merit function, such as $f_\infty(x, \xi) := \max_{z \in S} \langle F(x, \xi), x - z \rangle$. However, it would be necessary to discuss whether the DRERM with f_∞ is a reasonable method for solving the SVI. In fact, the ERM with f_∞ may be unsuitable to measure the distance to solutions of SVI (5.1) because $f_\infty(x, \xi)$ takes $+\infty$ for some $x \in S$ and is not differentiable in general. For such reasons, we did not adopt f_∞ for (DRERM).

5.3 Numerical experiments

This section provides numerical results to demonstrate the validity of the DRERM model. In particular, we first compare the DRERM with the ERM proposed by Luo and Lin [80] in terms of robustness. Second, we quantitatively investigate the robustness of solutions obtained from the DRERM model when the confidence parameters γ_1 and γ_2 for the mean and variance of the ambiguity set $\mathscr{P}$, respectively, are gradually changed.

Throughout this section, we use the following example.

Example 5.3.1 (Two-person noncooperative games)**.** *Two players are competing with each other to minimize their own cost functions. Each player $\nu \in \{1,2\}$ solves the following optimization problem:*

$$\begin{array}{ll} \min\limits_{x^\nu \in I\!R^{n_\nu}} & \frac{1}{2}\left(x^\nu\right)^\top M_\nu x^\nu + v^\nu(x^\nu, x^{-\nu}, \xi) + q^\nu(\xi)^\top x^\nu \\ s.t. & A_\nu x^\nu \le b^\nu, \end{array} \tag{5.34}$$

where $M_\nu \in \mathbb{S}^{n_\nu}_{++}$, $A_\nu \in I\!R^{l_\nu \times n_\nu}$, $b^\nu \in I\!R^{l_\nu}$, and $q^\nu(\xi) \in I\!R^{n_\nu}$. Here, $v^\nu(x^\nu, x^{-\nu}, \xi)$ is a zero-sum function defined by

$$v^\nu(x^\nu, x^{-\nu}, \xi) := \begin{cases} \left(x^1\right)^\top R(\xi) x^2 & \text{if } \nu = 1, \\ -\left(x^2\right)^\top R(\xi)^\top x^1 & \text{if } \nu = 2, \end{cases}$$

where $R(\xi) \in I\!R^{n_1 \times n_2}$, and $x^{-\nu} \in I\!R^{n_{-\nu}}$ is the decision variable of the rival player. □

The above noncooperative game can be reformulated as SVI (5.1) when the mapping $F(\cdot,\xi)\colon I\!R^n \to I\!R^n$ and the set $S \subset I\!R^n$ are given as follows:

$$\begin{aligned} F(x,\xi) &= \begin{bmatrix} M_1 & R(\xi) \\ -R(\xi)^\top & M_2 \end{bmatrix} x + \begin{bmatrix} q^1(\xi) \\ q^2(\xi) \end{bmatrix}, \\ S &= \left\{ x \in I\!R^n \;\middle|\; \begin{bmatrix} A_1 & O \\ O & A_2 \end{bmatrix} x \le \begin{bmatrix} b^1 \\ b^2 \end{bmatrix} \right\}, \\ x &= \left[(x^1)^\top,\ (x^2)^\top \right]^\top \in I\!R^{n_1+n_2}. \end{aligned} \tag{5.35}$$

Note that it is easy to verify that the coefficient matrix in (5.35) satisfies the assumption of Theorem 5.2.19; hence, we solve a convex NSDP in the experiments.

We generate numerical instances of problem (5.34) according to the following manners:

- We set $n_1 = n_2 = 2$, $m = n_1 n_2 + 2 = 6$, and $l_1 = l_2 = 2$.
- The matrix M_ν is generated by $L_\nu {L_\nu}^\top + I$, where the matrix $L_\nu \in I\!R^{2\times 2}$ is lower triangular and its elements are randomly generated from the interval $[-5, 5)$.
- Each element of the matrix $A_\nu \in I\!R^{2\times 2}$ and the vector $b^\nu \in I\!R^2$ is randomly generated from $[-2, 2)$ and $[0, 10)$, respectively.
- We set the regularization parameter α by $1/\beta_0$ to ensure that the derived NSDP is convex, where β_0 is the minimum eigenvalue of the matrix

$$\begin{bmatrix} M_1 & O_{n_1\times n_2} \\ O_{n_2 \times n_1} & M_2 \end{bmatrix} \in I\!R^{4\times 4}.$$

- We define the random variable $\xi \in I\!R^m$ by $\xi = [\xi_1, \dots, \xi_6]^\top$.
- The matrix $R(\xi)$ is defined by

$$R(\xi) := \begin{bmatrix} \xi_1 & \xi_2 \\ \xi_3 & \xi_4 \end{bmatrix} + R_0 \in I\!R^{2\times 2},\ R_0 := \begin{bmatrix} r_0^{1,1} & r_0^{1,2} \\ r_0^{2,1} & r_0^{2,2} \end{bmatrix} \in I\!R^{2\times 2}$$

where $r_0^{i,j}$, $i, j = 1, 2$ are nominal values generated randomly from $[-5, 5)$.

- The vector $q(\xi) := \left(q^1(\xi)^\top, q^2(\xi)^\top\right)^\top \in I\!R^4$ is defined by

$$q(\xi) = Q\xi + q_0,$$

where

$$Q = \begin{bmatrix} 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix},\ q_0 = -\begin{bmatrix} M_1 & R_0 \\ -R_0^\top & M_2 \end{bmatrix} x_0^*,$$

and the vector $x_0^* \in I\!R^4$ is randomly generated from $[-2, 2)$.

In the experiments, all programs are implemented with Python 3.8 and run on a machine with Intel Core i7-8700K @ 3.70GHz CPU and 32 GB RAM.

5.3.1 Comparison to the ERM model

Here, we suppose that $\Xi = I\!R^6$ and ξ follows the normal distribution $\mathcal{N}(\mu_0, \Sigma_0)$, where the mean μ_0 and the variance-covariance matrix Σ_0 are given as follows:

$$\mu_0 = 0,\ \Sigma_0 = \begin{bmatrix} 2 & 1.6 & \cdots & 1.6 \\ 1.6 & 2 & \cdots & 1.6 \\ \vdots & \vdots & \ddots & \vdots \\ 1.6 & 1.6 & \cdots & 2 \end{bmatrix}. \tag{5.36}$$

In the ERM model, we use the regularized gap function f_α proposed by Luo and Lin [80] as the merit function f. In the experiments, because it is difficult to exactly compute the expected value $\mathbb{E}[f_\alpha(x, \xi)]$, we obtain its approximate value using a quasi-Monte Carlo method described below:

$$\mathbb{E}[f_\alpha(x, \xi)] \approx \theta^k(x) := \frac{1}{N_k} \sum_{\hat{\xi}^k \in \Xi^k} f_\alpha(x, \hat{\xi}^k) p(\hat{\xi}^k),$$

where the uniform random vector $\hat{\xi}^k \in \Xi^k$ is generated by

$$\hat{\xi}^k = \left((\mu_0 - 3\sqrt{2}) + (\mu_0 + 3\sqrt{2})\zeta^i\right) \mathbf{1}_m,$$

and ζ^i is a Sobol point from the interval $[0, 1)$. The set $\Xi^k := \{\hat{\xi}^i \mid i = 1, 2, \dots, N_k\} \subset \Xi$ is the collection of the samples $\hat{\xi}^k$, which approximates the support Ξ, and $p(\cdot)$ is the probability density function of the normal distribution $\mathcal{N}(\mu_0, \Sigma_0)$. Note that as the number of samples N_k and dimensions m increased, it may face underflow and subsequently fail to evaluate $\theta^k(x)$.

To avoid this, we multiply $\theta^k(x)$ by $1/p(\mu_0)$. Summarizing the above arguments, we solve the following approximate problem for (ERM) with the regularized gap function:

$$\begin{array}{ll} \min & \theta^k(x)/p(\mu_0) \\ \text{s.t.} & x \in S, \end{array} \tag{5.37}$$

We use the SLSQP package, which is based on sequential quadratic programming methods, in the `Scipy.Optimize` module to obtain a solution to problem (5.37). The initial point is set to 0, and the termination criterion for the residual of the Karush–Kuhn–Tucker condition is set to 10^{-7}.

In the DRERM, because we know the exact values μ_0 and Σ_0 in advance, the ambiguity set $\mathscr{P}$ is given by (5.5). When $\Xi = I\!R^6$ and $\mathscr{P}$ is given as (5.5), (DRERM) can be reformulated as the following NSDP, which can be regarded as the special case of (NSDP′):

$$\begin{array}{ll} \displaystyle\min_{(x,\lambda,y_0,y,Y)} & y_0 + \mu_0^\top y + \langle \Sigma_0 + \mu_0\mu_0^\top, Y \rangle \\ \text{s.t.} & \mathcal{D}_\alpha(x,\lambda,y_0,y,Y) \succeq O, \\ & Ax \le b,\ \lambda \in I\!R^2_-, \end{array} \tag{5.38}$$

where $I\!R^2_- \coloneqq \{\lambda \in I\!R^2 \mid \lambda \le 0\}$. To solve (5.38), we utilize an interior point method, which is a hybrid method of [125] and [126]. The initial point and termination criterion are the same as the method for (5.37).

We prepare 10 numerical instances of SVI (5.1) and solve them via (5.37) and (5.38), where we set two cases where $N_k = 80$ and $N_k = 10000$ in (5.37). Let x^{i*}_{ERM} and x^{i*}_{DRERM} be solutions to (5.37) and (5.38) at the i-th instance, respectively. In what follows, for a realization $\bar{\xi}^j$ of the random variable ξ, f^{ij}_{ERM} and f^{ij}_{DRERM} respectively denote $f_\alpha(x^{i*}_{\mathrm{ERM}}, \bar{\xi}^j)$ and $f_\alpha(x^{i*}_{\mathrm{DRERM}}, \bar{\xi}^j)$ for simplicity.

To quantitatively evaluate the solutions x^{i*}_{ERM} and x^{i*}_{DRERM}, we conduct the following steps:

(i) Generate $N \coloneqq 5000$ realizations $\{\bar{\xi}^j\}_{j=1}^N$, where each realization $\bar{\xi}^j$ follows the normal distribution $\mathcal{N}(\mu_1, \Sigma_1)$. Here, μ_1 and Σ_1 are respectively the perturbations of μ_0 and Σ_0 as follows:

$$\mu_1 \coloneqq \mu_0 + \delta_\mu,\ \Sigma_1 \coloneqq \Sigma_0 + \Delta_\Sigma,$$

where each element of $\delta_\mu \in I\!R^6$ and $\Delta_\Sigma \in \mathbb{S}^6$ are uniformly generated from the interval $[-0.1, 0.1)$.

(ii) Compute the regularized gap function values $\{f^{ij}_{\mathrm{ERM}}\}_{j=1}^N$ and $\{f^{ij}_{\mathrm{DRERM}}\}_{j=1}^N$ by using the realizations $\{\bar{\xi}^j\}_{j=1}^N$ for each solution.

(iii) Evaluate the solutions x^{i*}_{ERM} and x^{i*}_{DRERM} by using the following five indicators, which represent the rates of change (RC):

- Minimum:
$$(\min_j f^{ij}_{\mathrm{DRERM}} - \min_j f^{ij}_{\mathrm{ERM}}) / \min_j f^{ij}_{\mathrm{ERM}}, \tag{5.39}$$

- Maximum:
$$(\max_j f^{ij}_{\mathrm{DRERM}} - \max_j f^{ij}_{\mathrm{ERM}}) / \max_j f^{ij}_{\mathrm{ERM}}, \tag{5.40}$$

- Mean:

$$(\text{mean } f^i_{\text{DRERM}} - \text{mean } f^i_{\text{ERM}})/\text{mean } f^i_{\text{ERM}}, \tag{5.41}$$

where mean $f^i_\cdot := \frac{1}{N}\sum_{j=1}^N f^{ij}_\cdot$.

- Median:

$$(\text{med } f^i_{\text{DRERM}} - \text{med } f^i_{\text{ERM}})/\text{med } f^i_{\text{ERM}}, \tag{5.42}$$

where med $f^i_\cdot := (f^{i[N/2]}_\cdot + f^{i[N/2+1]}_\cdot)/2$, and $f^{i[j]}_\cdot$ denotes the j-th largest regularized gap function value in the 5000 realizations.

- Standard deviation (SD):

$$(\text{sd } f^i_{\text{DRERM}} - \text{sd } f^i_{\text{ERM}})/\text{sd } f^i_{\text{ERM}}, \tag{5.43}$$

where sd $f^i_\cdot := \sqrt{\frac{1}{N-1}\sum_{j=1}^N (f^{ij}_\cdot - \text{mean } f^i_\cdot)^2}$.

The computational results are shown in Figure 5.1. In each graph, the horizontal and the vertical axes represent the instance number and the RC, respectively. Figures 5.1(a) and 5.1(b) indicate the RC evaluated by (5.39) for $N_k = 80$ and $N_k = 10000$, respectively, and Figures 5.1(c) and 5.1(d) represent the RC evaluated by (5.40)–(5.43) for each N_k. Note that the vertical axis of Figure 5.1(a) is a logarithmic scale.

First, we focus on the minimum values, i.e., Figures 5.1(a) and 5.1(b). We observe that for most of the instances of $N_k = 80$ and $N_k = 10000$, the minimum values of the ERM tend to be small compared with the DRERM. In particular, the 8th instance in Figure 5.1(a) indicates a significant difference between the ERM and DRERM models. Indeed, $\min_j f^{8j}_{\text{ERM}} = 0.0023$ and $\min_j f^{8j}_{\text{DRERM}} = 1.5784$, and they have a 690-fold difference. In the case of $N_k = 10000$, the gaps between the ERM and DRERM are small for all instances compared with $N_k = 80$.

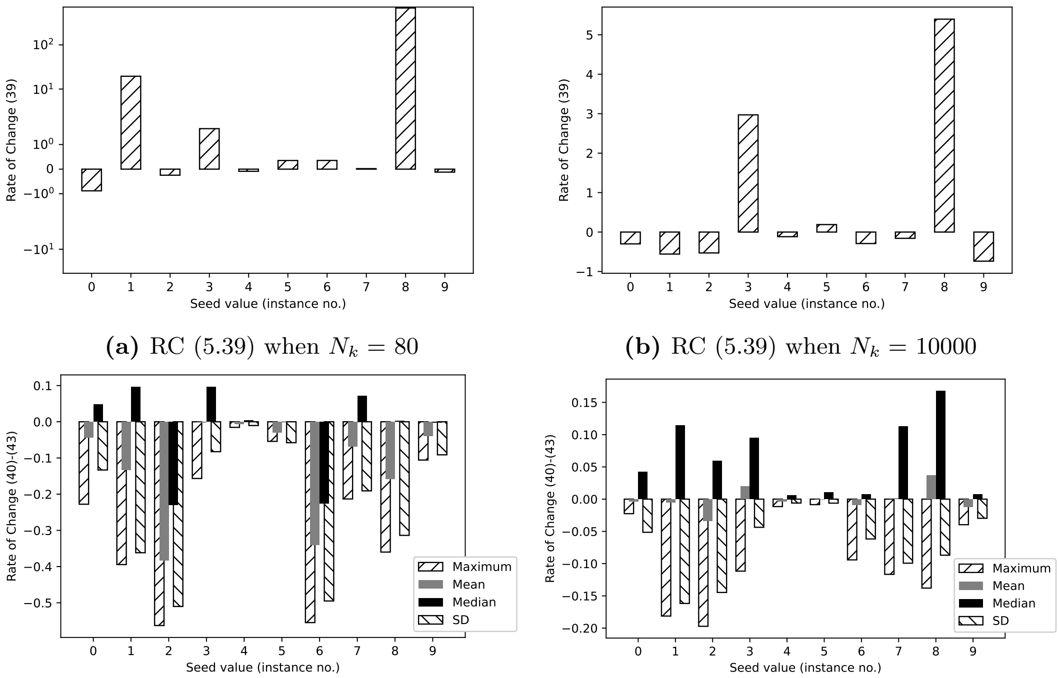
Next, we focus on Figures 5.1(c) and 5.1(d). Notably, the values of the gap function of maximum and SD on the DRERM are smaller than the ERM for all instances for $N_k = 80$ and $N_k = 10000$. This is an important result that shows that the DRERM is reasonably designed to consider the distributionally worst case in terms of the expected value of the regularized gap function.

From the above results, we confirm that the DRERM can obtain more robust solutions that consider outliers, while the ERM is not as robust as the DRERM even when N_k is sufficiently large in spite of using the exact distribution function for evaluating the expected value. This is because the ERM is designed to minimize the expected value of the regularized gap function; hence, it cannot directly consider the variance and maximum value. In fact, the median of f^{ij}_{ERM} with $N_k = 10000$ is less than the DRERM; however, outliers of realizations $\bar{\xi}^j$ adversely affect the mean of the regularized gap values. As a result, the difference between the mean of f^{ij}_{ERM} with $N_k = 10000$ and that of f^{ij}_{DRERM} is insignificant.

5.3.2 Analysis of solution by varying confidence parameters

In this section, we assume that $\Xi = \mathbb{R}^6$, and the estimated mean $\tilde{\mu}_0$ and variance-covariance matrix $\tilde{\Sigma}_0$ are given as follows:

$$\tilde{\mu}_0 := \mu_0 + u^6,\ \tilde{\Sigma}_0 := \Sigma_0 + U_6,$$



(a) RC (5.39) when $N_k = 80$ **(b)** RC (5.39) when $N_k = 10000$

(c) RC (5.40)–(5.43) when $N_k = 80$ **(d)** RC (5.40)–(5.43) when $N_k = 10000$

Figure 5.1: The rate of change between the ERM and the DRERM.

where each element of $u^6 \in I\!R^6$ and $U_6 \in \mathbb{S}^6$ are uniformly generated from $[-0.25, 0.25)$ and $[-0.2, 0.2)$, respectively. Here, the true μ_0 and Σ_0 are the same as (5.36), and the confidence regions of $\tilde{\mu}_0$ and $\tilde{\Sigma}_0$ in the ambiguity set $\mathscr{P}$ are given as follows:

$$\left(\mathbb{E}_P\left[\xi\right] - \tilde{\mu}_0\right)^\top \tilde{\Sigma}_0^{-1}\left(\mathbb{E}_P\left[\xi\right] - \tilde{\mu}_0\right) \le \gamma_1, \tag{5.44}$$

$$\mathbb{E}_P\left[\left(\xi - \tilde{\mu}_0\right)\left(\xi - \tilde{\mu}_0\right)^\top\right] \preceq \gamma_2 \tilde{\Sigma}_0. \tag{5.45}$$

In this setting, we solve the following NSDP:

$$\begin{array}{ll} \displaystyle\min_{(x,\lambda,y_0,y,Y,z_0)} & z_0 + y_0 + \tilde{\mu}_0^\top y + \langle \gamma_2 \tilde{\Sigma}_0 + \tilde{\mu}_0 \tilde{\mu}_0^\top, Y \rangle \\ \text{s.t.} & z_0 \ge \sqrt{\gamma_1}\left\|\tilde{\Sigma}_0^{1/2}(y + 2Y\tilde{\mu}_0)\right\|, \\ & \mathcal{D}_\alpha(x, \lambda, y_0, y, Y) \succeq O, \\ & Ax \le b, \ \lambda \in I\!R^2_-. \end{array} \tag{5.46}$$

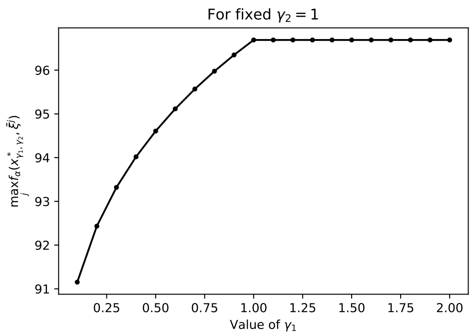
Here, we solve (5.46) using the interior point method, which is the same method for solving (5.38). The initial point is set as 0, and the stopping criterion is 10^{-7}. Note that we set $\alpha > 0$ to ensure that problem (5.46) is convex. Let $x^*_{\gamma_1,\gamma_2}$ be a solution of problem (5.46) for given γ_1 and γ_2.

In the first experiment, we quantitatively analyze the characteristics of the solutions in the case where γ_1 is incremented by 0.1 from 0.1 to 2, and γ_2 is set to 1 or 2. We prepare realizations $\{\bar{\xi}^j\}_{j=1}^N$, where each $\bar{\xi}^j$ follows $\mathcal{N}(\mu_0, \Sigma_0)$ and $N = 5000$. After obtaining a solution $x^*_{\gamma_1,\gamma_2}$, we compute the maximum, mean, and SD of $\{f_\alpha(x^*_{\gamma_1,\gamma_2}, \bar{\xi}^j)\}_{j=1}^N$.

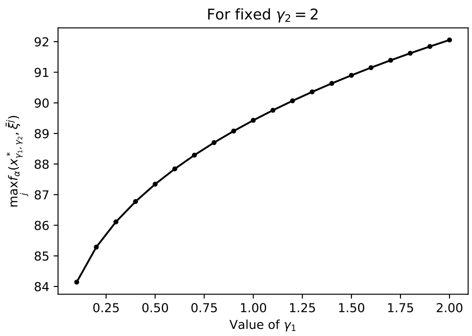
Figure 5.2 shows the results of the first experiment. In each graph, the horizontal and vertical axes represent the values of γ_1 and the regularized gap function, respectively. The curves in Figures 5.2(a) and 5.2(b) indicate the maximum of $\{f_\alpha(x^*_{\gamma_1,\gamma_2}, \bar{\xi}^j)\}_{j=1}^N$ for fixed $\gamma_2 = 1$ and $\gamma_2 = 2$, respectively, and Figures 5.2(c) and 5.2(d) represent the mean and SD of $\{f_\alpha(x^*_{\gamma_1,\gamma_2}, \bar{\xi}^j)\}_{j=1}^N$ for fixed $\gamma_2 = 1$ and $\gamma_2 = 2$, respectively.

In Figures 5.2(a) and 5.2(c) (when $\gamma_2 = 1$), the maximum, mean, and SD of regularized gap function values increase as γ_1 increases. However, Figures 5.2(b) and 5.2(d) (when $\gamma_2 = 2$) indicate that the values of the maximum and SD are entirely smaller than the case where $\gamma_2 = 1$; we will discuss the reason in the next experiment. In particular, from Figure 5.2(d), the curve of the mean gradually decreases for $0.1 \leq \gamma_1 \leq 1$, unlike the case where $\gamma_2 = 1$. Moreover, Figures 5.2(a) and 5.2(c) indicate that the optimal solutions $x^*_{\gamma_1,1}$ to problem (5.46) are not changed for $1 \leq \gamma_1 \leq 2$.

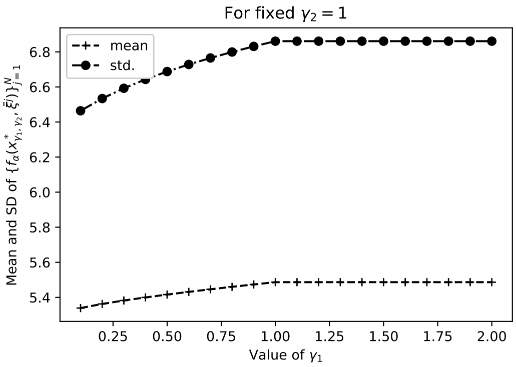
To summarize the first experiment, as γ_1 increases, the solution $x^*_{\gamma_1,\gamma_2}$ tends to focus on decreasing the mean of realizations of f_α for the case of $\gamma_2 = 2$. Moreover, the mean increases as γ_1 becomes larger when $\gamma_2 = 1$. This implies that the uncertainty of the estimated variance-covariance $\tilde{\Sigma}_0$ is not sufficiently considered for the case of $\gamma_2 = 1$.



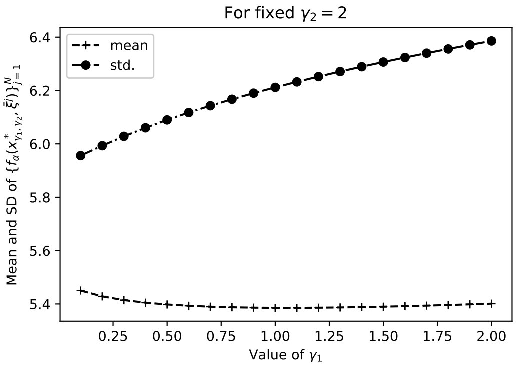
(a) Maximum for fixed $\gamma_2 = 1$



(b) Maximum for fixed $\gamma_2 = 2$



(c) Mean and SD for fixed $\gamma_2 = 1$



(d) Mean and SD for fixed $\gamma_2 = 2$

Figure 5.2: Maximum, mean, and SD of 5000 realizations of the regularized gap function when γ_1 is varied.

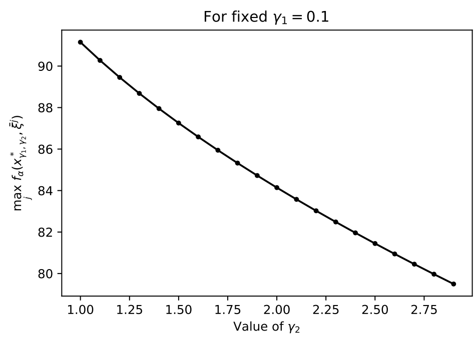
In the second experiment, we investigate the characteristics of the solutions in the case where γ_2 is incremented by 0.1 from 1 to 3, and γ_1 is set to 0.1 or 1. We prepare 5000 realizations $\{\bar{\xi}^j\}_{j=1}^N$, which are the same samples used in the first experiment and compute the maximum, mean, and SD of $\{f_\alpha(x^*_{\gamma_1,\gamma_2}, \bar{\xi}^j)\}_{j=1}^N$ for the solution $x^*_{\gamma_1,\gamma_2}$ to problem (5.46).

Figure 5.3 depicts the results of the second experiment. In particular, Figures 5.3(a) and 5.3(b) are $\max_j\ f_\alpha(x^*_{\gamma_1,\gamma_2}, \bar{\xi}^j)$ for fixed $\gamma_1 = 0.1$ and $\gamma_1 = 1$, respectively. Figures 5.3(c) and 5.3(d) are the mean and SD of $\{f_\alpha(x^*_{\gamma_1,\gamma_2}, \bar{\xi}^j)\}_{j=1}^N$ for fixed $\gamma_1 = 0.1$ and $\gamma_1 = 1$, respectively.

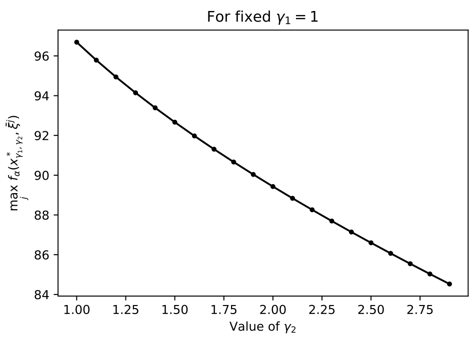
For fixed $\gamma_1 = 0.1$, the maximum and SD gradually decrease as γ_2 increases, whereas the mean increases. For fixed $\gamma_1 = 1$, the maximum and SD also decrease; however, the values of f_α are larger than the case where $\gamma_1 = 0.1$ entirely. Moreover, there is a diminutive change in the curve of the mean in Figure 5.3(d) compared with that of Figure 5.3(c).

To summarize the second experiment, as γ_2 increases, the DRERM outputs the solutions $x^*_{\gamma_1,\gamma_2}$ that tend to decrease the maximum and SD of f_α. This is because, by the definition of the moment ambiguity set (5.10), increasing γ_2 leads to conservative behavior regarding the variance of ξ. Consequently, f_α also behaves conservatively, and its outlier tends to be decreased as well. Meanwhile, when γ_2 is very large, the mean increases.

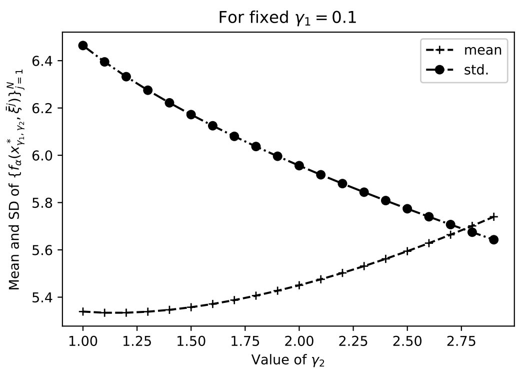
Consequently, from the results of both experiments, we confirm that there are trade-off relations between the mean and the SD, and the mean and the maximum, respectively, in response to the confidence parameters γ_1 and γ_2.



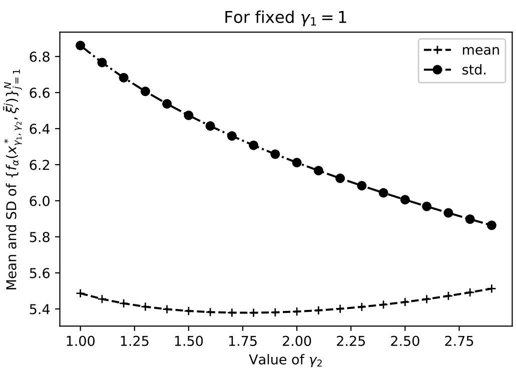
(a) Maximum for fixed $\gamma_1 = 0.1$



(b) Maximum for fixed $\gamma_1 = 1$



(c) Mean and SD for fixed $\gamma_1 = 0.1$



(d) Mean and SD for fixed $\gamma_1 = 1$

Figure 5.3: Maximum, mean, and SD of 5000 realizations of the regularized gap function when γ_2 is varied.

Remark 5.3.2. *When the support Ξ is compact, the reasonable γ_1 and γ_2 can be analytically obtained depending on the number of observations (refer to* [23]*). However, if Ξ is not compact, such as in this experiment, one can obtain desired γ_1, γ_2, and solutions to SVI* (5.1) *by* *approximating Ξ into a compact set.*

5.4 Concluding remarks

We have proposed a DRERM model for an SVI under uncertainty of distribution by incorporating the idea of the DRO into the ERM model with the regularized gap function. In particular, we have shown that the DRERM can be conservatively approximated into a deterministic NSDP, and under suitable assumptions, the solution of the NSDP also solves the DRERM. Furthermore, for the SVI whose mapping F is affine with respect to x, we have provided a sufficient condition of the regularization parameter of the regularized gap function to ensure that the reformulated NSDP is a convex optimization problem. Meanwhile, the reformulated NSDP proposed in the existing research is not convex in general. In numerical experiments, we have confirmed the reasonability of the DRERM model by comparing it with the ERM in terms of robustness, and we have analyzed their solutions by varying confidence parameters γ_1 and γ_2 included in the ambiguity set $\mathscr{P}$.

A remaining challenge is an NSDP approximation for more general cases of the following ambiguity sets described in [124]:

$$\mathscr{P}' = \left\{ P \in \mathscr{M}_\Xi \;\middle|\; \begin{array}{l} \mathbb{E}_P\left[\Psi_i(\xi)\right] = O, \;\; i = 1, 2, \ldots, t' \\ \mathbb{E}_P\left[\Psi_i(\xi)\right] \preceq O, \;\; i = t'+1, t'+2, \ldots, t \end{array} \right\},$$

where Ψ_i $(i = 1, 2, \ldots, t)$ is a symmetric matrix- or scalar-valued function over Ξ with measurable random components. We expect that our approach can be extended into the case of $\mathscr{P}'$ because the DRO with $\mathscr{P}'$ can be equivalently reformulated to a semi-infinite programming problem, such as (SIP), by assuming a 'Slater-type' condition on $\mathscr{P}'$.

Acknowledgments

This chapter is based on A. Hori, Y. Yamakawa, and N. Yamashita, "Distributionally robust expected residual minimization for stochastic variational inequality problem," *Optimization Methods and Software*, 2023. DOI: https://doi.org/10.1080/10556788.2023.2167995.

Chapter 6

Conclusions and future works

6.1 Conclusions

In this thesis, we studied some Nash games and those extensions under uncertainty, and we addressed stochastic variational inequalities with uncertain probability distributions.

In Chapter 3, we studied a multi-leader–follower game and reformulated it to an equilibrium problem with equilibrium constraints, which is essentially and numerically difficult to solve because of the complementarity constraints on each leader's problem, and they depend on other leaders' strategy vectors. To tackle such difficulties, we proposed a penalization technique. Using this approach, the game consisting of the penalized problems on each player is regarded as a classical standard differentiable Nash game. We solved the penalized Nash game with the Gauss–Seidel approach, and then we discussed convergence of the sequence generated by the method to a stationary point of the multi-leader–follower game under suitable assumptions. Finally, we demonstrated the validity of the algorithm through numerical experiments. Moreover, we considered a wholesale electricity market as an application of the game and analyzed the behavior of the market through numerical experiments.

In Chapter 4, we considered a two-stage distributionally robust Nash game in which each player solves a two-stage distributionally robust optimization parametrized by the other players' strategies. Existing studies had a limitation such as the linear decision rule in second-stage decisions and have only analyzed the game from the perspective of ex-post equilibrium. We established a more general result to demonstrate the sufficient condition for the existence of Nash equilibrium in the game and introduced a two-stage distributionally robust variational inequality to construct a solution method for finding the Nash equilibrium. As an application of the game, we considered a two-stage distributionally robust Cournot–Nash competition that appeared in the theory of industrial organization in microeconomics. We proved the existence of the equilibrium in the market under economically standard assumptions. Finally, we conducted some numerical experiments and analyzed the behavior of each player in a duopoly market from some perspectives. Particularly, we found that when one of the players unilaterally knows the almost exact probability distribution, the rival player behaves in passive involvement in the market because the shadow price of the player decreases.

In Chapter 5, we proposed a distributionally robust optimization model for stochastic variational inequality problems with uncertain probability distributions via expected residual minimization, referred to as a distributionally robust expected residual minimization (DRERM). In general, the DRERM is computationally much demanding because of the evaluation of expected residual functions and maximization with respect to the probability distribution. However, for a certain mapping $F(x, \xi)$, set S, and moment ambiguity set, we proved that the DRERM can be conservatively approximated as a deterministic nonlinear semidefinite programming problem. Furthermore, we showed that when the support Ξ of the probability distribution is given by $\mathbb{R}^m$ or a single ellipsoid, the globally optimal point of the semidefinite programming coincides with the optimal solution of the DRERM. We also provided a sufficient condition for the convexity of the nonlinear semidefinite programming when $F(x, \xi)$ is also affine with respect to x. In numerical experiments, we considered a stochastic Nash game in an almost sure formulation and compared our method and the existing expected residual minimization method that does not consider the uncertainty of distributions. We confirmed the validity of the proposed method in terms of robustness against the perturbation of probability distributions. We also obtained knowledge of how the optimal solution of the DRERM changes as the parameters of the moment ambiguity set change. These results may be useful for model designers when adjusting the robustness of the solution to the DRERM.

6.2 Future works

Although the main direction of future works may be the generalization of the concepts in the thesis to more inclusive decision-making situations, it is also essential to rigorously consider details that we ignored, as indicated in the following research issues, in order to apply the proposed method to real-world problems.

First, in the multi-leader–follower game presented in Chapter 3, when discussing the convergence of the proposed algorithm (Theorem 3.4.5), we assumed that the followers' response is uniquely determined for any strategies of leaders. However, this assumption does not necessarily hold in many real-world applications, and then an additional issue arises: whether the leaders should choose a strategy that is most or least convenient for themselves from the set of followers' responses. The solution concepts are called an *optimistic* and *pessimistic* leader–follower equilibrium, respectively, and few studies have addressed this issue in multi-leader–follower games. In the literature on bilevel optimization (Stackelberg games), the special case of multi-leader–follower games, some studies tackled the issue in recent years [68, 74, 75]. The multi-leader–follower games under uncertainty, such as the concept of distributional robustness in Chapter 4, are quite limited, while we found an approach in terms of robust Nash games [54]. Finally, a sensitivity analysis of the game would be interesting as well.

The future works on two-stage distributionally robust Nash games presented in Chapter 4 are desired to extend the framework to a generalized Nash game in which the strategy set depends on the other players' strategies. Along with this development, multistage (distributionally robust) stochastic variational inequalities should be extended to the framework of quasi-variational inequalities. In this thesis, we focused on the case in which random variables are independent of the decisions of players. However, in many real-world situations more general cases can be considered such as the games with private information and the game in which the probability of future scenarios depends on the decisions of players who have large impact on the market. In the course of Chapter 4, we introduced a reformulation of the game to two-stage distributionally robust variational inequalities, but the alternating-type method to solve them does not have global convergence in general; there are no studies that address this issue as with the one-stage case to the best of our knowledge. Other future studies include perturbation and sensitivity analysis of the equilibrium and extensions to multistage games.

As for the stochastic variational inequalities in Chapter 5, the proposed DRERM is limited to moment-based ambiguity sets. Hence, it would be interesting to adopt metric-based ambiguity sets such as the Wasserstein ambiguity set, which has gained attention in recent years in fields such as machine learning. We assumed that the function $F(x, \xi)$ is linear with respect to ξ, as in Assumption 5.2.4, for the purpose of obtaining an SDP approximation. However, this assumption may not hold in many real-world applications, such as when $F(x, \xi)$ is nonlinear, as in the BPR (Bureau of Public Roads) function used in transportation engineering. In more general cases, it may no longer be able to write a closed-form expression for $F(x, \xi)$ with respect to ξ. We think, however, such a closed-form assumption may be necessary for obtaining an SDP approximation. Therefore, it would be important to establish a method for approximating $F(x, \xi)$ linearly with respect to ξ for general stochastic variational inequalities so that Assumption 5.2.4 holds. Another interesting topic is the statistical robustness of the expected residual minimization method to tackle research issue (RI2) from a different perspective. Finally, it would be valuable to compare the robustness of the solution of a distributionally robust variational inequality [118] with the solution of our model.

Bibliography

[1] R.P. Agdeppa, N. Yamashita, and M. Fukushima, Convex expected residual models for stochastic affine variational inequality problems and its application to the traffic equilibrium problem, *Pacific J. Optim.*, **6** (2010), 3–19.

[2] M. Aghassi and D. Bertsimas, Robust game theory, *Math. Program.*, **107** (2006), 231–273.

[3] R. Andreani, L. D. Secchin and P. J. S. Silva, Convergence properties of a second order augmented Lagrangian method for mathematical programs with complementarity constraints, *SIAM J. Optim.*, **28** (2018), 2574–2600.

[4] K.J. Arrow and G. Debreu, Existence of an equilibrium for a competitive economy, *Econometrica*, **22** (1954), 265–290.

[5] A. Auslender, *Optimisation: méthodes numériques*, Masson, Paris, 1976.

[6] A. Bensoussan, Points de Nash dans le cas de fonctionnelles quadratiques et jeux differentiels lineaires a N personnes, *SIAM J. Control Optim.*, **12** (1974), 460–499.

[7] A. Ben-Tal and A. Nemirovski, Robust solutions of linear programming problems contaminated with uncertain data, *Math. Program.*, **88** (2000), 411–424.

[8] A. Ben-Tal and A. Nemirovski, Robust optimization - methodology and applications, *Math. Program.*, **92** (2002), 453–480.

[9] A. Ben-Tal and J. Zowe, Necessary and sufficient optimality conditions for a class of nonsmooth minimization problems, *Math. Program.*, **24** (1982), 70–91.

[10] K. Bertsimas, K. Natarajan, and C.P. Teo, Persistence in discrete optimization under data uncertainty, *Math. Program.*, **108** (2006), 251–274.

[11] J. R. Birge and F. Louveaux, *Introduction to Stochastic Programming*, Springer, New York, 2011.

[12] F. Bonnans and A. Shapiro, *Perturbation Analysis of Optimization Problems*, Springer, New York, 2000.

[13] S. Boyd and L. Vandenberghe, *Convex Optimization*, Cambridge University Press, 2004.

[14] X. Chen and M. Fukushima, Expected residual minimization method for stochastic linear complementarity problems, *Math. Oper. Res.*, **30** (2005), 1022–1038.

[15] X. Chen, T.K. Pong and R.J. Wets, Two-stage stochastic variational inequalities: an ERM solution procedure, *Math. Program.*, **165** (2017), 71–111.

[16] X. Chen, A. Shapiro and H. Sun, Convergence analysis of sample average approximation of two-stage stochastic generalized equations, *SIAM J. Optim.*, **29** (2019), 135–161.

[17] X. Chen, H. Sun and H. Xu, Discrete approximation of two-stage stochastic and distributionally robust linear complementarity problems, *Math. Program.*, **177** (2019), 255–289.

[18] X. Chen, R.J.B. Wets, and Y. Zhang, Stochastic variational inequalities: Residual minimization smoothing sample average approximations, *SIAM J. Optim.*, **22** (2012), 649–673.

[19] X. Chen, C. Zhang and M. Fukushima, Robust solution of monotone stochastic linear complementarity problems, *Math. Program.*, **117** (2009), 51–80.

[20] Y. Chen, B.F. Hobbs, S. Leyffer and T. Munson, Leader-follower equilibria for electric power and NO_x allowances markets, *Compt. Manag. Sci.*, **3** (2006), 307–330.

[21] F. H. Clarke, *Optimization and Nonsmooth Analysis*, Wiley, New York, 1983.

[22] P. Dasgupta and J. Collins, A survey of game theoretic approaches for adversarial machine learning in cybersecurity tasks, *AI Magazine*, **40** (2019), 31–43.

[23] E. Delage and Y. Ye, Distributionally robust optimization under moment uncertainty with application to data-driven problems, *Oper. Res.*, **58** (2010), 595–612.

[24] V. DeMiguel and H. Xu, A stochastic multiple-leader Stackelberg model: Analysis, computation, and application, *Oper. Res.*, **57** (2009), 1220–1235.

[25] K. Derinkuyu and M.Ç. Pınar, On the S-procedure and some variants, *Math. Method Oper. Res.*, **64** (2006), 55-77.

[26] G. Dhaene and J. Bouckaert, Sequential reciprocity in two-player, two-stage games: An experimental analysis, *Games Econ. Behav.*, **70** (2010), 289–303.

[27] S. Dirkse and M. Ferris, The path solver: a nommonotone stabilization scheme for mixed complementarity problems, *Optim. Methods Softw.*, **5** (1995), 123–156.

[28] F. Facchinei, A. Fischer and V. Piccialli, On generalized Nash games and variational inequalities, *Oper. Res. Lett.*, **35** (2007), 159–164.

[29] F. Facchinei and C. Kanzow, Generalized Nash equilibrium problems, *4OR*, **5** (2007), 173–210.

[30] F. Facchinei and J. S. Pang, *Finite-Dimensional Variational Inequalities and Complementarity Problems*, Springer, New York, 2003.

[31] F. Facchinei and J. Soares, A new merit function for nonlinear complementarity problems and a related algorithm, *SIAM J. Optim.*, **7** (1997), 225–247.

[32] C. Finn, P. Abbeel and S. Levine, Model-agnostic meta-learning for fast adaptation of deep networks, In *Proceedings of the 34th International Conference on Machine Learning* (ICML), Sydney, Australia, 2017.

[33] C. Finn, K. Xu and S. Levine, Probabilistic model-agnostic meta-learning, In *the 32nd Conference on Neural Information Processing Systems* (NeurIPS 2018), Montréal, Canada, 2018.

[34] L. Franceschi, P. Frasconi, S. Salzo, R. Grazzi and M. Pontil, Bilevel programming for hyperparameter optimization and meta-learning, In *Proceedings of the 35th International Conference on Machine Learning* (ICML), Stockholm, Sweden, 2018.

[35] M. Fukushima, Equivalent differentiable optimization problems and descent methods for asymmetric variational inequality problems, *Math. Program.*, **53** (1992), 99–110.

[36] M. Fukushima, Restricted generalized Nash equilibria and controlled penalty algorithm, *Comput. Manag. Sci.*, **8** (2011), 201–218.

[37] M. Fukushima and J.-S. Pang, Convergence of a smoothing continuation method for mathematical programs with complementarity constraints, In: M. Théra, R. Tichatschke (eds.) *Ill-posed Variational Problems and Regularization Techniques*, Lecture Notes in Economics and Mathematical Systems, **447** (1999), 99–110.

[38] R. S. Gibbons, *Game theory for applied economists*, Princeton University Press, New Jersey, 1992.

[39] I. Goodfellow, J. P-. Abadie, M. Mirza, B. Xu, D. W-. Farley, S. Ozair, A. Courville and Y. Bengio, Generative adversarial networks, In *Proceedings of Neural Information Processing Systems* (NeurIPS 2014), Montréal, Canada, 2014.

[40] S. Guo and H. Xu, Statistical robustness in utility preference robust optimization models, *Math. Program.*, **190** (2021), 679–720.

[41] G. Gürkan, A.Y. Özge, and S.M. Robinson, Sample-path solution of stochastic variational inequalities, *Math. Program.*, **84** (1999), 313–333.

[42] P.T. Harker, Generalized Nash games and quasi-variational inequalities, *Eur. J. Oper. Res.*, **54** (1991), 81–94.

[43] A. Haurie, G. Zaccour and Y. Smeers, Stochastic equilibrium programming for dynamic oligopolistic markets, *J. Optim. Theory Appl.*, **66** (1990), 243–253.

[44] S. Hayashi, N. Yamashita and M. Fukushima, Robust Nash equilibria and second-order cone complementarity problems, *J. Nonlinear Conv. Anal.*, **6** (2005), 283–296.

[45] X. He, A. Prasad, S.P. Sethi and G.J. Gutierrez, A survey of Stackelberg differential game models in supply and marketing channels, *J. Sys. Sci. Sys. Eng.*, **16** (2007), 385–413.

[46] J. P. Hespanha, *Noncooperative game theory: An introduction for engineers and computer scientists*, Princeton University Press, New Jersey, 2017.

[47] B.F. Hobbs, C. Metzler and J.-S. Pang, Strategic gaming analysis for electric power networks: An MPEC approach, *IEEE Trans. Power Syst.*, **15** (2000), 638–645.

[48] A. Hori and M. Fukushima, Gauss–Seidel method for multi-leader–follower games, *J. Optim. Theory Appl.*, **180** (2019), 651–670.

[49] A. Hori, Y. Yamakawa and N. Yamashita, Distributionally robust expected residual minimization for stochastic variational inequality problem, *Optim. Method Softw.*, to appear.

[50] A. Hori and N. Yamashita, Two-stage distributionally robust noncooperative games: Existence of Nash equilibrium and its application to Cournot–Nash competition, *J. Ind. Manag. Optim.*, to appear.

[51] T. Hospedales, A. Antoniou, P. Micaelli and A. Storkey, Meta-learning in neural networks: A survey, *IEEE Trans. Pattern Anal. Mach. Intell.*, **44** (2022).

[52] M. Hu and M. Fukushima, Variational inequality formulation of a class of multi-leader-follower games, *J. Optim. Theory Appl.*, **151** (2011), 455–473.

[53] M. Hu and M. Fukushima, Smoothing approach to Nash equilibrium formulations for a class of equilibrium problems with shared complementarity constraints, *Compt. Optim. Appl.*, **52** (2012), 415–437.

[54] M. Hu and M. Fukushima, Existence, uniqueness, and computation of robust Nash equilibria in a class of multi-leader–follower games, *SIAM J. Optim.*, **23** (2013), 894–916.

[55] M. Hu and M. Fukushima, Multi-leader-follower games: Models, methods and applications, *J. Oper. Res. Soc. Japan*, **58** (2015), 1–23.

[56] Z. Hu and L. Hong, Kullback–Leibler divergence constrained distributionally robust optimization, preprint Optimization Online, 2012, `https://optimization-online.org/2012/11/3677/`.

[57] X. Hu and D. Ralph, Using EPECs to model bilevel games in restructured electricity markets with locational prices, *Oper. Res.*, **55** (2007), 808–827.

[58] X. Huang, X. Yang and D. Zhu, A sequential smooth penalization approach to mathematical programs with complementarity constraints, *Numer. Func. Anal. Optim.*, **27** (2006), 71–98.

[59] J. Jiang and S. Li, Regularized sample average approximation approach for two-stage stochastic variational inequalities, *J. Optim. Theory Appl.*, **190** (2021), 650–671.

[60] J. Jiang and S. Li, Statistical robustness of two-stage stochastic variational inequalities, *Optim. Lett.*, **16** (2022), 2591–2605.

[61] J. Jiang, Y. Shi, X. Wang and X. Chen, Regularized two-stage stochastic variational inequalities for Cournot–Nash equilibrium under uncertainty, *J. Compt. Math.*, **37** (2019), 813–842.

[62] J. Jiang, H. Sun and B. Zhou, Convergence analysis of sample average approximation for a class of stochastic nonlinear complementarity problems: from two-stage to multistage, *Numer. Algorithms*, **89** (2022), 167–194.

[63] R. Jiang and Y. Guan, Data-driven chance constrained stochastic program, *Math. Program.*, **158** (2016), 291–327.

[64] A. Jofré, R. T. Rockafellar and R. J. Wets., Variational inequalities and economic equilibrium, *Math. Oper. Res.*, **32** (2007), 32–50.

[65] S. Kakutani, A generalization of Brouwer's fixed point theorem, *Duke Math. J.*, **8** (1941), 457–459.

[66] C. Kanzow and D. Steck, Augmented Lagrangian methods for the solution of generalized Nash equilibrium problems, *SIAM J. Optim.*, **26** (2016), 2034–2058.

[67] V. Krätschmer, A. Schied, and H. Zähle, Qualitative and infinitesimal robustness of tail-dependent statistical functionals, *J. Multivariate Anal.*, **103** (2012), 35–47.

[68] L. Lampariello, S. Sagratella and O. Stein, The standard pessimistic bilevel problem, *SIAM J. Optim.*, **29** (2019), 1634–1656.

[69] J. Lei and U. V. Shanbhag, Stochastic Nash equilibrium problems: Models, analysis, and algorithms, *IEEE Control Syst.*, **42** (2022), 103–124.

[70] J. Lei, U. V. Shanbhag, J. S. Pang and S. Sen, On synchronous, asynchronous, and randomized best-response schemes for stochastic Nash games, *Math. Oper. Res.*, **45** (2020), 157–190.

[71] S. Leyffer and T. Munson, Solving multi-leader-common-follower games, *Optim. Method Softw.*, **25** (2010), 601–623.

[72] B. Li, J. Sun, H. Xu and M. Zhang, A class of two-stage distributionally robust games, *J. Ind. Manag. Optim.*, **15** (2019), 387–400.

[73] X. Liang and Y. Xiao, Game theory for network security, in *IEEE Communications Surveys & Tutorials*, **15** (2013), 472–486.

[74] R. Liu, Y. Liu, S. Zeng and J. Zhang, Towards gradient-based bilevel optimization with non-convex followers and beyond, In *the 34th Conference on Neural Information Processing Systems* (NeurIPS 2021), Virtual, 2021.

[75] R. Liu, P. Mu, X. Yuan, S. Zeng and J. Zhang, A generic first-order algorithmic framework for bi-level programming beyond lower-level singleton, In *Proceedings of the 37th International Conference on Machine Learning* (ICML), Vienna, Austria, 2020.

[76] Y. Liu and H. Xu, Entropic approximation for mathematical programs with robust equilibrium constraints, *SIAM J. Optim.*, **24** (2014), 933-958.

[77] Y. Liu, H. Xu, S. J. S. Yang and J. Zhang, Distributionally robust equilibrium for continuous games: Nash and Stackelberg models, *Eur. J. Oper. Res.*, **265** (2018), 631–643.

[78] N. Loizou, *Distributionally robust game theory*, Master's thesis, Imperial College London, London, 2015.

[79] N. Loizou, Distributionally robust games with risk-averse players, *Proceedings of the 5th International Conference on Operations Research and Enterprise Systems (ICORES)*, Rome, Italy, 2016, 186–196.

[80] M.J. Luo and G.H. Lin, Expected residual minimization method for stochastic variational inequality problems, *J. Optim. Theory Appl.*, **140** (2009), 103–106.

[81] M.J. Luo and G.H. Lin, Convergence results of the ERM method for nonlinear stochastic variational inequality problems, *J. Optim. Theory Appl.*, **142** (2009), 569–581.

[82] Z.Q. Luo, J.-S. Pang and D. Ralph, *Mathematical Programs with Equilibrium Constraints*, Cambridge University Press, 1996.

[83] S. Mehrotra and D. Papp, A cutting surface algorithm for semi-infinite convex programming with an application to moment robust optimization, *SIAM J. Optim.*, **24** (2014), 1670–1697.

[84] D. Monderer and L. S. Shapley, Potential games, *Games Econ. Behav.*, **14** (1996), 124–143.

[85] J. Nash, Equilibrium points in n-person games, *Proceedings of Natl. Acad. Sci. USA*, **36** (1950), 48–49.

[86] J. Nash, Non-cooperative games, *Ann. Math.*, **54** (1951), 286–295.

[87] J. von Neumann and O. Morgenstern, *Theory of games and economic behavior*, Princeton University Press, 1944.

[88] R. Nishimura, S. Hayashi and M. Fukushima, Robust Nash equilibria in N-person noncooperative games: Uniqueness and reformulation, *Pacific J. Optim.*, **5** (2009), 237–259.

[89] J. Outrata, A note on a class of equilibrium problems with equilibrium constraints, *Kybernetika*, **40** (2004), 585–594.

[90] J. Outrata, M. Kočvara and J. Zowe, *Nonsmooth Approach to Optimization Problems with Equilibrium Constraints: Theory, Applications and Numerical Results*, Springer, New York, 1998.

[91] J.-S. Pang, A posteriori error bounds for the linearly constrained variational inequality problem, *Math. Oper. Res.*, **12** (1987), 474–484.

[92] J.-S. Pang and M. Fukushima, Quasi-variational inequalities, generalized Nash equilibria and multi-leader-follower games, *Compt. Manag. Sci.*, **2** (2005), 21–56. Erratum. *ibid.*, **6** (2005), 373–375.

[93] J.-S. Pang, S. Sen and U.V. Shanbhag, Two-stage non-cooperative games with risk-averse players, *Math. Program.*, **165** (2017), 235–290.

[94] G. Peng, T. Zhang and Q. Zhu, A data-driven distributionally robust game using Wasserstein distance. In *Decision and Game Theory for Security*, Springer, 2020, 405–421.

[95] I. Pólik and T. Terlaky, A survey of the S-lemma, *SIAM Rev.*, **49** (2007), 371–418.

[96] S. Qu and M. Goh, Distributionally robust games with an application to supply chain, Harbin Institute of Technology, Technical report, 2012.

[97] H. Rahimian and S. Mehrotra, Frameworks and results in distributionally robust optimization, *Open J. Math. Optim.*, 2022, available online.

[98] D. Ralph and Y. Smeers, Risk trading and endogenous probabilities in investment equilibria, *SIAM J. Optim.*, **25** (2015), 2589–2611.

[99] D. Ralph and H. Xu, Convergence of stationary points of sample average two-stage stochastic programs: A generalized equation approach, *Math. Oper. Res.*, **36** (2011), 568–592.

[100] U. Ravat and U. V. Shanbhag, On the characterization of solution sets of smooth and nonsmooth convex stochastic Nash games, *SIAM J. Optim.*, **21** (2011), 1168–1199.

[101] U. Ravat and U. V. Shanbhag. On the existence of solutions to stochastic quasi-variational inequality and complementarity problems, *Math. Program.*, **165** (2017), 291–330.

[102] R. T. Rockafellar, Monotone operators and the proximal point algorithm, *SIAM J. Control Optim.*, **14** (1976), 877–898.

[103] R. T. Rockafellar and J. Sun, Solving monotone stochastic variational inequalities and complementarity problems by progressive hedging, *Math. Program.*, **174** (2019), 453–471.

[104] R. T. Rockafellar and S. Uryasev, Optimization of conditional value-at-risk, *J. Risk*, **2** (2000), 21–41.

[105] R. T. Rockafellar and R. J. Wets, Scenarios and policy aggregation in optimization under uncertainty, *Math. Oper. Res.*, **16** (1991), 119–147.

[106] R. T. Rockafellar and R. J. Wets, *Variational Analysis*, Springer, Berlin, 1998.

[107] R. T. Rockafellar and R. J. Wets, Stochastic variational inequalities: single-stage to multistage, *Math. Program.*, **165** (2017), 331–360.

[108] T. Roughgarden, Algorithmic game theory, *Commun. ACM*, **53** (2010), 78–86.

[109] J. B. Rosen, Existence and uniqueness of equilibrium points for concave N-person games, *Econometrica*, **33** (1965), 520–534.

[110] H. Scheel and S. Scholtes, Mathematical programs with complementarity constraints: Stationarity, optimality, and sensitivity, *Math. Oper. Res.*, **25** (2000), 1–22.

[111] A. Shapiro, First and second order analysis of nonlinear semidefinite programs, *Math. Program.* **77** (1997), 301–320.

[112] A. Shapiro and A. Nemirovski, On complexity of stochastic programming problems, In V. Jeyakumar and A. Rubinov (eds), *Continuous Optimization: Current Trends and Modern Applications*, Springer, Boston, MA, 2005, 111–146.

[113] H. D. Sherali, A. L. Soyster and F. H. Murphy, Stackelberg-Nash-Cournot equilibria: characterizations and computations, *Oper. Res.*, **31** (1983), 253–276.

[114] V. V. Singh and A. Lisser, Variational inequality formulation for the games with random payoffs, *J. Glob. Optim.*, **72** (2018), 743–760.

[115] H. von Stackelberg, Marktform und Gleichgewicht, Springer Verlag, Berlin, 1934.

[116] J.F. Sturm and S. Zhang, On cones of nonnegative quadratic functions, *Math. Oper. Res.*, **28** (2003), 246–267.

[117] C.L. Su, A sequential NCP algorithm for solving equilibrium problems with equilibrium constraints, Technical report, Stanford University, Stanford, US, 2004.

[118] H. Sun, A. Shapiro and X. Chen, Distributionally robust stochastic variational inequalities, *Math. Program.* (2022).

[119] H. Sun and H. Xu, Convergence analysis for distributionally robust optimization and equilibrium problems, *Math. Oper. Res.*, **41** (2016), 377–401.

[120] H. Sun and X. Chen, Two-stage stochastic variational inequalities: Theory, algorithms and applications, *J. Oper. Res. Soc. China* (2019).

[121] D. M. Topkis, Equilibrium points in nonzero-sum n-person submodular games, *SIAM J. Control Optim.*, **17** (1979), 773–787.

[122] W. Wiesemann, D. Kuhn and M. Sim, Distributionally robust convex optimization, *Oper. Res.*, **62** (2014), 1358–1376.

[123] H. Xu, An MPCC approach for stochastic Stackelberg–Nash–Cournot equilibrium, *Optimization*, **54** (2005), 27–57.

[124] H. Xu, Y. Liu, and H. Sun, Distributionally robust optimization with matrix moment constraints: Lagrange duality and cutting plane methods, *Math. Program.*, **169** (2018), 489–529.

[125] H. Yamashita and H. Yabe, A primal-dual interior point method for nonlinear optimization over second-order cones, *Optim. Method Softw.*, **24** (2009), 407–426.

[126] H. Yamashita, H. Yabe, and K. Harada, A primal-dual interior point method for nonlinear semidefinite programming, *Math. Program.*, **135** (2012), 89–121.

[127] N. Yamashita, M. Fukushima, Equivalent unconstrained minimization and global error bounds for variational inequality problems, *SIAM J. Control Optim.*, **35** (1997), 273–284.

[128] N. Yamashita, K. Taji and M. Fukushima, Unconstrained optimization reformulations of variational inequality problems, *J. Optim. Theory Appl.*, **92** (1997), 439–456.

[129] Y. Yang and J. Wang, An overview of multi-agent reinforcement learning from game theoretical perspective, arXiv: 2011.00583, 2020.

[130] D. Zhang, H. Xu and Y. Wu, A two stage stochastic equilibrium model for electricity markets with two way contracts, *Math. Methods Oper. Res.*, **71** (2010), 1–45.

[131] M. Zhang, J. Sun and H. Xu, Two-stage quadratic games under uncertainty and their solution by progressive hedging algorithms, *SIAM J. Optim.*, **29** (2019), 1799–1818.

[132] B. Zhou and H. Sun, Two-stage stochastic variational inequalities for Cournot–Nash equilibrium with risk-averse players under uncertainty, *Numer. Algebra, Control. Optim.*, **10** (2020), 521–535.

[133] L. Zhu, B. Yu, and L. Xu, The distributionally robust complementarity problem, *Optim. Method Softw.*, **32** (2017), 650–668.

Appendix A

Details of the proof of Lemma 3.3.3 and Theorem 3.3.4–(b)

By the continuity of $\nabla_z f, \nabla_z^2 f$, etc., we can prove the lemma and theorem by appropriate modifications of the arguments in Lemma 4.3 and Theorem 4.4 in Huang et al. [58].

The difference between our arguments and those of [58] is that our problem is parametric while the problem in [58] is parameter-free. However, since our problem functions along with their first- and second-derivatives are continuous and the sequence $\{a^k\}$ of parameters is convergent, the proof technique in [58] can be adapted to our theorem in a straightforward manner. Moreover, we have checked the proofs in [58] carefully and found no errors apart from a few typos.

However, a complete proof of our theorem would be lengthy. In fact, Huang et al. [58] spent 8 pages to prove their Lemma 4.3 and Theorem 4.4. So we will give a brief sketch of the proof for Lemma 3.3.3 and Theorem 3.3.4 below.

Proof of Lemma 3.3.3

Proof. First, the feasibility at $\bar{z}$ is ensured by Lemma 3.3.1. Next, we show $\bar{z}$ satisfies weak stationarity. If we set

$$\xi_j^k = -\rho_k \phi(G_j(z^k, a^k), H_j(z^k, a^k)) \left(1 - \frac{G_j(z^k, a^k)}{\sqrt{G_j(z^k, a^k)^2 + H_j(z^k, a^k)^2}}\right), \ j \in \mathcal{J}'(z^k, a^k), \tag{A.1}$$

$$\eta_j^k = -\rho_k \phi(G_j(z^k, a^k), H_j(z^k, a^k)) \left(1 - \frac{H_j(z^k, a^k)}{\sqrt{G_j(z^k, a^k)^2 + H_j(z^k, a^k)^2}}\right), \ j \in \mathcal{J}'(z^k, a^k), \tag{A.2}$$

$$\xi_j^k = \eta_j^k = 0, \quad j \in \mathcal{J}_{0+}(z^k, a^k) \cup \mathcal{J}_{+0}(z^k, a^k) \cup \mathcal{J}_{00}(z^k, a^k),$$
$$\lambda_i^k = \rho_k g_i(z^k, a^k), \quad i \in \mathcal{I}_g^+(z^k, a^k),$$
$$\lambda_i^k = 0, \quad i \in \{1, 2, \ldots, r\} \setminus \mathcal{I}_g^+(z^k, a^k),$$
$$\mu_i^k = \rho_k h_i(z^k, a^k), \quad i \in \mathcal{I}_h^C(z^k, a^k),$$
$$\mu_i^k = 0, \quad i \in \mathcal{I}_h(z^k, a^k),$$

then the first-order condition in Lemma 3.3.2 yields

$$\nabla f_z(z^k,a^k)+\sum_{i=1}^{r}\lambda_i^k\nabla_z g_i(z^k,a^k)+\sum_{i=1}^{s}\mu_i^k\nabla_z h_i(z^k,a^k)$$
$$-\sum_{j=1}^{m}\xi_j^k\nabla_z G_j(z^k,a^k)-\sum_{j=1}^{m}\eta_j^k\nabla_z H_j(z^k,a^k)=0$$

and

$$\lambda_i^k\geq 0,\quad i=1,2,\ldots,r.$$

In a similar manner to the proof of Lemma 4.3 in [58], we can show that the sequences $\{\xi_j^k\}_{j=1}^m$, $\{\eta_j^k\}_{j=1}^m$, $\{\lambda_i^k\}_{i=1}^r$ and $\{\mu_i^k\}_{i=1}^s$ are bounded under the MPEC-LICQ condition, and so we may assume without loss of generality that those sequences converge to $\bar{\xi}_j$ $(j=1,\ldots,m)$, $\bar{\eta}_j$ $(j=1,\ldots,m)$, $\bar{\lambda}_i$ $(i=1,\ldots,r)$ and $\bar{\mu}_i$ $(i=1,\ldots,s)$, respectively. Moreover, we can show that

$$\begin{aligned}\bar{\xi}_j&=0,\quad j\in\mathcal{J}_{+0}(\bar{z},\bar{a}),\\ \bar{\eta}_j&=0,\quad j\in\mathcal{J}_{0+}(\bar{z},\bar{a}),\\ \bar{\lambda}_i&=0,\quad i\in\mathcal{I}_g^+(\bar{z},\bar{a}).\end{aligned}$$

Consequently, we obtain

$$\nabla f_z(\bar{z},\bar{a})+\sum_{i\in\mathcal{I}_g(\bar{z},\bar{a})}\bar{\lambda}_i\nabla_z g_i(\bar{z},\bar{a})+\sum_{i=1}^{s}\bar{\mu}_i\nabla_z h_i(\bar{z},\bar{a})$$
$$-\sum_{j\in\mathcal{J}_{0+}(\bar{z},\bar{a})\cup\mathcal{J}_{00}(\bar{z},\bar{a})}\bar{\xi}_j\nabla_z G_j(\bar{z},\bar{a})-\sum_{j\in\mathcal{J}_{+0}(\bar{z},\bar{a})\cup\mathcal{J}_{00}(\bar{z},\bar{a})}\bar{\eta}_j\nabla_z H_j(\bar{z},\bar{a})=0.$$

This completes the proof of Lemma 3.3.3. □

Proof of Theorem 3.3.4–(b) Indeed, the proof technique used in [58] is an adaptation of that used in Fukushima and Pang [37].

Proof. First, we can assume without loss of generality that

$$j^*\in\mathcal{J}'(z^k,a^k)\quad\text{for all }k.\tag{A.3}$$

Moreover, by the definitions of (A.1), (A.2), and the fact that $\xi_{j^*}^k\to\bar{\xi}_{j^*}(<0)$ and $\eta_{j^*}^k\to\bar{\eta}_{j^*}(<0)$ from Lemma 3.3.3, we have

$$\phi(G_{j^*}(z^k,a^k),H_{j^*}(z^k,a^k))>0$$

and hence

$$G_{j^*}(z^k,a^k)>0,\ H_{j^*}(z^k,a^k)>0$$

for all k sufficiently large. In addition, $\{\rho_k\phi(G_{j^*}(z^k,a^k),H_{j^*}(z^k,a^k))\}$ is shown to be bounded. So we may assume without loss of generality that $\{\rho_k\phi(G_{j^*}(z^k,a^k),H_{j^*}(z^k,a^k))\}$ converges, and let

$$\theta^*=\lim_{k\to\infty}\rho_k\phi(G_{j^*}(z^k,a^k),H_{j^*}(z^k,a^k)).$$

On the other hand, we define

$$a^k_{j^*} = 1 - \frac{G_{j^*}(z^k, a^k)}{\sqrt{G_{j^*}(z^k, a^k)^2 + H_{j^*}(z^k, a^k)^2}}, \quad b^k_{j^*} = 1 - \frac{H_{j^*}(z^k, a^k)}{\sqrt{G_{j^*}(z^k, a^k)^2 + H_{j^*}(z^k, a^k)^2}},$$

and assume without loss of generality that $\lim_{k\to\infty} a^k_{j^*} = a^*$ and $\lim_{k\to\infty} b^k_{j^*} = b^*$. Then, by the definitions of ξ^k_j, $a^k_{j^*}$ and θ^*, we have

$$0 > \bar{\xi}_{j^*} = -\theta^* a^*.$$

Since $a^k_{j^*} \geq 0$ for all k, we obtain $\theta^* > 0$ as well as $a^* > 0$. However, this leads to a contradiction as sketched below, and so $\bar{z}$ must be an M-stationary point.

In fact, the contradiction is derived by considering a bounded sequence of vectors $\{d^k\}$ satisfying the system

$$\begin{aligned}
&\nabla G_{j^*}(z^k, a^k)^\top d^k = 1 - \frac{H_{j^*}(z^k, a^k)}{\sqrt{G_{j^*}(z^k, a^k)^2 + H_{j^*}(z^k, a^k)^2}},\\
&\nabla H_{j^*}(z^k, a^k)^\top d^k = -\left(1 - \frac{G_{j^*}(z^k, a^k)}{\sqrt{G_{j^*}(z^k, a^k)^2 + H_{j^*}(z^k, a^k)^2}}\right),\\
&\nabla G_j(z^k, a^k)^\top d^k = \nabla H_j(z^k, a^k)^\top d^k = 0, \quad j \in \mathcal{J}'(z^k, a^k) \setminus \{j^*\},\\
&\nabla G_j(z^k, a^k)^\top d^k = 0, \quad j \in \mathcal{J}_{0+}(z^k, a^k) \cup \mathcal{J}_{00}(z^k, a^k),\\
&\nabla H_j(z^k, a^k)^\top d^k = 0, \quad j \in \mathcal{J}_{+0}(z^k, a^k) \cup \mathcal{J}_{00}(z^k, a^k),\\
&\nabla g_i(z^k, a^k)^\top d^k = 0, \quad i \in \mathcal{I}^+_g(z^k, a^k),\\
&\nabla h_i(z^k, a^k)^\top d^k = 0, \quad i \in \mathcal{I}^C_h(z^k, a^k).
\end{aligned}$$

(Recall $j^* \in \mathcal{J}'(z^k, a^k)$; see (A.3)). By the MPEC-LICQ at $(\bar{z}, \bar{a})$, there exists d^k satisfying the above system for all k sufficiently large. However, we can show that the second-order optimality condition shown in Lemma 3.3.2 does not hold for such d^k when k becomes large, since $\theta^* > 0$ ensures that the left-hand side of (3.3) tends to $-\infty$ as $k \to \infty$. This contradicts the assumption that weak second-order condition (3.3) holds at $\bar{z}$. Consequently, $\bar{z}$ is an M-stationary point of PMPCC($\bar{a}$). $\square$

Appendix B

Detailed results of numerical experiments in Section 3.6

In this appendix, we show that the detail of the B-stationarity of each leader's MPCC in Examples 3.6.1 to 3.6.7 shown in Section 3.6.

Example 3.6.1 Here we discuss the detail of the results in Example 3.6.1. We checked each player's optimality conditions of MPCC. The active constraints of the complementarity conditions at the solution $(x^{1,*}, x^{2,*}, y^{1,*}, y^{2,*})$ to Example 3.6.1 are $(My+N_1x^1+N_2x^2+q)_i \geq 0$ $(i = 1, 2)$, hence we have to check the KKT conditions for the following each player's optimization problem:

$$\begin{array}{ll} \underset{x^\nu \in I\!R^2, y \in I\!R^2}{\text{minimize}} & \theta^\nu(x^\nu, x^{-\nu}, y) \\ \text{subject to} & A_\nu x^\nu \leq b^\nu, \\ & My + N_1 x^1 + N_2 x^2 + q = 0, \\ & y \geq 0. \end{array} \tag{B.1}$$

For player 1:

$$\begin{aligned}
&H_1 x^{1,*} + G_1 x^{-1,*} + A_1^\top \lambda^{1,*} + N_1^\top \xi^{1,*} = \begin{bmatrix} 0.1332\text{e}-14 \\ 0.0444\text{e}-14 \end{bmatrix}, \\
&c^1 + M^\top \xi^{1,*} - \eta^{1,*} = \begin{bmatrix} -0.1033\text{e}-15 \\ 0.4441\text{e}-15 \end{bmatrix}, \\
&\text{with} \lambda^{1,*} = 0.5972,\ \xi^{1,*} = \begin{bmatrix} 1.3575 \\ 1.7246 \end{bmatrix},\ \eta^{1,*} = \begin{bmatrix} 0.5871\text{e}-05 \\ 0 \end{bmatrix}, \\
&A_1 x^{1,*} - b^1 = -3.3493\text{e}-06,\ My^{1,*} + N_1 x^{1,*} + N_2 x^{2,*} + q = \begin{bmatrix} -0.1148\text{e}-12 \\ -0.1115\text{e}-12 \end{bmatrix}.
\end{aligned}$$

For player 2:

$$H_2 x^{2,*} + G_2 x^{-2,*} + A_2^\top \lambda^{2,*} + N_2^\top \xi^{2,*} = \begin{bmatrix} 0 \\ 0.8882\text{e}-15 \end{bmatrix},$$
$$c^2 + M^\top \xi^{2,*} - \eta^{2,*} = \begin{bmatrix} -0.0888\text{e}-14 \\ 0.1501\text{e}-14 \end{bmatrix},$$
$$\text{with } \lambda^{2,*} = 3.0775,\ \xi^{2,*} = \begin{bmatrix} 1.2802 \\ 1.4058 \end{bmatrix},\ \eta^{2,*} = \begin{bmatrix} 0 \\ 0.1956\text{e}-05 \end{bmatrix},$$
$$A_2 x^{2,*} - b^2 = -3.2132\text{e}-06,\ My^{2,*} + N_1 x^{1,*} + N_2 x^{2,*} + q = \begin{bmatrix} 0.1875\text{e}-07 \\ -0.1423\text{e}-07 \end{bmatrix},$$

where $\lambda^{\nu,*}, \xi^{\nu,*}, \eta^{\nu,*}$ are the Lagrange multipliers for the constraints $A_\nu x^\nu \le b^\nu, My + N_1 x^1 + N_2 x^2 + q = 0, y \ge 0$, respectively. From the above considerations, we confirmed that the algorithm successfully obtained a B-stationary equilibrium point.

Example 3.6.2 Here we discuss the detail of the results in Example 3.6.2. The active constraints of the complementarity constraints are the same as Example 3.6.1, hence we have to check the KKT conditions for the problem (B.1). For player 1:

$$H_1 x^{1,*} + G_1 x^{-1,*} + A_1^\top \lambda^{1,*} + N_1^\top \xi^{1,*} = \begin{bmatrix} -0.0444\text{e}-14 \\ 0.4940\text{e}-14 \\ -0.2887\text{e}-14 \end{bmatrix},$$
$$c^1 + M^\top \xi^{1,*} - \eta^{1,*} = \begin{bmatrix} -0.1332\text{e}-14 \\ 0.0888\text{e}-14 \\ -0.1450\text{e}-14 \end{bmatrix},$$
$$\text{with } \lambda^{1,*} = \begin{bmatrix} 0.8540 \\ 4.3951 \end{bmatrix},\ \xi^{1,*} = \begin{bmatrix} 0.7138 \\ 0.2580 \\ 0.6795 \end{bmatrix}, \eta^{1,*} = \begin{bmatrix} 0 \\ 0 \\ -0.1831\text{e}-06 \end{bmatrix},$$
$$A_1 x^{1,*} - b^1 = \begin{bmatrix} -0.1171\text{e}-06 \\ -0.0228\text{e}-06 \end{bmatrix},\ My^{1,*} + N_1 x^{1,*} + N_2 x^{2,*} + q = \begin{bmatrix} 0.2471\text{e}-08 \\ 0.1780\text{e}-08 \\ -0.3818\text{e}-08 \end{bmatrix}.$$

For player 2:

$$H_2 x^{2,*} + G_2 x^{-2,*} + A_2^\top \lambda^{2,*} + N_2^\top \xi^{2,*} = \begin{bmatrix} 0.1998\text{e}-14 \\ -0.2276\text{e}-14 \\ 0.0722\text{e}-14 \end{bmatrix},$$
$$c^2 + M^\top \xi^{2,*} - \eta^{2,*} = \begin{bmatrix} -0.0513\text{e}-14 \\ 0.1332\text{e}-14 \\ -0.0888\text{e}-14 \end{bmatrix},$$
$$\text{with } \lambda^{2,*} = \begin{bmatrix} 3.0130 \\ 1.1435 \end{bmatrix},\ \xi^{2,*} = \begin{bmatrix} 0.6488 \\ 0.2213 \\ 0.6341 \end{bmatrix}, \eta^{2,*} = \begin{bmatrix} 0.1123\text{e}-07 \\ 0 \\ 0 \end{bmatrix},$$
$$A_2 x^{2,*} - b^2 = \begin{bmatrix} -0.3319\text{e}-07 \\ -0.8745\text{e}-07 \end{bmatrix},\ My^{2,*} + N_1 x^{1,*} + N_2 x^{2,*} + q = \begin{bmatrix} -0.1688\text{e}-13 \\ 0.0266\text{e}-13 \\ 0.0977\text{e}-13 \end{bmatrix}.$$

Example 3.6.3 Here we discuss the detail of the results in Example 3.6.3. The active constraints of the complementarity constraints for the example are $(My+N_1x^1+N_2x^2+q)_{\{1,3\}} \geq 0, y_2 \geq 0$. We checked the KKT conditions for the following each player's optimization problem:

$$\begin{array}{ll} \text{minimize} & \theta^\nu(x^\nu, x^{-\nu}, y) \\ \text{subject to} & A_\nu x^\nu \leq b^\nu, \\ & (My + N_1x^1 + N_2x^2 + q)_1 = 0,\ y_1 \geq 0, \\ & (My + N_1x^1 + N_2x^2 + q)_2 \geq 0,\ y_2 = 0, \\ & (My + N_1x^1 + N_2x^2 + q)_3 = 0,\ y_3 \geq 0. \end{array}$$

For player 1:

$$H_1x^{1,*} + G_1x^{-1,*} + A_1^\top\lambda^{1,*} + N_1^\top\xi^{1,*} = \begin{bmatrix} 0.6292\mathrm{e}-06 \\ 0.5938\mathrm{e}-06 \\ 0.1957\mathrm{e}-06 \end{bmatrix},$$

$$c^1 + M^\top\xi^{1,*} - \eta^{1,*} = \begin{bmatrix} 0.1295\mathrm{e}-06 \\ -0.0000\mathrm{e}-06 \\ 0.1343\mathrm{e}-06 \end{bmatrix},$$

$$\text{with } \lambda^{1,*} = \begin{bmatrix} 0.8877 \\ 4.5799 \end{bmatrix},\ \xi^{1,*} = \begin{bmatrix} 0.8220 \\ -0.0000 \\ 0.5573 \end{bmatrix}, \eta^{1,*} = \begin{bmatrix} 0.0000 \\ 3.1068 \\ 0.0000 \end{bmatrix},$$

$$A_1x^{1,*} - b^1 = \begin{bmatrix} -0.2255\mathrm{e}-05 \\ -0.0437\mathrm{e}-05 \end{bmatrix},\ My^{1,*} + N_1x^{1,*} + N_2x^{2,*} + q = \begin{bmatrix} 0.0000 \\ 0.1429 \\ -0.0000 \end{bmatrix}.$$

For player 2:

$$H_2x^{2,*} + G_2x^{-2,*} + A_2^\top\lambda^{2,*} + N_2^\top\xi^{2,*} = \begin{bmatrix} 0.0991\mathrm{e}-07 \\ -0.0137\mathrm{e}-07 \\ 0.1245\mathrm{e}-07 \end{bmatrix},$$

$$c^2 + M^\top\xi^{2,*} - \eta^{2,*} = \begin{bmatrix} 0.4880\mathrm{e}-08 \\ 0.0000\mathrm{e}-08 \\ -0.3825\mathrm{e}-08 \end{bmatrix},$$

$$\text{with } \lambda^{2,*} = \begin{bmatrix} 2.5454 \\ 1.1247 \end{bmatrix},\ \xi^{2,*} = \begin{bmatrix} 0.7416 \\ 0.0000 \\ 0.5293 \end{bmatrix}, \eta^{2,*} = \begin{bmatrix} -0.0000 \\ 2.8334 \\ 0.0000 \end{bmatrix},$$

$$A_2x^{2,*} - b^2 = \begin{bmatrix} -0.3929\mathrm{e}-07 \\ -0.8891\mathrm{e}-07 \end{bmatrix},\ My^{2,*} + N_1x^{1,*} + N_2x^{2,*} + q = \begin{bmatrix} -0.0000 \\ 0.1429 \\ 0.0000 \end{bmatrix}.$$

Example 3.6.5 Here we discuss the detail of the results in Example 3.6.5. The active constraint of the complementarity constraints is $c^\top y + (d^1)^\top x^1 + (d^2)^\top x^2 + a \geq 0$. We checked the KKT conditions for the following each leader's optimization problem:

$$\begin{array}{ll} \underset{x^\nu \in I\!R^2, y \in I\!R^3, \lambda \in I\!R}{\text{minimize}} & \theta^\nu(x^\nu, x^{-\nu}, y) \\ \text{subject to} & A_\nu x^\nu \leq b^\nu, \\ & My + q - D_1^\top x^1 - D_2^\top x^2 - c\lambda = 0, \\ & c^\top y + (d^1)^\top x^1 + (d^2)^\top x^2 + a = 0, \\ & \lambda \geq 0. \end{array}$$

For player 1:

$$H_1 x^{1,*} + G_{1,2} x^{-1,*} + D_1 y^{2,*} + A_1^\top \alpha^{1,*} - D_1 \gamma^{1,*} + d^1 \eta^{1,*} = \begin{bmatrix} -0.5574\text{e}-05 \\ -0.6135\text{e}-05 \end{bmatrix},$$

$$D_1^\top x^{1,*} + M^\top \gamma^{1,*} + c\,\eta^{1,*} = \begin{bmatrix} 0.0156\text{e}-06 \\ 0.1005\text{e}-06 \\ 0.0486\text{e}-06 \end{bmatrix},$$

$$-c^\top \gamma^{1,*} - \xi^{1,*} = 6.9036\text{e}-08,$$

$$\text{with } \alpha^{1,*} = \begin{bmatrix} 0.0000 \\ 0.6532 \\ 0.0000 \end{bmatrix}, \ \gamma^{1,*} = \begin{bmatrix} -0.1189 \\ 0.2373 \\ -0.1346 \end{bmatrix}, \ \xi^{1,*} = -2.7155e-09, \ \eta^{1,*} = 0.1470,$$

$$A_1 x^{1,*} - b^1 = \begin{bmatrix} -1.4936 \\ -0.0000 \\ -1.8196 \end{bmatrix}, c^\top y^{1,*} + (d_1)^\top x^{1,*} + (d_2)^\top x^{2,*} + a = -2.2707\text{e}-08,$$

For player 2:

$$H_2 x^{2,*} + G_{2,1} x^{-2,*} + D_2 y^{2,*} + A_2^\top \alpha^{2,*} - D_2 \gamma^{2,*} + d^2 \eta^{2,*} = \begin{bmatrix} -0.1567\text{e}-07 \\ 0.2881\text{e}-07 \end{bmatrix},$$

$$D_2^\top x^{2,*} + M^\top \gamma^{2,*} + c\,\eta^{2,*} = \begin{bmatrix} 0.1075\text{e}-07 \\ 0.0136\text{e}-07 \\ -0.1590\text{e}-07 \end{bmatrix},$$

$$-c^\top \gamma^{2,*} - \xi^{2,*} = -2.7892\text{e}-10,$$

$$\text{with } \alpha^{2,*} = \begin{bmatrix} 0.0000 \\ 0.1061 \\ 0.0000 \end{bmatrix}, \ \gamma^{2,*} = \begin{bmatrix} -0.3527 \\ 0.1367 \\ -0.2102 \end{bmatrix} \ \xi^{2,*} = -1.3078e-10, \ \eta^{2,*} = 0.2615$$

$$A_2 x^{2,*} - b^2 = \begin{bmatrix} -0.4273 \\ -0.0000 \\ -2.3430 \end{bmatrix}, c^\top y^{2,*} + (d_1)^\top x^{1,*} + (d_2)^\top x^{2,*} + a = -4.9960\text{e}-14,$$

where $\alpha^{\nu,*}, \gamma^{\nu,*}, \xi^{\nu,*}, \eta^{\nu,*}$ are Lagrange multipliers for the constraints $A_\nu x^\nu \leq b^\nu, My + q - D_1^\top x^1 - D_2^\top x^2 - c\,\lambda = 0, 0 \leq \lambda \perp c^\top y + (d^1)^\top x^1 + (d^2)^\top x^2 + a \geq 0$, respectively.

Example 3.6.6 Here we discuss the detail of the results in Example 3.6.5. The active constraints of the complementarity constraints for the example is $\lambda \geq 0$. We checked the

KKT conditions for the following each leader's optimization problem:

$$\begin{array}{ll} \text{minimize} & \theta^\nu(x^\nu, x^{-\nu}, y) \\ \text{subject to} & A_\nu x^\nu \leq b^\nu, \\ & \lambda = 0, \\ & c^\top y + (d^1)^\top x^1 + (d^2)^\top x^2 + (d^3)^\top x^3 + a \geq 0. \end{array}$$

For leader 1:

$$H_1 x^{1,*} + G_{1,2} x^{2,*} + G_{2,3} x^{3,*} + D_1 y^{1,*} + A_1^\top \alpha^{1,*} - D_1 \gamma^{1,*} - d^1 \eta^{1,*} = \begin{bmatrix} -0.2501\text{e}-07 \\ 0.3491\text{e}-07 \\ -0.2658\text{e}-07 \end{bmatrix},$$

$$D_1^\top x^{1,*} + M^\top \gamma^{1,*} - c\,\eta^{1,*} = \begin{bmatrix} -0.1230\text{e}-08 \\ -0.9576\text{e}-08 \\ -0.4648\text{e}-08 \end{bmatrix},$$

$$-c^\top \gamma^{1,*} + \xi^{1,*} = 0,$$

$$\text{with } \alpha^{1,*} = \begin{bmatrix} 0.3022\text{e}-08 \\ -0.4926\text{e}-08 \\ -0.7129\text{e}-08 \end{bmatrix},\ \gamma^{1,*} = \begin{bmatrix} 0.0493 \\ -0.2145 \\ 0.0476 \end{bmatrix},\ \xi^{1,*} = -0.0718,\ \eta^{1,*} = 5.1410\text{e}-09,$$

$$A_1 x^{1,*} - b^1 = \begin{bmatrix} -2.2634 \\ -1.4461 \\ -2.8254 \end{bmatrix},\ c^\top y^{1,*} + (d_1)^\top x^{1,*} + (d_2)^\top x^{2,*} + (d^3)^\top x^{3,*} + a = 1.2594.$$

For leader 2:

$$H_2 x^{2,*} + G_{2,1} x^{1,*} + G_{2,3} x^{3,*} + D_2 y^{2,*} + A_2^\top \alpha^{2,*} - D_2 \gamma^{2,*} - d^2 \eta^{2,*} = \begin{bmatrix} 0.2420\text{e}-06 \\ 0.0352\text{e}-06 \\ 0.0760\text{e}-06 \end{bmatrix},$$

$$D_2^\top x^{2,*} + M^\top \gamma^{2,*} - c\,\eta^{2,*} = \begin{bmatrix} 0.2149\text{e}-06 \\ -0.0708\text{e}-06 \\ 0.1586\text{e}-06 \end{bmatrix},$$

$$-c^\top \gamma^{2,*} + \xi^{2,*} = -5.5511\text{e}-17,$$

$$\text{with } \alpha^{2,*} = \begin{bmatrix} -0.0083\text{e}-05 \\ 0.6181\text{e}-05 \\ -0.0148\text{e}-05 \end{bmatrix},\ \gamma^{2,*} = \begin{bmatrix} -0.0363 \\ -0.0752 \\ -0.0847 \end{bmatrix},\ \xi^{2,*} = -0.3893,\ \eta^{2,*} = 5.4398\text{e}-07,$$

$$A_2 x^{2,*} - b^2 = \begin{bmatrix} -1.2185 \\ -0.0610 \\ -1.9965 \end{bmatrix},\ c^\top y^{2,*} + (d_1)^\top x^{1,*} + (d_2)^\top x^{2,*} + (d^3)^\top x^{3,*} + a = 1.2594.$$

For leader 3:

$$H_3x^{3,*} + G_{3,1}x^{1,*} + G_{3,2}x^{2,*} + D_3y^{3,*} + A_3^\top\alpha^{3,*} - D_3\gamma^{3,*} - d^3\eta^{3,*} = \begin{bmatrix} -0.3040\text{e}-06 \\ 0.5118\text{e}-06 \\ -0.3482\text{e}-06 \end{bmatrix},$$

$$D_3^\top x^{3,*} + M^\top\gamma^{3,*} - c\,\eta^{3,*} = \begin{bmatrix} -0.0528\text{e}-06 \\ -0.3414\text{e}-06 \\ 0.1187\text{e}-06 \end{bmatrix},$$

$$-\,c^\top\gamma^{3,*} + \xi^{3,*} = 0,$$

with $\alpha^{3,*} = \begin{bmatrix} -0.2038\text{e}-06 \\ -0.0584\text{e}-06 \\ 0.2004\text{e}-06 \end{bmatrix}$, $\gamma^{3,*} = \begin{bmatrix} -0.0414 \\ -0.0389 \\ -0.0951 \end{bmatrix}$, $\xi^{3,*} = -0.3790,\ \eta^{3,*} = 9.3602\text{e}-08,$

$$A_3x^{3,*} - b^3 = \begin{bmatrix} -1.6571 \\ -2.8412 \\ -1.6535 \end{bmatrix},\ c^\top y^{3,*} + (d_1)^\top x^{1,*} + (d_2)^\top x^{2,*} + (d^3)^\top x^{3,*} + a = 1.2594.$$

Example 3.6.7 Here we discuss the detail of the results in Example 3.6.6. First, we write each firm ν's MPCC as follows.

$$\begin{array}{ll} \underset{x^\nu\in I\!R^2,(y,\lambda)\in I\!R^6}{\text{minimize}} & \frac{1}{2}(x^\nu)^\top\text{diag}(\tau_1^\nu,\tau_2^\nu)x^\nu - (x^\nu)^\top D_\nu y \\ \text{subject to} & -x^\nu \leq 0,\ x^\nu - \sigma^\nu \leq 0, \\ & 0 \leq \begin{bmatrix} By + c + D_{\text{I}}^\top x^{\text{I}} + D_{\text{II}}^\top x^{\text{II}} + A^\top\lambda \\ a - Ay \end{bmatrix} \perp \begin{bmatrix} y \\ \lambda \end{bmatrix} \geq 0. \end{array}$$

where

$$D_{\text{I}} := \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix},\ D_{\text{II}} := \begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix},$$

$$B := \begin{bmatrix} \beta_1 + \frac{\zeta_1}{(a^{\text{I}})^2} & 0 & \beta_1 - \frac{\zeta_1}{a^{\text{I}}a^{\text{II}}} & 0 \\ 0 & \beta_2 + \frac{\zeta_2}{(a^{\text{I}})^2} & 0 & \beta_2 - \frac{\zeta_2}{a^{\text{I}}a^{\text{II}}} \\ \beta_1 - \frac{\zeta_1}{a^{\text{I}}a^{\text{II}}} & 0 & \beta_1 + \frac{\zeta_1}{(a^{\text{II}})^2} & 0 \\ 0 & \beta_2 - \frac{\zeta_2}{a^{\text{I}}a^{\text{II}}} & 0 & \beta_2 + \frac{\zeta_2}{(a^{\text{II}})^2} \end{bmatrix},\ c := \begin{bmatrix} -\alpha_1 \\ -\alpha_2 \\ -\alpha_1 \\ -\alpha_2 \end{bmatrix},$$

$$A := \begin{bmatrix} 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \end{bmatrix},\ a := \begin{bmatrix} a^{\text{I}} \\ a^{\text{II}} \end{bmatrix}.$$

At the solution we obtained by the algorithms, the active constraints of the complementarity constraints are $(By + c + D_{\text{I}}^\top x^{\text{I}} + D_{\text{II}}^\top x^{\text{II}} + A^\top\lambda)_i \geq 0$ $(i = 1,\dots,4)$ and $(a - Ay)_i \geq 0$ $(i = 1,2)$, hence we checked the KKT conditions for the following each firm's problem:

$$\begin{array}{ll} \underset{x^\nu\in I\!R^2,(y,\lambda)\in I\!R^6}{\text{minimize}} & \frac{1}{2}(x^\nu)^\top\text{diag}(\tau_1^\nu,\tau_2^\nu)x^\nu - (x^\nu)^\top D_\nu y \\ \text{subject to} & -x^\nu \leq 0,\ x^\nu - \sigma^\nu \leq 0, \\ & By + c + D_{\text{I}}^\top x^{\text{I}} + D_{\text{II}}^\top x^{\text{II}} + A^\top\lambda = 0, \\ & a - Ay = 0, \\ & y \geq 0,\ \lambda \geq 0. \end{array}$$

For firm I:

$$\operatorname{diag}(\tau_1^{\mathrm{I}},\tau_2^{\mathrm{I}})x^{\mathrm{I},*} - D_{\mathrm{I}}y^{\mathrm{I},*} - \chi_l^{\mathrm{I},*} + \chi_u^{\mathrm{I},*} + D_{\mathrm{I}}\xi^{\mathrm{I},*} = \begin{bmatrix} -0.1722\mathrm{e}-07 \\ -0.2299\mathrm{e}-07 \end{bmatrix},$$

$$-D_{\mathrm{I}}^{\top}x^{\mathrm{I},*} + B\xi^{\mathrm{I},*} - A^{\top}\eta^{\mathrm{I},*} - \kappa_y^{\mathrm{I},*} = \begin{bmatrix} 0.0503\mathrm{e}-07 \\ -0.1270\mathrm{e}-07 \\ -0.1214\mathrm{e}-07 \\ 0.0453\mathrm{e}-07 \end{bmatrix},$$

$$A\xi^{\mathrm{I},*} - \kappa_\lambda^{\mathrm{I},*} = \begin{bmatrix} 0.0281\mathrm{e}-07 \\ 0.0002\mathrm{e}-07 \end{bmatrix},$$

$$\text{with } \chi_l^{\mathrm{I},*} = \begin{bmatrix} 0.0000 \\ 0.0000 \end{bmatrix},\ \chi_u^{\mathrm{I},*} = \begin{bmatrix} 0.1557\mathrm{e}-07 \\ 0.1145\mathrm{e}-07 \end{bmatrix},\ \xi^{\mathrm{I},*} = \begin{bmatrix} -0.0606 \\ 0.0606 \\ 0.0571 \\ -0.0571 \end{bmatrix},\ \eta^{\mathrm{I},*} = \begin{bmatrix} -0.5458 \\ 0.0002 \end{bmatrix},$$

$$\kappa_y^{\mathrm{I},*} = \begin{bmatrix} 0.0152\mathrm{e}-06 \\ 0.0000 \\ 0.0000 \\ 0.0050\mathrm{e}-06 \end{bmatrix},\ \kappa_\lambda^{\mathrm{I},*} = \begin{bmatrix} 0.1842\mathrm{e}-06 \\ 0.0015\mathrm{e}-06 \end{bmatrix},$$

$$\begin{bmatrix} By^{\mathrm{I},*} + c + D_{\mathrm{I}}^{\top}x^{\mathrm{I},*} + D_{\mathrm{II}}^{\top}x^{\mathrm{II},*} + A^{\top}\lambda^{\mathrm{I},*} \\ a - Ay^{\mathrm{I},*} \end{bmatrix} = \begin{bmatrix} -0.1221\mathrm{e}-13 \\ 0.0155\mathrm{e}-13 \\ 0.0444\mathrm{e}-13 \\ 0.0644\mathrm{e}-13 \\ 0.0022\mathrm{e}-13 \\ 0.0000 \end{bmatrix}.$$

For firm II:

$$\operatorname{diag}(\tau_1^{\mathrm{II}},\tau_2^{\mathrm{II}})x^{\mathrm{II},*} - D_{\mathrm{I}}y^{\mathrm{II},*} - \chi_l^{\mathrm{II},*} + \chi_u^{\mathrm{II},*} + D_{\mathrm{II}}\xi^{\mathrm{II},*} = \begin{bmatrix} -0.4596\mathrm{e}-06 \\ -0.4444\mathrm{e}-06 \end{bmatrix},$$

$$-D_{\mathrm{II}}^{\top}x^{\mathrm{II},*} + B\xi^{\mathrm{II},*} - A^{\top}\eta^{\mathrm{II},*} - \kappa_y^{\mathrm{II},*} = \begin{bmatrix} -0.0232\mathrm{e}-06 \\ 0.0057\mathrm{e}-06 \\ 0.0065\mathrm{e}-06 \\ -0.0069\mathrm{e}-06 \end{bmatrix},$$

$$A\xi^{\mathrm{II},*} - \kappa_\lambda^{\mathrm{II},*} = \begin{bmatrix} 0.0000 \\ 0.1287\mathrm{e}-06 \end{bmatrix},$$

$$\text{with } \chi_l^{\mathrm{II},*} = \begin{bmatrix} 0.0000 \\ 0.0000 \end{bmatrix},\ \chi_u^{\mathrm{II},*} = \begin{bmatrix} 0.0997\mathrm{e}-05 \\ 0.1018\mathrm{e}-05 \end{bmatrix},\ \xi^{\mathrm{II},*} = \begin{bmatrix} 0.0267 \\ -0.0267 \\ -0.0291 \\ 0.0292 \end{bmatrix},\ \eta^{\mathrm{II},*} = \begin{bmatrix} 0.0001 \\ -0.6425 \end{bmatrix},$$

$$\kappa_y^{\mathrm{II},*} = \begin{bmatrix} 0.0000 \\ 0.0002\mathrm{e}-04 \\ 0.0001\mathrm{e}-04 \\ 0.0000 \end{bmatrix},\ \kappa_\lambda^{\mathrm{II},*} = \begin{bmatrix} 0.0000 \\ 0.1287\mathrm{e}-04 \end{bmatrix},$$

$$\begin{bmatrix} By^{\mathrm{II},*} + c + D_{\mathrm{I}}^\top x^{\mathrm{I},*} + D_{\mathrm{II}}^\top x^{\mathrm{II},*} + A^\top \lambda^{\mathrm{II},*} \\ a - Ay^{\mathrm{II},*} \end{bmatrix} = \begin{bmatrix} 0.6957\mathrm{e}-06 \\ -0.3518\mathrm{e}-06 \\ 0.0000 \\ 0.0000 \\ 0.0000 \\ 0.0000 \end{bmatrix},$$

where $\chi_l^\nu, \chi_u^\nu, \xi^\nu, \eta^\nu, \kappa_y^\nu$ and κ_λ^ν are the Lagrange multipliers for the constraints $-x^\nu \leq 0, x^\nu - \sigma^\nu \leq 0, By + c + D_{\mathrm{I}}^\top x^{\mathrm{I}} + D_{\mathrm{II}}^\top x^{\mathrm{II}} + A^\top \lambda = 0, a - Ay = 0, y \geq 0$ and $\lambda \geq 0$, respectively.

Appendix C

Distributionally robust Nash equilibrium and ex-post equilibrium in a one-stage stochastic Nash game

In Remark 4.3.1, we have described the difference of TSDRNE and ex-post equilibrium in the sense of the formulations of two-stage distributionally robust variational inequalities.

As a specific case in which a distributionally robust Nash equilibrium exists whereas the ex-post equilibrium does not, consider the following example borrowed from [77].

Example C.0.1 (Boxed pig (for details, see Liu et al. [77, Example 2.1])). *A large pig and a piglet are placed in a space with a lever at one end of the space and food dispenser at the other end. The pig that presses the lever must run to the other side to eat, and by the time it gets there, the other pig has eaten most, but not, all of the food. The large pig is dominant, and the piglet is subordinate. Therefore, the large pig is able to prevent the piglet from getting any food when both are at the food. When the large pig presses the lever, a disutility of $\alpha = 6$ units will be incurred for the large pig and $\alpha = 2$ units for the piglet (which can be interpreted as the energies to be consumed), and ξ units (a random variable taking integer values) of food will be released at the dispenser.*

The pigs have two choices, whether to press the lever or wait at the dispenser. Because the large pig dominates the game, if it gets to the dispenser first (wait at the dispenser) or at the same time (both press the lever and then run to the dispenser) as the piglet, it will receive the following amount of food:

$$p_d(\xi) := \begin{cases} \xi & \text{if } \xi \leq 9; \\ 9 + \log(\xi - 9) & \text{if } \xi \geq 10. \end{cases}$$

The piglet will receive the rest. Instead, if the piglet waits at the dispenser first, it will receive

$$p_s(\xi) := \begin{cases} \xi & \text{if } \xi \leq 4; \\ 4 + \log(\xi - 4) & \text{if } \xi \geq 5. \end{cases}$$

The game is summarized in Table C.1.

Now suppose that the random variable ξ follows the two potential distributions $P_1(\xi = 4) = 1/4$ and $P_1(\xi = 15) = 3/4$ or $P_2(\xi = 4) = 3/4$ and $P_2(\xi = 15) = 1/4$. The ambiguity set $\mathscr{P}$ of each pig is $\mathscr{P} = \{P_1, P_2\}$.

Table C.1: Boxed pigs

	(piglet) Pull the lever	(piglet) Wait
(big pig) Pull the lever	$(p_d(\xi) - 6, \xi - p_d(\xi) - 2)$	$(\xi - p_s(\xi) - 6, p_s(\xi))$
(big pig) Wait	$(p_d(\xi), \xi - p_d(\xi) - 2)$	$(0, 0)$

First, we consider a case in which both pigs initially know that ξ follows P_1. The expected utility under P_1 is shown in Table C.2, where the stochastic Nash equilibrium is displayed in bold. The table indicates that, as the stochastic Nash equilibrium, either one of pigs pulls the lever, and the other pig waits. Next, consider a case where neither pig knows the exact probability distribution. Suppose they play the game under the worst-case expected utilities. The table of the worst-case expected utilities over $\mathscr{P}$ is summarized in Table C.3. As the distributionally robust Nash equilibrium, both pigs wait.

Table C.2: Stochastic Nash equilibrium under P_1 (**bold fonts**)

	(piglet) Pull the lever	(piglet) Wait
(big pig) Pull the lever	$(3.0938, 1.1562)$	$\mathbf{(0.4516, 5.7984)}$
(big pig) Wait	$\mathbf{(9.0938, 1.1562)}$	$(0, 0)$

Table C.3: Distributionally robust Nash equilibrium (**bold fonts**)

	(piglet) Pull the lever	(piglet) Wait
(big pig) Pull the lever	$(-0.3021, -0.9479)$	$(-3.8495, 4.5995)$
(big pig) Wait	$(5.6979, -0.9479)$	$\mathbf{(0, 0)}$

These two results suggest that no ex-post equilibrium exists because the intersection between the sets of stochastic Nash equilibria under P_1 and the distributionally robust Nash equilibria is empty. Therefore, the concept of distributionally robust equilibria is weaker than that of ex-post equilibria.