Globally and Superlinearly Convergent Algorithms for Two-Stage Stochastic Variational Inequalities and Their Applications

Xiaozhou Wang

PhD

The Hong Kong Polytechnic University

2022

The Hong Kong Polytechnic University

Department of Applied Mathematics

Globally and Superlinearly Convergent Algorithms for Two-Stage Stochastic Variational Inequalities and Their Applications

Xiaozhou Wang

A thesis submitted in partial fulfilment of the requirements

for the degree of Doctor of Philosophy

December 2021

CERTIFICATE OF ORIGINALITY

I hereby declare that this thesis is my own work and that, to the best of my knowledge and belief, it reproduces no material previously published or written, nor material that has been accepted for the award of any other degree or diploma, except where due acknowledgement has been made in the text.

____________________________(Signed)

WANG Xiaozhou (Name of student)

Dedicate to my parents.

Abstract

The two-stage stochastic variational inequality (SVI) is a powerful modeling paradigm for many real applications in the fields of finance, engineering and economics, which characterizes the first-order optimality condition of the two-stage stochastic program and models some equilibrium problems under uncertain environments. The research for two-stage SVI has received much attention during past decades. Numerically, we solve the sample discretization problem of the two-stage SVI, which is a large-scale problem due to the large sample size. Many existing deterministic VI solvers fail to handle such large-scale problems. The well-known progressive hedging algorithm (PHA) proposed by Rockafellar and Sun is a competitive algorithm for the large-scale monotone two-stage SVI. However, only a linear convergence rate is established for the monotone affine SVI. So far, to the best of our knowledge, there are no superlinearly convergent algorithms being developed for the two-stage SVI. This thesis aims to develop globally and superlinearly convergent algorithms for the two-stage SVI.

Firstly, a projection semismooth Newton algorithm (PSNA) is proposed, which is a hybrid algorithm that combines the projection algorithm and the classic semismooth Newton algorithm. At each step of PSNA, the second stage problem is split into a number of small problems and solved in parallel for a fixed first stage decision iterate. The projection algorithm and the semismooth Newton algorithm are used to find a new first stage decision iterate. The global convergence and the superlinear convergence rate are established under suitable assumptions. Numerical results for

monotone problems show that PSNA outperforms PHA. Moreover, PSNA is efficient to solve some nonmonotone problems which PHA fails to solve.

Secondly, a regularized PSNA (rPSNA) is developed to solve a two-stage stochastic linear complementarity problem (SLCP) that describes the global crude oil market share under the impact of the COVID-19 pandemic. The existence, uniqueness and robustness of the solution to the model are analyzed. As the regularized parameter goes to zero, the sequence generated by rPSNA converges to the unique solution of the single-stage SVI reformulation of the original problem. Numerical results for randomly generated examples illustrate that rPSNA performs better than PHA in terms of the number of iterations as well as CPU time. In addition, the two-stage SLCP model is applied to recover and predict the crude oil market share under the influence of COVID-19 with related parameters determined by oil data from reliable sources. This problem is solved by rPSNA efficiently, and the solution obtained is suitable to explain and rationalize the behavior of main oil-producing countries.

Lastly, rPSNA is further applied to solve two classes of nonmonotone traffic assignment problems. One is the stochastic user equilibrium problem in the form of the two-stage SVI. The second is the stochastic dynamic user equilibrium problem, which is formulated as a differential linear stochastic complementarity system (DLSCS) with the discretization problem being a special two-stage SVI. Numerically, rPSNA is more efficient for solving these problems compared with PHA.

Publications Arising from the Thesis

- X. Chen, Y. Shi and X. Wang, Equilibrium oil market share under the COVID-19 pandemic, arXiv preprint arXiv:2007.15265, the 12th International Conference on Applied Energy (ICAE2020).
- J. Luo, X. Wang and Y. Zhao, Convergence of discrete approximation for differential linear stochastic complementarity systems, Numerical Algorithms, 87: 223-262, 2021.

Acknowledgements

In the first place, I would like to express my sincerest gratitude to my supervisor, Professor Xiaojun Chen, for her professional, patient and dedicated instruction over the past three and a half years. Without her professional guidance, this research would not be possible. Apart from this, I particularly thank that Professor Chen appointed me as her *Research Assistant* for half year before I officially registered as a PhD student. During this time, with the patient instruction of Professor Chen, I managed to lay a foundation for persuing my PhD degree. At the same time, Professor Chen's profound professional knowledge and rigorous academic attitude deeply inspire and encourage me both in academic research and life. I am greatly indebted to Professor Chen.

Special thanks to Professors Defeng Sun and Tim Kelley for listening to some of my research reports, and I have benefitted a lot from their professional suggestions. In particular, I want to express my heartfelt thanks to Dr. Jie Jiang, Dr. Yun Shi and Dr. Jianfeng Luo for many inspiring discussions, in which I have benefitted tremendously. Also, many thanks to Professor Liqun Qi for teaching me the course *Nonlinear optimization*, and all other teachers who have taught me a class in PolyU. I also would like to express my sincerest thanks to Professor Hailin Sun, Professor Dingtao Peng, Dr. Xiao Wang, Dr. Zaikun Zhang, Dr. Bo Wen, Dr. Lei Yang, Dr. Lin Chen, Dr. Zhenhua Peng, Dr. Liang Chen, Dr. Bo Jiang, Dr. Liaoyuan Zeng, Dr. Fang He, Dr. Chao Li, Mr. Andy Yat Ming Cheung, Mr. Wei Liu, Mr.

Shisen Liu, Mr. Shijie Yu, Mr. Zicheng Qiu, Mr. Yifan He, Ms. Yue Wang and Mr. Xiao Zha for their help and support. It greatly broadens my academic horizon by attending to their reports in our regular group seminars.

Finally, I want to express deepest thanks to my family, for their love and instilling in me an attitude of learning from the childhood. Especially, I want to express my sincerest gratitude and deepest love to my wife Ting Zou for her consistent love and encouragement during my PhD study.

Contents

List of Figures

List of Tables

List of Notation

$\mathbb{R}^n$	the n-dimensional real space
$\mathbb{R}^n_+$	the nonnegative orthant
$\mathbb{R}^{m\times n}$	the space of $m \times n$ real matrices
$x \perp y$	vectors x and y are perpendicular
A^T/x^T	the transpose of a matrix A /vector x
$\|x\|$	the ℓ_2-norm of vector x, unless otherwise specified
$\|A\|$	the Euclidean norm of a matrix A defined as the square root of the maximum eigenvalue of A^TA unless otherwise specified, i.e., $\sqrt{\lambda_{\max}(A^TA)}$,
$A_{\alpha\beta}$	the submatrix of A with row and column indexed by α and β, respectively
A^{-1}	the inverse of a nonsingular matrix A
$\min(a,b)$	the vector whose i-th component is $\min(a_i,b_i)$
$\mathrm{mid}(a,b,x)$	the mid function with given a and b defined by the projection operator $\Pi_{[a,b]}(x)$
I_n	the identity matrix of order n (subscript often omitted)
$\nabla F(x)$	the Jacobian of the vector-valued function F at x
$F'(x;d)$	the directional derivative of the vector-valued function F along the direction d

$\partial F(x)$	the Clarke generalized Jacobian of the vector-valued function F at x
$\mathcal{B}(x, \delta)$	the open neighborhood centered at x with radius $\delta > 0$
cl C	the topological closure of a set C
$\mathcal{N}_C(x)$	the normal cone of the set C at $x \in C$
$\Pi_C(x)$	the Euclidean projection of x onto the set C
LCP(q, M)	the linear complementarity problem defined by the vector q and matrix M
SOL(q, M)	the solution set of the LCP(q, M)
VI(C, F)	the variational inequality defined by the set C and function F
$o(t)$	any function $f(t)$ such that $\lim_{t \to 0} f(t)/t = 0$

Chapter 1

Introduction

1.1 Background

Let $(\Xi, \mathcal{A}, P)$ be a probability space induced by a random vector ξ, in which $\Xi \subseteq \mathbb{R}^d$ is the support set of ξ, $\mathcal{A}$ is the Borel sigma algebra of Ξ and P is a probability measure. Let $\mathcal{Y}$ be the space consisting of $\mathcal{A}$-measurable functions from Ξ to $\mathbb{R}^m$. Consider the following two-stage stochastic variational inequalities (SVI) [6]:

$$-\mathbb{E}[G(x, y(\xi), \xi)] \in \mathcal{N}_D(x), \tag{1.1}$$

$$-F(x, y(\xi), \xi) \in \mathcal{N}_{C(\xi)}(y(\xi)), \qquad \text{for almost every (a.e.) } \xi \in \Xi, \tag{1.2}$$

where

- $G : \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^d \to \mathbb{R}^n$ is a vector-valued map, Lipschitz continuous with respect to (x, y) with Lipschtiz constant $L_G(\xi)$ for a.e. $\xi \in \Xi$, and $\mathcal{A}$-measureable and integrable with respect to ξ;

- $F : \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^d \to \mathbb{R}^m$ is a vector-valued map, continuously differentiable with respect to (x, y) for a.e. $\xi \in \Xi$, and $\mathcal{A}$-measureable with respect to ξ;

- $\mathbb{E}[\cdot]$ is the expected operator over Ξ, $D \subseteq \mathbb{R}^n$ is a nonempty closed convex set and $C(\xi) \subseteq \mathbb{R}^m$ is a polyhedral set for a.e. $\xi \in \Xi$, and $\mathcal{N}_D(x)$ and $\mathcal{N}_{C(\xi)}(y(\xi))$ are normal cones to the set D at $x \in \mathbb{R}^n$ and the set $C(\xi)$ at $y(\xi) \in \mathbb{R}^m$, respectively.

According to the definition of the normal cone, (1.1)-(1.2) is equivalent to finding $(x, y(\cdot)) \in \mathbb{R}^n \times \mathcal{Y}$ such that the following inequalities are satisfied:

$$(x' - x)^T \mathbb{E}[G(x, y(\xi), \xi)] \geq 0, \qquad \forall\, x' \in D,$$

$$(y'(\xi) - y(\xi))^T F(x, y(\xi), \xi) \geq 0, \qquad \forall\, y'(\xi) \in C(\xi), \text{ for a.e. } \xi \in \Xi.$$

The two-stage SVI (1.1)-(1.2) is the generalization of the single-stage SVI [5, 12]. The random vector ξ is introduced to describe the stochastic factors from the unknown future or uncertain environments. A distinct feature for the two-stage SVI is that its decision variables are comprised of two types, the *here-and-now* decision variable x that does not depend on the realization of the random vector ξ and the *wait-and-see* decision variable $y(\xi)$ that depends on ξ. A solution $(x, y(\cdot))$ to the two-stage SVI is such that the collection of inclusions in (1.1)-(1.2) is satisfied, where the second stage inclusions (1.2) should hold for a.e. $\xi \in \Xi$. In practice, a here-and-now decision has to be determined before the observation of the realization of ξ. The two-stage SVI is a powerful mathematical model that allows decision-makers to make a decision here and now by taking into account every possible realization of ξ. This illustrates the reliability and robustness of the solution to the two-stage SVI compared with the single-stage SVI.

Numerically, we solve the discretization problem of (1.1)-(1.2). More specifically, given a set of samples $\{\xi_1, \ldots, \xi_\nu\}$ of the random vector ξ, the discrete approximation problem of (1.1)-(1.2) reads

$$-\sum_{\ell=1}^{\nu} p(\xi_\ell) G(x, y(\xi_\ell), \xi_\ell) \in \mathcal{N}_D(x), \tag{1.3}$$

$$-F(x, y(\xi_\ell), \xi_\ell) \in \mathcal{N}_{C(\xi_\ell)}(y(\xi_\ell)), \quad \ell = 1, \ldots, \nu, \tag{1.4}$$

where $p(\xi_\ell) > 0$ for $\ell = 1, \ldots, \nu$ and $\sum_{\ell=1}^{\nu} p(\xi_\ell) = 1$. If the sample set is independent and identically distributed (i.i.d.), then (1.3)-(1.4) is called the sample average approximation (SAA) discretization problem. One can refer to [6, 8, 10] for the convergence analysis of the solution of the SAA discretization problem to that of the two-stage SVI (1.1)-(1.2). (1.3)-(1.4) is a large-scale deterministic problem due to the large sample size ν. Thus, many deterministic VI solvers encounter difficulties in handling it, since they are designed to solve small- to medium-scale problems. So, it is meaningful and necessary to develop specialized algorithms that are globally and superlinearly convergent for solving the two-stage SVI. However, no such algorithms are available at present. This thesis is devoted to develop globally and superlinearly convergent algorithms for this problem.

1.2 Literature review

The two-stage SVI (1.1)-(1.2) is first proposed by Chen, Pong and Wets in [6], which is the generalization of the single-stage SVI. The two-stage SVI characterizes the first-order optimality conditions of the two-stage stochastic optimization problems [6, 44] and models some stochastic equilibrium problems in stochastic environments [9, 25, 45, 48]. The two-stage SVI was studied extensively; see [8, 10, 33, 42, 43, 44] for references. One can refer to [42] for the extension of the two-stage SVI to the multi-stage SVI.

In the case that $G(\cdot,\cdot,\xi)$ and $F(\cdot,\cdot,\xi)$ are both linear with respect to (x,y) for a.e. $\xi \in \Xi$, $D = \mathbb{R}^n_+$, and $C(\xi) = \mathbb{R}^m_+$ for a.e. $\xi \in \Xi$, (1.1)-(1.2) reduces to a two-stage SLCP as follows:

$$0 \leq x \perp \tilde{A}x + \mathbb{E}[B(\xi)y(\xi)] + c \geq 0, \tag{1.5}$$

$$0 \leq y(\xi) \perp N(\xi)x + M(\xi)y(\xi) + q(\xi) \geq 0, \quad \text{for a.e. } \xi \in \Xi, \tag{1.6}$$

where $\tilde{A} \in \mathbb{R}^{n\times n}$, $c \in \mathbb{R}^n$, $B : \mathbb{R}^d \to \mathbb{R}^{n\times m}$, $N : \mathbb{R}^d \to \mathbb{R}^{m\times n}$, $M : \mathbb{R}^d \to \mathbb{R}^{m\times m}$ and $q : \mathbb{R}^d \to \mathbb{R}^m$. (1.5)-(1.6) is the first-order optimality condition for the two-stage stochastic linear program [6]. In [10], the existence and uniqueness of the solution of the two-stage SLCP are studied under the strong monotonicity assumption. In addition, a new discretization scheme is proposed and a distributionally robust two-stage SLCP is studied. In [8], the existence and uniqueness of the solution are established for a two-stage SVI-NCP, in which the first stage problem is an SVI and the second stage is a stochastic nonlinear complementarity problem (NCP). The properties of the Lipschitz continuity and monotonicity of the second stage solution function for any fixed x and ξ are also discussed. In addition, the expression for the generalized Jacobian of the second stage solution function is given. The convergence of the solution of the SAA discretization problem to that of the original problem is studied, in which the convergence rate is shown to be exponential with respect to the sample size.

The well-known progressive hedging algorithm (PHA) was first proposed by Rockafellar and Wets [41] to solve convex multi-stage stochastic optimization problems. Recently, it was extended to solve the monotone multi-stage SVI by Rockafellar and Sun with a finite support set [40]. PHA is a competitive algorithm since it decomposes the original large-scale problem into a sequence of independent small sample-based subproblems solved in parallel. Theoretically, PHA is globally convergent for the monotone SVI. However, only the linear convergence rate is established for the affine monotone SVI and it is not applicable to nonmonotone problems. Recently, an elicited PHA was proposed by Zhang, Sun and Xu [48] to solve the elicited monotone (not necessarily monotone) SVI. However, it is difficult to verify the elicited monotonicity of the problem, and the convergence rate is still linear. So far, little attention has been paid to develop a superlinearly convergent algorithm for solving the two-stage SVI.

In order to design a superlinearly convergent algorithm, the most natural choice is the Newton-type algorithms. One desirable property is that the local convergence rate can be superlinear if the problem is semismooth [35]. One issue is that only the local convergence can be obtained. To guarantee the global convergence, one can exploit the line search technique when the starting point is far from the solution, which is called the damped Newton algorithm. Newton-type algorithms have received much attention for solving nonsmooth equations, complementarity problems and variational inequalities; see [7, 15, 24, 31, 35] for details. One may simply apply the classical damped semismooth Newton algorithms [35, 36] for solving the two-stage SVI. However, the storage and calculation cost of large-scale Jacobian matrices are prohibitively high, and the line search is inefficient for large-scale problems. So far, no specialized Newton-type algorithms are developed for the two-stage SVI.

Next, we review some important applications from the deterministic VI to the two-stage SVI. The deterministic VI plays an important role in modeling many equilibrium problems in finance, economic and engineering [17, 19, 20]. As a generalization of the deterministic VI, the single-stage SVI is capable of describing equilibrium problems under uncertain environments; see [5, 18, 21, 23, 26, 38]. The two-stage SVI as a further generalization of the single-stage SVI, is able to model equilibrium problems under stochastic environments and involves a dynamic decision process; see [6, 42]. In [6], a Warlas equilibrium problem under uncertain environments is formulated as a two-stage SVI, which has a variety of applications in finance, international commodity trading and species interaction in ecological models. In [25], the two-stage SVI is used to characterize the first-order optimality condition of a convex two-stage noncooperative multi-agent quadratic game. In addition, the two-stage quadratic game model is applied to describe the two-stage stochastic production and supply planing problem in the oligopolistic crude oil market. Numerical simulations with historical data from the crude oil market show the efficiency of the model to recover and predict market shares of the main oil-producing countries over the world. Recently, the COVID-19 pandemic has brought a significant impact on the global crude oil market. During the period, the crude oil price and demand are violently volatile. However, market shares of main oil-producing countries remained relative stable despite the influence of COVID-19. This arouses the question if the stable oil market share can be explained by the results of some kinds of stochastic Nash equilibrium problems. Most importantly, calculating a solution of some stochastic game problems leads to solving a large-scale two-stage SVI, which can be difficult to solve if no superlinear convergence algorithms are developed.

The stochastic Wardrop flow equilibrium in the stochastic traffic assignment is an important class of applications of the two-stage SVI. One can refer to [6, 12, 19, 37] for references. The applications of the single-stage SVI for the stochastic Wardrop flow equilibrium have been studied by Zhang, Chen and Sumalee in [46, 47]. In [12], Chen, Wets and Zhang proposed an expected residual minimization (ERM) formulation for a new residual function, which is a two-stage stochastic program that involves the here-and-now decision variable and the wait-and-see decision variable. This new model is then applied to describe the stochastic Wardrop flow equilibrium. In [6], the stochastic traffic assignment problem on the link flow is formulated as a two-stage SVI, which is further rewritten as a two-stage stochastic program with recourse and solved by the Douglas-Rachford splitting method [19]. In general, the two-stage SVI model can produce a more desirable solution than that of the single-stage model. However, to solve two-stage transportation problems poses a challenge for the well-known solver PHA since these problems may be nonmonotone and nonlinear.

On the other hand, the differential variational inequality (DVI) [34], which is comprised of ordinary differential equations (ODEs) and dynamic complementarity systems, formulates the dynamic Wardrop flow equilibrium that predicts the traffic states over a short-term time horizon. One can refer to [2, 3, 4, 32, 37] for references. To deal with stochastic factors, one may introduce a random vector in DVI, which can model the stochastic dynamic Wardrop flow equilibrium. This thesis also gives the differential linear stochastic complementarity systems (DLSCS) [29] to model the stochastic dynamic Wardrop flow equilibrium. Its discretization problem can be viewed as a special two-stage SVI in which the first stage problem is a mixed LCP (mLCP) and the second stage problem is an LCP. Nevertheless, since the discretization problem is large-scale and nonmonotone, this brings difficulties to many existing algorithms. Therefore, developing new efficient and superlinear convergence algorithms is needed.

1.3 Summary of contributions of the thesis

The contributions of the thesis are summarized as follows.

- A globally and superlinearly convergent projection semismooth Newton algorithm (PSNA) is developed to solve the two-stage SVI (1.1)-(1.2) based on its single-stage SVI reformulation. PSNA has the decomposition feature like PHA and detours the difficulties of calculating and storing large-scale Jacobian matrices, which enable it to solve large-scale problems. The global convergence and superlinear convergence rate of PSNA are well-established under suitable assumptions. Numerically, PSNA outperforms PHA for solving monotone problems and it is competitive for some nonmonotone problems.

- A regularized PSNA (rPSNA) is proposed to solve a new two-stage SLCP model which describes the global crude oil market share problem under the COVID-19 pandemic. For any fixed first stage variable and realization of a random variable, the existence and uniqueness of the solution to the second stage problem are studied. We give an expression for the unique least-norm solution of the second stage problem, and the robustness of the least-norm solution is explored. The global convergence of rPSNA is established as the regularized parameter goes to zero. Numerically, rPSNA is efficient to calculate a solution of the two-stage SLCP model, and the solution is used to explain and forecast the stability of the crude oil market share under the COVID-19 pandemic.

- We further apply rPSNA to solve a nonmonotone stochastic user equilibrium problem and a nonmonotone stochastic dynamic user equilibrium problem. The application of rPSNA to the latter problem extends the usage of the algorithm. Numerically, rPSNA outperforms PHA in solving these problems.

1.4 Organization of the thesis

The thesis is organized as follows.

- In Chapter 1, we introduce the background of the thesis and give a literature review for the related topics.

- In Chapter 2, we develop PSNA for the two-stage SVI (1.1)-(1.2). The original two-stage problem is first reformulated as a single-stage SVI by substituting the second stage Lipschitz solution function into the first stage problem. We then investigate the solvability, Lipschitz continuity, semismoothness, linear Newton approximation scheme and monotonicity of the single-stage problem. Based on the single-stage problem, PSNA is developed. The global convergence and superlinear convergence rate are established under suitable assumptions. Next, PSNA is applied to the two-stage semi-linear SVI, which is a special case of (1.1)-(1.2). In this case, assumptions for guaranteeing the global convergence and superlinear convergence rate can be verified under suitable conditions. Numerical experiments are conducted to show the competitiveness of PSNA compared with PHA.

- In Chapter 3, we propose rPSNA to solve a two-stage SLCP which models the global crude oil market share problem under the impact of the COVID-19 pandemic. The existence, uniqueness and robustness of the least-norm solution of the second stage problem are studied. The global convergence for rPSNA is established. Numerical results for randomly generated problems show that rPSNA is more efficient than PHA in terms of the number of iterations and CPU time. The proposed rPSNA is also applied to calculate a solution of the model with parameters being determined by real data from the crude oil market. The solution obtained by the model is used to explain and forecast the global crude oil market share under the COVID-19 pandemic.

- In Chapter 4, rPSNA is applied to solve a nonmonotone stochastic user equilibrium problem and a nonmonotone stochastic dynamic user equilibrium problem. Numerically, rPSNA is promising in solving these nonmonotone problems compared with PHA.

- In Chapter 5, we summarize the main conclusions of the thesis and discuss some future work.

Chapter 2

A globally and superlinerly convergent projection semismooth Newton algorithm for two-stage stochastic variational inequalities

In this chapter, we propose PSNA to solve the two-stage SVI (1.1)-(1.2), which has the global convergence property and superlinear convergence rate under suitable assumptions.

We assume that (1.1)-(1.2) has the relatively complete recourse [8]; that is, for any $x \in D$ and a.e. $\xi \in \Xi$, the second stage problem (1.2) has at least one solution. Among these solutions, we further assume that there exists a Lipshitz continuous solution function with respect to x for a.e. ξ. By substituting the Lipschitz solution function into the first stage problem (1.1), we obtain a single-stage SVI, in which x is the only decision variable. Then we propose PSNA based on the single-stage SVI for solving (1.1)-(1.2), which is a hybrid algorithm that combines the semismooth Newton algorithm with the projection algorithm [19]. The advantages for using the single-stage formulation are that PSNA detours the difficulties of handling the large-scale original problem directly, and the computations of the second stage solution can be executed in parallel. The solvability, Lipschitz continuity, semimsoothness, linear Newton approximation scheme and monotonicity of the single-stage SVI are analyzed under proper assumptions, which provide the foundation for the development of PSNA. The Lipschitz continuity and the monotonicity guarantee that the projection algorithm is well executed, and the semismoothness and the linear Newton approximation scheme properties are essential for the semismooth Newton algorithm. It is worth noting that if the two-stage SVI (1.1)-(1.2) is monotone, then its single-stage SVI formulation is monotone, but conversely it is not true. Hence the conditions for the global convergence of PSNA are weaker than the conditions for the global convergence of PHA [40].

2.1 A projection semismooth Newton algorithm (PSNA)

In this section, we first reformulate the two-stage SVI (1.1)-(1.2) into a single-stage SVI under suitable assumptions. Based on the single-stage formulation, we discuss its properties and develop PSNA.

Let $K : \mathbb{R}^k \to \mathbb{R}^s$ be a locally Lipschitz continuous function. According to Rademacher's Theorem, K is differentiable almost everywhere. Let Ω_K be the set of differentiable points of K. The generalized Jacobian of K at x in the sense of Clarke [16] is defined as follows:

$$\partial K(x) := \text{conv}\{V \in \mathbb{R}^{s\times k} \ : \ V = \lim_{x^t \in \Omega_K, x^t \to x} \nabla_x K(x^t)\},$$

where "conv" denotes the convex hull. K is said to be semismooth at x if K is locally Lipschitz continuous around x and the limit [36]

$$\lim_{\substack{V\in\partial K(x+th'),\\ h'\downarrow h,\ t\downarrow 0}} \{Vh\}$$

exists for any $h \in \mathbb{R}^k$.

For any given $(x,\xi) \in D \times \Xi$, let $\mathcal{S}(x,\xi)$ be the solution set of the second stage problem (1.2). Thus, we can define a set-valued map $\mathcal{S}(x,\cdot) : \Xi \rightrightarrows \mathcal{Y}$. Then, the two-stage SVI (1.1)-(1.2) can be reformulated as a single-stage problem finding an $x \in D$ and an integrable selection $\bar{y}(x,\cdot) \in \mathcal{S}(x,\cdot)$ which solve

$$-\mathbb{E}[G(x,\bar{y}(x,\xi),\xi)] \in \mathcal{N}_D(x).$$

If the second stage problem (1.2) admits a unique solution $\bar{y}(x,\xi)$ for any $x \in D$ and a.e. $\xi \in \Xi$, then $\mathcal{S}(x,\xi) = \{\bar{y}(x,\xi)\}$ is a singleton.

Assumption 2.1. *(i) The two-stage SVI* (1.1)-(1.2) *is solvable.*
(ii) The Lipschitz constant of $G(\cdot,\cdot,\xi)$ *satisfies* $\mathbb{E}[L_G^2(\xi)] < \infty$.
(iii) (1.1)-(1.2) *has the relatively complete recourse. In addition, there exists an integrable selection* $\hat{y}(x,\cdot) \in \mathcal{S}(x,\cdot)$ *for any* $x \in D$*, and* $\hat{y}(\cdot,\xi)$ *is Lipschitz continuous for a.e.* $\xi \in \Xi$ *which satisfies*

$$\|\hat{y}(x,\xi) - \hat{y}(x',\xi)\| \le L_{\hat{y}}(\xi)\|x - x'\|, \quad \forall\, x, x' \in D,$$

with $\mathbb{E}[L_{\hat{y}}^2(\xi)] < \infty$.

Among all the solutions in $\mathcal{S}(x,\xi)$, the Lipschitz continuous solution function $\hat{y}(x,\xi)$ is important both in studying the solvability of the single-stage problem and in developing the numerical algorithm under proper conditions. Substituting $\hat{y}(x,\xi)$ into (1.1), we can get the single-stage problem

$$-H(x) \in \mathcal{N}_D(x), \tag{2.1}$$

where

$$H(x) := \mathbb{E}[\hat{G}(x,\xi)] \quad \text{with} \quad \hat{G}(x,\xi) := G(x,\hat{y}(x,\xi),\xi).$$

(2.1) can be viewed as a deterministic problem. Denote its solution set by S^*. It is not hard to see that if $x^* \in S^*$, then

$$(x^*, \hat{y}(x^*,\cdot)) \;\text{ solves }\; (1.1) - (1.2).$$

Based on (2.1), we propose PSNA in Subsection 2.1.2.

2.1.1 Properties analysis

In this subsection, we study the solvability of (2.1), and the Lipschtiz continuity, semismoothness, linear Newton approximation scheme and monotonicity of H.

The following proposition studies the Lipschitz continuity of H and the solvability of (2.1).

Proposition 2.1. *Under Assumption 2.1, the following assertions hold.*

(i) The function H is Lipschitz continuous on D with Lipschitz constant L_H.

(ii) If D is bounded, then (2.1) *is solvable.*

(iii) If $\mathcal{S}(x,\xi)$ is a singleton for any $x \in D$ and a.e. $\xi \in \Xi$, then (2.1) *is solvable.*

Proof. (i) By the Lipschitz continuity of $G(\cdot,\cdot,\xi)$ and $\hat{y}(\cdot,\xi)$ for a.e. $\xi \in \Xi$, we have for any $x, x' \in D$

$$\begin{aligned}
\|H(x) - H(x')\| &= \|\mathbb{E}[\hat{G}(x,\xi) - \hat{G}(x',\xi)]\| \\
\leq& \mathbb{E}[\|G(x,\hat{y}(x,\xi),\xi) - G(x',\hat{y}(x',\xi),\xi)\|] \\
\leq& \mathbb{E}[L_G(\xi)L_{\hat{y}}(\xi) + L_G(\xi)]\|x - x'\| \\
=:& \mathbb{E}[L_{\hat{G}}(\xi)]\|x - x'\|.
\end{aligned}$$

By Hölder's inequality, from the last inequality we have

$$L_H := \mathbb{E}[L_{\hat{G}}(\xi)] \leq \mathbb{E}[L_G^2(\xi)]^{\frac{1}{2}}\mathbb{E}[L_{\hat{y}}^2(\xi)]^{\frac{1}{2}} + \mathbb{E}[L_G(\xi)] < \infty.$$

(ii) Since D is bounded and that H is Lipschitz continuous, from [19, Corollary 2.2.5], we immediately know that (2.1) is solvable.

(iii) $\mathcal{S}(x,\xi)$ being a singleton implies that the second stage problem has a unique solution for any $x \in D$ and a.e. $\xi \in \Xi$. The solvability of (1.1)-(1.2) is equivalent to that of (2.1) in the sense that

$$x^* \text{ solves (2.1)} \Leftrightarrow (x^*, \hat{y}(x^*,\cdot)) \text{ is a solution of (1.1)-(1.2)}.$$

$\square$

Proposition 2.1 states that (2.1) is an equivalent reformulation of (1.1)-(1.2) if $\mathcal{S}(x,\xi)$ is a singleton for any fixed x and ξ. In the case that $\mathcal{S}(x,\xi)$ has multiple elements, if D is bounded, solution to (2.1) is also a solution of the original problem.

Next, we will discuss the semismoothness and the linear Newton approximation scheme of H (see Definition 2.1).

Let $\mathcal{D} \supseteq D$ be an open set. Define the set-valued mapping $\mathcal{H} : \mathcal{D} \rightrightarrows \mathbb{R}^{n\times n}$ as

$$\mathcal{H}(x) := \mathbb{E}[\partial \hat{G}(x,\xi)],$$

with the expectation taken in the sense of Aumann [1] as

$$\mathbb{E}[\partial \hat{G}(x,\xi)] := \{\mathbb{E}[V(x,\xi)] : V(x,\xi) \text{ is an integrable selection from } \partial \hat{G}(x,\xi)\},$$

where $V(x,\xi) \in \partial \hat{G}(x,\xi)$ for a.e. $\xi \in \Xi$ and it is integrable with respect to ξ.

To analyze the properties of $\mathcal{H}$, we make the following technical assumptions.

Assumption 2.2. *(i) $\partial \hat{G}(x,\xi)$ is $\mathcal{A}$-measurable with respect to ξ at any $x \in \mathcal{D}$.*

(ii) For any $x \in \mathcal{D}$, there exists an open neighborhood such that for any x' in this neighborhood, $\partial \hat{G}(x',\xi)$ is bounded by an integrable function $\hat{\kappa}_x : \Xi \to \mathbb{R}_+$; that is, for any x' in this neighborhood

$$\|V(\xi)\|_F \leq \hat{\kappa}_x(\xi), \quad \forall\ (\xi, V(\xi)) \in \Xi \times \mathbb{R}^{n\times n} \text{ with } V(\xi) \in \partial G(x',\xi).$$

The following proposition studies the properties of $\mathcal{H}$.

Proposition 2.2. *Under Assumption 2.2, $\mathcal{H}(x)$ is nonempty, convex and compact for any $x \in \mathcal{D}$. Moreover, $\mathcal{H}$ is outer semicontinuous and closed at any $x \in \mathcal{D}$; that is, if $x^k \to x$, $W^k \in \mathcal{H}(x^k)$ and $W^k \to W$, then $W \in \mathcal{H}(x)$.*

Proof. Since $\partial \hat{G}(x,\xi)$ is closed-valued at any $x \in \mathcal{D}$ for a.e. $\xi \in \Xi$, by Assumption 2.2, it follows directly from Theorems 1, 2 and 4 in [1] that $\mathcal{H}(x)$ is nonempty, convex and compact at any $x \in \mathcal{D}$. The last assertion follows from [1, Corollary 5.2]. □

The following definition of linear Newton approximation scheme is important for the development of Newton-type algorithms.

Definition 2.1 ([19]). *Let $K : \mathbb{R}^s \to \mathbb{R}^s$ be a locally Lipschitz continuous function. We say that K admits a linear Newton approximation at $\bar{x}$, if there is a set-valued mapping $\Psi : \mathbb{R}^s \rightrightarrows \mathbb{R}^{s\times s}$ such that Ψ has nonempty compact images, outer semicontinuous at $\bar{x}$, and for any $h \to 0$, $W \in \Psi(\bar{x}+h)$,*

$$K(\bar{x}+h) - K(\bar{x}) - Wh = o(\|h\|). \tag{2.2}$$

We also say that Ψ is a linear Newton approximation scheme of K at $\bar{x}$.

By Definition 2.1, ∂H is a linear Newton approximation scheme of H if H is semismooth. However, the calculation of ∂H is difficult since the explicit form of H is not available. In some cases, the calculation of $\mathcal{H}(x)$ is practical. So, we turn to discuss its linear Newton approximation scheme.

To establish that $\mathcal{H}$ is a linear Newton approximation scheme of H, the semismoothness of $\hat{G}(\cdot,\xi)$ is needed. If $\hat{G}(\cdot,\xi)$ is semismooth at x, by the properties of semismoothness [36], there exists a function $\Delta_\xi : (0,\infty) \to [0,\infty)$, depending on ξ, with $\lim_{t\to 0}\Delta_\xi(t) = 0$, such that for all $h \neq 0$ sufficiently small, and all $V(\xi) \in \partial\hat{G}(x+h,\xi)$,

$$\frac{\|\hat{G}(x+h,\xi) - V(\xi)h - \hat{G}(x,\xi)\|}{\|h\|} \leq \Delta_\xi(\|h\|). \tag{2.3}$$

The following technical assumption is needed.

Assumption 2.3. *There exists an integrable function $\gamma : \Xi \to \mathbb{R}_+$ such that $\Delta_\xi(\|h\|) \leq \gamma(\xi)$ for all sufficiently small $\|h\|$.*

Note that $\hat{G}(\cdot,\xi) = G(\cdot,\hat{y}(\cdot,\xi),\xi)$. The semismoothness of $\hat{G}(\cdot,\xi)$ is related to the semismoothness of the second stage solution function $\hat{y}(\cdot,\xi)$. To this end, we introduce the *Strong Regularity Condition* (SRC) proposed by Robinson [39]. Facchinei and Pang also thoroughly discussed this property in the monograph [19]. In the case of the VI with a polyhedral set, the SRC condition is equivalently defined as in [27], which we will use in the following Assumption 2.4.

Without loss of generality, let

$$C(\xi) = \{y \in \mathbb{R}^m : A_1(\xi)y \leq b_1(\xi), A_2(\xi)y = b_2(\xi)\},$$

with $A_1 : \mathbb{R}^d \to \mathbb{R}^{l_1 \times m}, A_2 : \mathbb{R}^d \to \mathbb{R}^{l_2 \times m}$, $b_1 : \mathbb{R}^d \to \mathbb{R}^{l_1}$ and $b_2 : \mathbb{R}^d \to \mathbb{R}^{l_2}$. For any given $x \in D$ and $\xi \in \Xi$, define the critical cone of the pair $(C(\xi), F(x, \cdot, \xi))$ at $\hat{y}(x, \xi) \in C(\xi)$ as follows

$$\mathcal{C}_x(\hat{y}; C(\xi), F) = \{v \in \mathbb{R}^m : F(x, \hat{y}(x, \xi), \xi)^T v = 0, \bar{A}_1(\xi)v \leq 0, A_2(\xi)v = 0\},$$

where $\bar{A}_1(\xi)$ is a sub-matrix of $A_1(\xi)$ consisting of rows of $A_1(\xi)$ satisfying $\bar{A}_1(\xi)y = \bar{b}_1(\xi)$ with $\bar{b}_1(\xi)$ being an appropriate sub-vector.

We make the following SRC assumption for the second stage problem.

Assumption 2.4. *For a.e. $\xi \in \Xi$, the SRC holds at $\hat{y}(x, \xi)$ for* $\mathrm{VI}(C(\xi), F(x, \cdot, \xi))$ *for any $x \in D$; that is, for a.e. $\xi \in \Xi$ and any $x \in D$, the following affine VI admits a unique solution for each $q \in \mathbb{R}^m$*

$$0 \in q + \nabla_y F(x, \hat{y}(x, \xi), \xi)v + \mathcal{N}_{\mathcal{C}_x(\hat{y}; C(\xi), F)}(v).$$

It is clear that Assumption 2.4 holds if $F(x, \cdot, \xi)$ is strongly monotone on $C(\xi)$ for any $x \in D$ and a.e. $\xi \in \Xi$. In the case that $C(\xi) = \mathbb{R}^m_+$ for a.e. $\xi \in \Xi$, a sufficient condition for guaranteeing Assumption 2.4 is that $F(x, \cdot, \xi)$ is a uniformly P function for any $x \in D$ and a.e. $\xi \in \Xi$; see [19, Corollary 5.3.20]. If further let $F(\cdot, \cdot, \xi)$ be a linear function as in (1.5), then $M(\xi)$ being a P-matrix for a.e. $\xi \in \Xi$ is sufficient; see [14].

The following proposition establishes the semismoothness of H at x and shows that $\mathcal{H}$ is a linear Newton approximation scheme of H at x.

Proposition 2.3. *Suppose that Assumptions 2.1-2.4 hold, and that* $G(\cdot,\cdot,\xi)$ *is semismooth at* $(x, \hat{y}(x,\xi))$ *for a.e.* $\xi \in \Xi$. *Then we have the following assertions.*

(i) $\hat{G}(\cdot,\xi)$ *is semismooth at* x *for a.e.* $\xi \in \Xi$ *and* H *is semismooth at* x.

(ii) $\mathcal{H}$ *is a linear Newton approximation scheme of* H *at* x.

Proof. (i) With Assumption 2.4, by [19, Theorem 5.4.6], we know that for a.e. $\xi \in \Xi$, $\hat{y}(\cdot,\xi)$ is a piecewise smooth function on D, and hence it is semismooth on D. By [19, Proposition 7.4.4], the composition of semismooth functions is also semismooth. Then, we deduce that $\hat{G}(\cdot,\xi)$ is semismooth at x for a.e. $\xi \in \Xi$.

With Assumption 2.1, H is Lipschitz continuous on D by Proposition 2.1. The semismoothness of $\hat{G}(\cdot,\xi)$ at x implies that $\hat{G}(\cdot,\xi)$ is directionally differentiable at x for a.e. $\xi \in \Xi$. In addition, $\hat{G}(\cdot,\xi)$ is Lipschitz continuous on D with Lipschitz constant satisfying $\mathbb{E}[L_{\hat{G}}(\xi)] < \infty$. Then H is also directionally differentiable at x by [43, Proposition 2].

Under Assumption 2.2, $\mathcal{H}$ has nonempty compact images and is outer semicontinuous on D. For any $h \to 0$, $W \in \mathcal{H}(x+h)$, let $V(\xi)$ be an integrable selection from $\partial\hat{G}(x+h,\xi)$ such that $W = \mathbb{E}[V(\xi)]$. It follows that

$$\begin{aligned}
&\lim_{\substack{h\to 0,\\ W\in\partial\mathcal{H}(x+h)}} \frac{\|H(x+h)-Wh-H(x)\|}{\|h\|} \\
=&\lim_{\substack{h\to 0,\\ V(\xi)\in\partial\hat{G}(x+h,\xi)}} \frac{\|\mathbb{E}[\hat{G}(x+h,\xi)-V(\xi)h-\hat{G}(x,\xi)]\|}{\|h\|} && \text{by Assumption 2.2} \\
\leq&\lim_{\substack{h\to 0,\\ V(\xi)\in\partial\hat{G}(x+h,\xi)}} \frac{\mathbb{E}[\|\hat{G}(x+h,\xi)-V(\xi)h-\hat{G}(x,\xi)\|]}{\|h\|} \\
\leq&\lim_{h\to 0}\mathbb{E}[\Delta_\xi(\|h\|)] && \text{by (2.3)} \\
=&\mathbb{E}[\lim_{h\to 0}\Delta_\xi(\|h\|)] \\
=&0, && \text{by } \lim_{h\to 0}\Delta_\xi(\|h\|)=0 \text{ for a.e. } \xi\in\Xi
\end{aligned}$$

where the second equality is due to Assumption 2.3 and Lebesgue's dominated convergence theorem. Due to $\partial H(x) \subseteq \mathcal{H}(x)$, the last inequality implies the semismoothness of H at x by [19, Theorem 7.4.3].

(ii) In the proof of (i), we show that

$$\lim_{h \to 0, W \in \partial\mathcal{H}(x+h)} \|H(x+h) - Wh - H(x)\| = o(\|h\|),$$

which implies that $\mathcal{H}$ is a linear approximation scheme of H at x. $\square$

Next, we discuss the monotonicity of H. We first discuss the monotonicity of (1.1)-(1.2). Define a mapping $\mathcal{T} : \mathbb{R}^n \times \mathcal{Y} \to \mathbb{R}^n \times \mathcal{Y}$ as

$$\mathcal{T}(x, y(\cdot)) = \begin{pmatrix} \mathbb{E}[G(x, y(\xi), \xi)] \\ F(x, y(\cdot), \cdot) \end{pmatrix}.$$

We say that $\mathcal{T}$ is monotone on $D \times C(\cdot)$ if for any $(x, y(\cdot))$, $(x', y'(\cdot)) \in D \times C(\cdot)$[1], it holds that

$$\left\langle \mathcal{T}(x, y(\cdot)) - \mathcal{T}(x', y'(\cdot)), \begin{pmatrix} x - x' \\ y(\cdot) - y'(\cdot) \end{pmatrix} \right\rangle \geq 0, \tag{2.4}$$

where the scalar product $\langle \cdot, \cdot \rangle$ is defined as

$$\langle (x, y(\cdot)), (x', y'(\cdot)) \rangle := x^T x' + \mathbb{E}[y(\xi)^T y'(\xi)].$$

(1.1)-(1.2) is said to be monotone if (2.4) is satisfied, while the monotonicity of H is defined under the Euclidean scalar product.

Let $\Theta : \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^d \to \mathbb{R}^n \times \mathbb{R}^m$ be

$$\Theta(x, y(\xi), \xi) = \begin{pmatrix} G(x, y(\xi), \xi) \\ F(x, y(\xi), \xi) \end{pmatrix}.$$

We make the following assumptions.

[1] $(x, y(\cdot)) \in D \times C(\cdot)$ if $(x, y(\xi)) \in D \times C(\xi)$ for a.e. $\xi \in \Xi$.

Assumption 2.5. *(i) For a.e.* $\xi \in \Xi$, $\Theta(\cdot,\cdot,\xi)$ *is monotone on* $D \times C(\xi)$.

(ii) For any $x \in D$ *and a.e.* $\xi \in \Xi$, $\nabla_y F(x, \hat{y}(x,\xi), \xi) v \in \mathrm{span}(\nabla_x F(x, \hat{y}(x,\xi), \xi))$ *for any* $v \in \mathcal{C}_x(\hat{y}; C(\xi), F)$, *where* $\mathrm{span}(\nabla_x F(x, \hat{y}(x,\xi), \xi))$ *denotes the subspace spanned by the columns of* $\nabla_x F(x, \hat{y}(x,\xi), \xi)$.

Following [10, Proposition 2.1], we can easily show that Assumption 2.5(i) implies the monotinicity of (1.1)-(1.2). To investigate the convergence of sample average approximation (SAA) for the two-stage SVI, the authors in [8] assumed the strong monotonicity of $\Theta(\cdot,\cdot,\xi)$ for a.e. $\xi \in \Xi$. The strong monotonicity assumption of $\Theta(\cdot,\cdot,\xi)$ implies Assumptions 2.4 and 2.5(i). If

$$F(x, y(\xi), \xi) = \tilde{F}(y(\xi), \xi) + N(\xi)x,$$

where $\mathrm{rank}(N(\xi)) = m$, then Assumption 2.5(ii) holds. The following proposition is an extension of [8, Theorem 3.9] which assumes that Θ is strongly monotone and $C(\xi) = \mathbb{R}^n_+$ for a.e. $\xi \in \Xi$.

Proposition 2.4. *Suppose that Assumptions 2.4 and 2.5 hold. Then* H *is monotone on* D.

Proof. It suffices to show that every element of $\partial \hat{G}_x(x,\xi)$ is positive semidefinite for any $x \in D$ and a.e. $\xi \in \Xi$. Under Assumption 2.5(i), for any $(x, y(\xi)) \in D \times C(\xi)$ and a.e. $\xi \in \Xi$, it holds

$$\begin{pmatrix} V_x(x,y(\xi),\xi) & V_y(x,y(\xi),\xi) \\ \nabla_x F(x,y(\xi),\xi) & \nabla_y F(x,y(\xi),\xi) \end{pmatrix} \succeq 0, \tag{2.5}$$

where $V_x(x,y(\xi),\xi) \in \partial_x G(x,y(\xi),\xi)$ and $V_y(x,y(\xi),\xi) \in \partial_y G(x,y(\xi),\xi)$.

If $\nabla_y F(x,y(\xi),\xi) = 0$, by (2.5), we have $V_y(x,y(\xi),\xi) = 0$, $\nabla_x F(x,y(\xi),\xi) = 0$ and $V_x(x,y(\xi),\xi) \succeq 0$. Hence every element of $\partial \hat{G}_x(x,\xi)$ is positive semidefinite. In the rest of the proof, we consider the case $\mathrm{rank}(\nabla_y F(x,y(\xi),\xi)) = r > 0$.

Define a set

$$\mathcal{Z}(x, y(\xi), \xi) = \{Z \in \mathbb{R}^{m\times l} : \ [Z^T \nabla_y F(x, y(\xi), \xi) Z] \text{ is nonsingular with } l = 1, \dots, r\}. \tag{2.6}$$

Let

$$U_Z(x, y(\xi), \xi) = -Z[Z^T \nabla_y F(x, y(\xi), \xi) Z]^{-1} Z^T \nabla_x F(x, y(\xi), \xi)$$

for a $Z \in \mathcal{Z}(x, y(\xi), \xi)$.

For any $u \in \mathbb{R}^n$, let $v = U_Z(x, y(\xi), \xi)u \in \mathbb{R}^m$. Then from (2.5), we have $u^T(V_x(x, y(\xi), \xi) + V_y(x, y(\xi), \xi)U_Z(x, y(\xi), \xi))u \geq 0$. Hence

$$V_x(x, y(\xi), \xi) + V_y(x, y(\xi), \xi)U_Z(x, y(\xi), \xi) \succeq 0. \tag{2.7}$$

Denote the linearity space of $\mathcal{C}_x(\hat{y}; C(\xi), F)$ by $\mathcal{L}_x(\hat{y}; C(\xi), F)$, which is defined as

$$\mathcal{L}_x(\hat{y}; C(\xi), F) = \mathcal{C}_x(\hat{y}; C(\xi), F) \cap -\mathcal{C}_x(\hat{y}; C(\xi), F).$$

Let $\bar{x} \in \Omega_{\hat{y}(\cdot,\xi)}$ be arbitrary with $\Omega_{\hat{y}(\cdot,\xi)}$ being the set of differentiable points of $\hat{y}(\cdot, \xi)$. It is obvious that $\mathcal{L}_{\bar{x}}(\hat{y}; C(\xi), F) \subseteq \mathcal{C}_{\bar{x}}(\hat{y}; C(\xi), F)$. Under Assumption 2.5(ii), it follows that for any $v \in \mathcal{C}_{\bar{x}}(\hat{y}; C(\xi), F)$, there exists a $d \in \mathbb{R}^n$ such that

$$\nabla_x F(\bar{x}, \hat{y}(\bar{x}, \xi), \xi)d + \nabla_y F(\bar{x}, \hat{y}(\bar{x}, \xi), \xi)v = 0,$$

which implies $v = \hat{y}'(\bar{x}, \xi; d) \in \mathcal{L}_{\bar{x}}(\hat{y}; C(\xi), F)$ by [27, Theorem 2.2 (3)]. Thus,

$$\mathcal{C}_{\bar{x}}(\hat{y}; C(\xi), F) = \mathcal{L}_{\bar{x}}(\hat{y}; C(\xi), F) \tag{2.8}$$

Then, for a.e. $\xi \in \Xi$, (2.8) holds for any $x \in \Omega_{\hat{y}(\cdot,\xi)}$. Therefore, by [27, Theorem 2.2] again, the Jacobian $\nabla_x \hat{y}(x, \xi)$ at any $x \in \Omega_{\hat{y}(\cdot,\xi)}$ can be represented as

$$\nabla_x \hat{y}(x, \xi) = U_Z(x, \hat{y}(x, \xi), \xi), \ \ Z \in \hat{\mathcal{Z}}(x, \hat{y}(x, \xi), \xi), \tag{2.9}$$

where $\hat{\mathcal{Z}}(x, \hat{y}(x, \xi), \xi)$ is a set consisting of matrices in $\mathbb{R}^{m\times l}$ with l being the dimension of $\mathcal{L}_x(\hat{y}; C(\xi), F)$, and each element $Z \in \hat{\mathcal{Z}}(x, \hat{y}(x, \xi), \xi)$ satisfies that $Z^T Z$ and

$Z^T \nabla_y F(x, \hat{y}(x,\xi), \xi) Z$ are nonsingular and $z \in \mathcal{L}_x(\hat{y}; C(\xi), F)$ if and only if $z = Zv$ for some $v \in \mathbb{R}^l$. It is clear that $\hat{\mathcal{Z}}(x, \hat{y}(x,\xi), \xi) \subseteq \mathcal{Z}(x, \hat{y}(x,\xi), \xi)$.

Let $\mathcal{B}(x)$ be an open neighborhood of $x \in D$. Since $\hat{G}(\cdot, \xi)$ and $\hat{y}(\cdot, \xi)$ are Lipschitz continuous, thus they are differentiable almost everywhere over $\mathcal{B}(x)$. Let $\hat{\Omega}_{\hat{y}}(x,\xi)$ and $\hat{\Omega}_{\hat{G}}(x,\xi)$ be the sets of differentiable points of $\hat{y}(\cdot,\xi)$ and $\hat{G}(\cdot,\xi)$ over the neighbourhood $\mathcal{B}(x)$, respectively. By the Lipschitz continuity of $G(\cdot,\cdot,\xi)$, we know that $\nabla G(x, \hat{y}(x,\xi), \xi)$ exists almost everywhere over $\mathcal{B}(x)$, and we denote this set by $\hat{\Omega}_G(x,\xi)$. Let $\hat{\Omega}(x,\xi) = \hat{\Omega}_{\hat{y}}(x,\xi) \cap \hat{\Omega}_{\hat{G}}(x,\xi) \cap \hat{\Omega}_G(x,\xi)$. It is clear that

$$\hat{\Omega}(x,\xi) \subseteq \hat{\Omega}_{\hat{y}}(x,\xi),\ \hat{\Omega}(x,\xi) \subseteq \hat{\Omega}_{\hat{G}}(x,\xi),\ \hat{\Omega}(x,\xi) \subseteq \hat{\Omega}_G(x,\xi),$$

and the measures of $\hat{\Omega}_{\hat{y}}(x,\xi) \setminus \hat{\Omega}(x,\xi)$, $\hat{\Omega}_{\hat{G}}(x,\xi) \setminus \hat{\Omega}(x,\xi)$ and $\hat{\Omega}_G(x,\xi) \setminus \hat{\Omega}(x,\xi)$ over the neighbourhood $\mathcal{B}(x)$ are all zero. Then, it follows that

$$\begin{aligned}
&\partial_x \hat{G}(x,\xi) \\
=&\mathrm{conv}\{\lim_{\bar{x}\to x} \nabla_x \hat{G}(\bar{x},\xi) : \bar{x} \in \hat{\Omega}_{\hat{G}}(x,\xi)\} \\
=&\mathrm{conv}\{\lim_{\bar{x}\to x} \nabla_x G(\bar{x}, \hat{y}(\bar{x},\xi), \xi) + \nabla_y G(\bar{x}, \hat{y}(\bar{x},\xi), \xi) \nabla_x \hat{y}(\bar{x},\xi) : \bar{x} \in \hat{\Omega}(x,\xi)\} \\
=&\mathrm{conv}\{\lim_{\bar{x}\to x} \nabla_x G(\bar{x}, \hat{y}(\bar{x},\xi), \xi) + \nabla_y G(\bar{x}, \hat{y}(\bar{x},\xi), \xi) U_{\bar{Z}}(\bar{x}, \hat{y}(\bar{x},\xi), \xi) : \\
&\qquad \bar{x} \in \hat{\Omega}(x,\xi), \bar{Z} \in \hat{\mathcal{Z}}(\bar{x}, \hat{y}(\bar{x},\xi), \xi)\} \\
\subseteq&\mathrm{conv}\{V_x(x, \hat{y}(x,\xi), \xi) + V_y(x, \hat{y}(x,\xi), \xi) U_Z(x, \hat{y}(x,\xi), \xi) : Z \in \mathcal{Z}(x, \hat{y}(x,\xi), \xi)\},
\end{aligned}$$

where the third equality is due to (2.9) and the last inclusion is due to $\hat{\mathcal{Z}}(x, \hat{y}(x,\xi), \xi) \subseteq \mathcal{Z}(x, \hat{y}(x,\xi), \xi)$. By (2.7), we know that for a.e. $\xi \in \Xi$, all elements in $\partial_x \hat{G}(x,\xi)$ are positive semidefinite for any $x \in D$, which implies the monotonicity of $\hat{G}(\cdot,\xi)$ on D for a.e. $\xi \in \Xi$. Therefore, we conclude that H is monotone on D. $\square$

2.1.2 The algorithm and convergence analysis

In this subsection, we propose the projection semismooth Newton algorithm (PSNA) for the discretization problem (1.3)-(1.4), which combines the semismooth Newton algorithm with the projection algorithm. The global convergence and superlinear convergence rate are established under suitable assumptions.

Note that the discretization problem (1.3)-(1.4) can be viewed as a two-stage SVI with the finite support set $\Xi_\nu = \{\xi_1, \dots, \xi_\nu\}$. Thus, all the results established in Subsection 2.1 hold without Assumptions 2.2 and 2.3. In this case, we have

$$H(x) = \mathbb{E}[\hat{G}(x,\xi)] = \sum_{\ell=1}^{\nu} p(\xi_\ell)\hat{G}(x,\xi_\ell), \quad \mathcal{H}(x) = \mathbb{E}[\partial\hat{G}(x,\xi)] = \sum_{\ell=1}^{\nu} p(\xi_\ell)\partial\hat{G}(x,\xi_\ell).$$

Define the residual function of (2.1) as

$$\hat{Q}(x) = x - \Pi_D(x - H(x)). \tag{2.10}$$

Proposition 1.5.8 in [19] claims that x^* solves (2.1) if and only if $\hat{Q}(x^*) = 0$. The function $\hat{Q}$ is also Lipschitz continuous since H is Lipschitz continuous and the nonexpansiveness of the projection operator. Let $L_{\hat{Q}}$ denote the Lipschitz constant of $\hat{Q}$.

The classic semismooth Newton algorithm (SNA) applying to (2.10) takes the form [36]

$$x^{k+1} = x^k - V_k^{-1}\hat{Q}(x^k), \tag{2.11}$$

where $V_k \in \partial\hat{Q}(x^k)$. Instead of using (2.11), we use a different semismooth Newton iteration to solve (2.1). We define a linear approximation of H and let the solution of the corresponding linear VI subproblem

$$-H(x^k) - (W^k + \epsilon_k I)(x - x^k) \in \mathcal{N}_D(x), \quad W^k \in \mathcal{H}(x^k), \tag{2.12}$$

be x^{k+1}, where $\epsilon_k > 0$ with $\epsilon_k \to 0$ as $k \to \infty$ is a regularized parameter forcing the linear VI (2.12) to be strongly monotone provided that W^k is positive semidefinite. (2.12) requires to solve a linear VI at each iteration while (2.11) solves a linear equation. But (2.12) is more desirable than (2.11) for solving (2.10). The reason is that $\partial \hat{Q}$ is difficult to calculate, since it is a composite function of two nonsmooth functions, the projection function Π_D and the implicit nonsmooth function H. On the other hand, by Proposition 2.3, we use $W^k \in \mathcal{H}(x^k)$ in (2.12) to avoid the difficulty of computing $\partial \hat{Q}(x^k)$. It seems that $\mathcal{H}(x^k)$ may still be difficult to calculate. We consider a special case of (1.1)-(1.2) in the next section in which one particular element of $\mathcal{H}(x^k)$ is easy to calculate.

Another main issue for Newton-type algorithms is that they are locally convergent in general. Since H is a nonsmooth and implicit function, the line search technique frequently used in Newton-type algorithms cannot be directly applied to our problem. Therefore, we turn to the projection algorithm to globalize the semismooth Newton iteration (2.12).

Define a projection operator

$$\tilde{\Pi}_{D,\alpha}(x) := \Pi_D[x - \alpha H(\pi(x))], \quad \text{with } \pi(x) = \Pi_D[x - \alpha H(x)], \tag{2.13}$$

where $\alpha > 0$ is the step size.

Under Assumptions 2.1 and 2.5, choosing $0 < \alpha < \frac{1}{L_H}$, by [19, Lemma 12.1.10] the projection operator $\tilde{\Pi}_{D,\alpha}$ is nonexpansive. Then, a natural fixed-point iteration is as follows

$$x^{k+1} = \tilde{\Pi}_{D,\alpha}(x^k).$$

It is shown in [19, Theorem 12.1.11] that $\{x^k\}$ generated by the above iteration globally converges to a fixed point x^* of $x = \tilde{\Pi}_{D,\alpha}(x)$ from any starting point $x^0 \in \mathbb{R}^n$, where x^* is also a solution of (2.1). However, the convergence rate is linear. To achieve a superlinear convergence rate, a hybrid algorithm with the semismooth Newton algorithm (2.12) is proposed in Algorithm 2.1.

Algorithm 2.1. *Projection Semismooth Newton Algorithm (PSNA)*

Step 0: *Choose an initial point $x^0 \in \mathbb{R}^n$, $\eta \in (0,1)$, step size $0 < \alpha < \frac{1}{L_H}$ and initial regularized parameter $\epsilon_0 > 0$. Set $k = 0$.*

Step 1: *For $\ell = 1, \ldots, \nu$, compute a Lipschitz continuous solution $\hat{y}(x^k, \xi_\ell)$ that solves the second stage problem* (1.4).

Step 2: *If $\|\hat{Q}(x^k)\| = 0$, stop. Otherwise, calculate a $W^k \in \mathcal{H}(x^k)$ and compute $\hat{x}^{k+1}$ that solves*

$$-H(x^k) - (W^k + \epsilon_k I)(x - x^k) \in \mathcal{N}_D(x). \tag{2.14}$$

If $\|\hat{Q}(\hat{x}^{k+1})\| \leq \eta \|\hat{Q}(x^k)\|$, let $x^{k+1} = \hat{x}^{k+1}$ and go to Step 4. Otherwise, go to Step 3.

Step 3: *Let $x^{k,0} = x^k$. Compute*

$$x^{k,j+1} = \tilde{\Pi}_{D,\alpha}(x^{k,j}), \quad j = 0, 1, \ldots, \tag{2.15}$$

until $\|\hat{Q}(x^{k,j+1})\| \leq \eta \|\hat{Q}(x^k)\|$ is satisfied. Set $x^{k+1} = x^{k,j+1}$.

Step 4: *Let $\epsilon_{k+1} = \min\{1, \|\hat{Q}(x^{k+1})\|\}$. Set $k := k+1$; go back to Step 1.*

Under Assumptions 2.4 and 2.5, any element of $\mathcal{H}(x)$ is positive semidefinite for any $x \in D$. Thus, the subproblem (2.14) is strongly monotone for any $\epsilon_k > 0$, which has a unique solution and is easy to solve. In Step 3, the projection iteration (2.15) is well-defined and is equivalent to solving a strongly convex program.

Lemma 2.1. *Under Assumptions 2.1, 2.4 and 2.5, for any x^k with $\|\hat{Q}(x^k)\| > 0$, there is a finite integer $J(x^k) > 0$ such that Step 3 of PSNA is terminated in finite times, i.e.,*

$$\|\hat{Q}(x^{k,j})\| \leq \eta \|\hat{Q}(x^k)\|, \ \forall j \geq J(x^k).$$

Proof. By [19, Theorem 12.1.11], we know that $\{x^{k,j}\}_{j=1}^{\infty}$ generated by (2.15) converges to a solution of (2.1), without loss of generality denoted by x^*. By the Lipschitz continuity of $\hat{Q}$, we have

$$\|\hat{Q}(x^{k,j})\| = \|\hat{Q}(x^{k,j}) - \hat{Q}(x^*)\| \leq L_{\hat{Q}} \|x^{k,j} - x^*\|.$$

Then, there exists a finite positive integer $J(x^k)$ such that the assertion of the lemma holds. □

The following standard assumption has been widely used in establishing the global convergence of Newton-type algorithms.

Assumption 2.6. *There exists a constant $\delta > 0$ such that the level set*

$$\mathcal{L}_0 = \{x \in D : \|\hat{Q}(x)\| \leq \delta\}$$

is bounded.

It is clear that if D is bounded, then $\mathcal{L}_0$ is bounded. By [19, Corollary 3.6.5(c)], Assumption 2.6 is satisfied if H is monotone and the solution set of (2.1) is nonempty and compact. Moreover, if D is a box, then H being a P_0 function with a bounded solution set can ensure Assumption 2.6.

Theorem 2.5. *Suppose that Assumptions 2.1, 2.4-2.6 hold. Let $\{x^k\}$ be an infinite sequence generated by PSNA. Then every accumulation point of $\{x^k\}$ is a solution of* (2.1). *In particular, if the Newton iteration is performed finite times, then $\{x^k\}$ converges to a solution of* (2.1).

Proof. Let

$$\mathcal{K} := \{k \ : \ \|\hat{Q}(\hat{x}^{k+1})\| \leq \eta\|\hat{Q}(x^k)\|, k \geq 0\}.$$

If $\mathcal{K}$ is finite, there exists an integer $\bar{k} > 0$ such that for all $k \geq \bar{k}$ the projection iteration (2.15) is always executed. By [19, Theorem 12.1.11], it follows that $\{x^k\}$ converges to a solution of (2.1).

If $\mathcal{K}$ is infinite, let $\mathcal{K}$ consist of $0 \leq k_0 < k_1 \cdots$. For any $k_{j+1}, k_j \in \mathcal{K}$, it follows that

$$\|\hat{Q}(x^{k_{j+1}})\| \leq \eta\|\hat{Q}(x^{k_{j+1}-1})\| \leq \ldots \leq \eta^{k_{j+1}-k_j}\|\hat{Q}(x^{k_j})\|,$$

which implies that $\lim_{j\to\infty, k_j\in\mathcal{K}} \|\hat{Q}(x^{k_j})\| = 0$. By the construction of the algorithm, it is easy to see that $\{x^k\} \in \mathcal{L}_0$ for sufficiently large k and $\lim_{k\to\infty} \|\hat{Q}(x^k)\| = 0$. Then, by the boundedness of $\{x^k\}$ and the continuity of $\hat{Q}$, for arbitrary convergent subsequence of $\{x^k\}$, we deduce that the limit point is a solution of (2.1). □

Next, we study the superlinear convergence rate of PSNA.

Theorem 2.6. *Suppose that assumptions in Theorem 2.5 hold and x^* is an accumulation point of $\{x^k\}$ generated by PSNA. If $G(\cdot,\cdot,\xi)$ is semismooth at x^* for any $\xi \in \Xi_\nu$, D is a polyhedral set, and all $W^* \in \mathcal{H}(x^*)$ are positive definite, then $\{x^k\}$ converges to x^* superlinearly.*

Proof. By Proposition 2.3, we know that H is semismooth at x^* and $\mathcal{H}$ is a linear Newton approximation scheme of H at x^*. Let $\mathcal{K}_0$ be the subsequence such that $\lim_{k\to\infty, k\in\mathcal{K}_0} x^k = x^*$. By Theorem 2.5, x^* is a solution of (2.1), or equivalently a zero of $\hat{Q}$.

The positive definiteness of all $W^* \in \mathcal{H}(x^*)$ implies that there exists a constant $\lambda > 0$ and a neighborhood $\mathcal{B}(x^*)$ of x^* such that for all $x \in \mathcal{B}(x^*)$, all $W \in \mathcal{H}(x)$ are positive definite with $v^TWv \geq \frac{1}{2}\lambda\|v\|^2, \forall v \in \mathbb{R}^n$. This implies that H is strongly monotone around x^*, and x^* is an isolated zero of $\hat{Q}$. Let $W^k_{\epsilon_k} = W^k + \epsilon_k I$. For all sufficiently large $k \in \mathcal{K}_0$, $x^k \in \mathcal{B}(x^*)$. Thus, the subproblem (2.14) has a unique solution, denoted by $\hat{x}^{k+1}$. For all $k \in \mathcal{K}_0$ sufficiently large, we have $x^k \in \mathcal{B}(x^*)$. Thus, the subproblem (2.14) has a unique solution, denoted by $\hat{x}^{k+1}$. Since

$$-H(x^k) - W^k_{\epsilon_k}(\hat{x}^{k+1} - x^k) \in \mathcal{N}_D(\hat{x}^{k+1}), \quad -H(x^*) \in \mathcal{N}_D(x^*),$$

by the definition of normal cone and $\hat{x}^{k+1}, x^* \in D$, we have

$$(H(x^k) + W^k_{\epsilon_k}(\hat{x}^{k+1} - x^k))^T(x^* - \hat{x}^{k+1}) \geq 0, \quad H(x^*)^T(\hat{x}^{k+1} - x^*) \geq 0,$$

which implies that

$$
\begin{aligned}
& 0 \leq [H(x^k) + W_{\epsilon_k}^k(\hat{x}^{k+1} - x^k) - H(x^*)]^T(x^* - \hat{x}^{k+1}) \\
\Leftrightarrow\ & 0 \leq [H(x^k) - H(x^*) + W_{\epsilon_k}^k(\hat{x}^{k+1} - x^* + x^* - x^k)]^T(x^* - \hat{x}^{k+1}) \\
\Leftrightarrow\ & (\hat{x}^{k+1} - x^*)^T W_{\epsilon_k}^k(\hat{x}^{k+1} - x^*) \leq [H(x^k) - H(x^*) + W_{\epsilon_k}^k(x^* - x^k)]^T(x^* - \hat{x}^{k+1}) \\
\Rightarrow\ & \frac{1}{2}\lambda\|\hat{x}^{k+1} - x^*\|^2 \leq (\|H(x^k) - H(x^*) - W^k(x^k - x^*)\| \\
& + \epsilon_k\|x^k - x^*\|)\|\hat{x}^{k+1} - x^*\| \\
\Rightarrow\ & \|\hat{x}^{k+1} - x^*\| \leq o(\|x^k - x^*\|), \qquad (2.16)
\end{aligned}
$$

where the last inequality is due to (2.2) and $\epsilon_k \to 0$. Next, we will prove that for all k sufficiently large,

$$\|\hat{Q}(\hat{x}^{k+1})\| = o(\|\hat{Q}(x^k)\|). \qquad (2.17)$$

By (2.16), we have

$$\|\hat{x}^{k+1} - x^k\| = \|x^k - x^*\| + o(\|x^k - x^*\|).$$

Since H is strongly monotone around x^* and is Lipschitz continuous, by [19, Theorem 2.3.3], there exists a positive constant $c' > 0$ such that

$$\|x^k - x^*\| \leq c'\|\hat{Q}(x^k)\|.$$

The last two inequalities imply that

$$\|\hat{x}^{k+1} - x^k\| \leq c'\|\hat{Q}(x^k)\|, \qquad (2.18)$$

(2.16) also implies that

$$\|\hat{x}^{k+1} - x^*\| \leq \varepsilon\|x^k - x^*\|, \qquad (2.19)$$

where $\varepsilon > 0$ is arbitrarily small as $k \to \infty$. Since H is semismooth at x^* and D is polyhedral, then $\hat{Q}$ is semismooth at x^* and directionally differentiable at x^* by [19, Theorem 4.1.1]. Since $\hat{Q}$ is directionally differentiable at x^* and Lipschitz continuous, by [36], we have

$$\|\hat{Q}(\hat{x}^{k+1}) - \hat{Q}(x^*) - \hat{Q}'(x^*; \hat{x}^{k+1} - x^*)\| \leq \varepsilon \|\hat{x}^{k+1} - x^*\|,$$

which means

$$\|\hat{Q}'(x^*; \hat{x}^{k+1} - x^*)\| \leq (L_{\hat{Q}} + \varepsilon)\|\hat{x}^{k+1} - x^*\|.$$

By the last three inequalities, we have

$$\begin{aligned}\|\hat{Q}(\hat{x}^{k+1})\| \leq & \|\hat{Q}'(x^*; \hat{x}^{k+1} - x^*)\| + \varepsilon\|\hat{x}^{k+1} - x^*\| \\ \leq & (L_{\hat{Q}} + 2\varepsilon)\|\hat{x}^{k+1} - x^*\| \\ \leq & (L_{\hat{Q}} + 2\varepsilon)\varepsilon\|x^k - x^*\|. \end{aligned} \tag{2.20}$$

From (2.18) and (2.19), it follows

$$\begin{aligned}\|x^k - x^*\| \leq & \|\hat{x}^{k+1} - x^k\| + \|\hat{x}^{k+1} - x^*\| \\ \leq & c'\|\hat{Q}(x^k)\| + \varepsilon\|x^k - x^*\|, \end{aligned}$$

which implies that

$$\|x^k - x^*\| \leq \frac{c'}{1 - \varepsilon}\|\hat{Q}(x^k)\|. \tag{2.21}$$

Combining (2.20) with (2.21), it holds that

$$\|\hat{Q}(\hat{x}^{k+1})\| \leq \frac{(L_{\hat{Q}} + 2\varepsilon)\varepsilon c'}{1 - \varepsilon}\|\hat{Q}(x^k)\|.$$

Since ε can be arbitrarily small when k is sufficiently large, the last inequality implies (2.17). This means that $\hat{x}^{k+1}$ computed from the Newton iteration (2.14) is always accepted when x^k is sufficiently close to x^*. Then, $x^{k+1} = \hat{x}^{k+1}$. Therefore, (2.16) becomes

$$\|x^{k+1} - x^*\| \leq o(\|x^k - x^*\|),$$

which means that x^k converges to x^* superlinearly. □

Remark 2.1. *The assumption that D is a polyhedron in Theorem 2.6 can be extended to $D = \{x \in R^n : g(x) \leq 0\}$ where g is twice continuously differentiable and convex with the constant rank constraint qualification at x^*. From [19, Theorem 4.5.2], $\hat{Q}$ is piecewise smooth around x^* in such case.*

2.2 A two-stage semi-linear SVI

In this section, we apply PSNA to solve a two-stage semi-linear SVI, which is a special case of (1.1)-(1.2) as follows:

$$-A(x) - \mathbb{E}[B(\xi)y(\xi)] \in \mathcal{N}_D(x), \tag{2.22}$$

$$0 \leq y(\xi) \perp M(\xi)y(\xi) + N(\xi)x + q(\xi) \geq 0, \qquad \text{for all } \xi \in \Xi_\nu, \tag{2.23}$$

where the function $A : \mathcal{D} \supset D \to \mathbb{R}^n$ is is semismooth and Lipschitz continuous on an open set $\mathcal{D}$ with Lipschitz constant L_A, $B : \mathbb{R}^d \to \mathbb{R}^{n\times m}$, $M : \mathbb{R}^d \to \mathbb{R}^{m\times m}$, $N : \mathbb{R}^d \to \mathbb{R}^{m\times n}$ and $q : \mathbb{R}^d \to \mathbb{R}^m$.

Under proper assumptions, we show that the Lipschitz continuity, semismoothness, linear Newton approximation scheme and monotonicity properties for the single-stage problem (2.1) hold, which are important to establish the global convergence and superlinear convergence rate of PSNA.

Problem (2.22)-(2.23) can be written in a compact form as follows:

$$-A(x) - \sum_{\ell=1}^{\nu} p_\ell B_\ell y_\ell \in \mathcal{N}_D(x), \tag{2.24}$$

$$0 \leq \begin{pmatrix} y_1 \\ \vdots \\ y_\nu \end{pmatrix} \perp \begin{pmatrix} N_1 & M_1 & \cdots & 0 \\ \vdots & 0 & \ddots & 0 \\ N_\nu & 0 & 0 & M_\nu \end{pmatrix} \begin{pmatrix} x \\ y_1 \\ \vdots \\ y_\nu \end{pmatrix} + \begin{pmatrix} q_1 \\ \vdots \\ q_\nu \end{pmatrix} \geq 0, \tag{2.25}$$

where $y_\ell = y(\xi_\ell)$, $q_\ell = q(\xi_\ell)$, $M_\ell = M(\xi_\ell)$, $B_\ell = B(\xi_\ell)$ and $N_\ell = N(\xi_\ell)$, $\ell = 1, 2, \ldots, \nu$.

Assumption 2.7. *(i) M_ℓ is a P-matrix for all ℓ; or,*

(ii) M_ℓ is a Z-matrix for all ℓ, and (2.25) is feasible for any $x \in \mathcal{D}$ and $\xi_\ell \in \Xi_\nu$.

Lemma 2.2. *For any fixed $x \in \mathcal{D}$ and $\xi_\ell \in \Xi_\nu$, the second stage problem (2.25) has a unique solution (or a unique least-element solution[2]) $\hat{y}(x, \xi_\ell)$ if Assumption 2.7 (i) (or Assumption 2.7 (ii)) holds, which reads*

$$\hat{y}(x, \xi_\ell) = -U(x, \xi_\ell)(N_\ell x + q_\ell), \tag{2.26}$$

with

$$U(x, \xi_\ell) := (I - \Lambda(x, \xi_\ell) + \Lambda(x, \xi_\ell)M_\ell)^{-1}\Lambda(x, \xi_\ell), \tag{2.27}$$

where $\Lambda(x, \xi_\ell)$ is a diagonal matrix with

$$\Lambda(x, \xi_\ell)_{ii} = \begin{cases} 1, & \text{if } (M_\ell \hat{y}(x, \xi_\ell) + N_\ell x + q_\ell)_i < (\hat{y}(x, \xi_\ell))_i, \\ 0, & \text{otherwise.} \end{cases}$$

Moreover, $\hat{y}(\cdot, \xi_\ell)$ is piecewise affine, strongly semismooth[3] and globally Lipschitz continuous on $\mathcal{D}$ with Lipschitz constant written as

$$L_\ell := \|N_\ell\| \max\{\|(M_\ell)^{-1}_{JJ}\| : (M_\ell)_{JJ} \text{ is nonsingular for } J \subseteq [m]\},$$

and

$$-U(x, \xi_\ell)N_\ell \in \partial \hat{y}(x, \xi_\ell). \tag{2.28}$$

Proof. When M_ℓ is a P-matrix, for any given (x, ξ_ℓ) the existence and uniqueness of $\hat{y}(x, \xi_\ell)$ are due to [17, Theorem 3.3.7]. When M_ℓ is a Z-matrix and the LCP$(N_\ell x + q_\ell, M_\ell)$ is feasible for all $x \in D$, the existence of the unique least-element solution follows from [17, Theorem 3.11.6]. The expression (2.26) follows from Lemma 2.1

[2] A solution y^* of the LCP(q, M) is called the least-element solution if $y^* \leq y$ (componentwise) for any $y \in \mathrm{SOL}(q, M)$, and the least-element solution can be computed by solving a linear program [17].

[3] A function K is called strongly semismooth at x if $\limsup_{\substack{x+h\in\Omega_K,\\ h\to 0}} \|K'(x+h; h) - K'(x; h)\|/\|h\|^2 < \infty$; see [35].

and Theorem 2.2 in [14]. It is clear that $\hat{y}(\cdot, \xi)$ is piecewise affine from the expression (2.26). According to [19, Proposition 7.4.7], every piecewise affine function is strongly semismooth.

When M_ℓ is a P-matrix or a Z-matrix, the Lipschitz continuity property of $\hat{y}(\cdot, \xi_\ell)$ follows from [14, Corollary 2.1] and [14, Theorem 2.3], respectively.

The generalized Jacobian (2.28) is due to [14, Theorem 3.1]. □

As in the last section, substituting the Lipschitz continuous function $\hat{y}(x, \xi_\ell)$ into (2.24), we get the single-stage SVI formulation (2.1), in which

$$H(x) = A(x) + \mathbf{B}_\nu \hat{\mathbf{y}}_\nu(x), \tag{2.29}$$

where

$$\mathbf{B}_\nu = (p_1 B_1, \ldots, p_\nu B_\nu) \in \mathbb{R}^{n\times \nu m}, \;\; \hat{\mathbf{y}}_\nu(x) = (\hat{y}^T(x, \xi_1), \ldots, \hat{y}^T(x, \xi_\nu))^T \in \mathbb{R}^{\nu m}$$

with $\hat{y}(x, \xi_\ell) \in \mathrm{SOL}(N_\ell x + q_\ell, M_\ell),\ \ell = 1, \ldots, \nu$. Moreover, H is Lipschitz continuous on $\mathcal{D}$ with Lipschitz constant

$$L_H := L_A + \bar{\sigma}, \quad \text{where } \bar{\sigma} := \sum_{\ell=1}^{\nu} p_\ell \|B_\ell\| L_\ell. \tag{2.30}$$

In addition, the corresponding residual function $\hat{Q}$ is Lipschitz continuous on $\mathcal{D}$. Under Assumption 2.7(i), as discussed in Proposition 2.1, (2.29) is an equivalent formulation to (2.24)-(2.25). Under Assumption 2.7(ii), if D is bounded, then (2.29) is solvable. Thus, if x^* solves (2.29), then $(x^*, \hat{y}(x^*, \xi_1), \ldots, \hat{y}(x^*, \xi_\nu))$ is a solution to (2.24)-(2.25).

Denote

$$\Theta(x, y_\ell, \xi_\ell) = \begin{pmatrix} A(x) + B_\ell y_\ell \\ N_\ell x + M_\ell y_\ell + q_\ell \end{pmatrix}.$$

To establish the monotonicity of H, we further make the following assumption.

Assumption 2.8. $\Theta(\cdot,\cdot,\xi_\ell)$ *is monotone on* $\mathcal{D}\times\mathbb{R}^m$ *for* $\ell=1,\ldots,\nu$.

The above assumption implies that all generalized Jacobian matrices in $\partial\Theta(x,y_\ell,\xi_\ell)$ are positive semidefinite for any $(x,y_\ell)\in\mathcal{D}\times\mathbb{R}^n$; that is,

$$\begin{pmatrix} V_A & B_\ell \\ N_\ell & M_\ell \end{pmatrix} \text{ is positive semidefinite for any } V_A\in\partial A(x). \tag{2.31}$$

Notice that Assumption 2.8 is a sufficient condition for (2.22)-(2.23) being monotone in the sense of (2.4). Assumptions 2.7-2.8 together mean that each M_ℓ should be a positive semidefinite P-matrix or a positive semidefinite Z-matrix.

Proposition 2.7. *Under Assumptions 2.7-2.8,* H *is monotone on* $\mathcal{D}$.

Proof. It follows from Proposition 2.4. We omit the details. □

Remark 2.2. *The monotonicity of* H *does not imply that the original problem* (2.24)-(2.25) *is monotone in the sense of* (2.4). *For example, without Assumption 2.8, the monotonicity of* H *can be guaranteed if* A *is strongly monotone on* D *such that*

$$(x-x')^T(A(x)-A(x'))\geq\bar{\sigma}\|x-x'\|^2,\quad\forall\ x,x'\in D, \tag{2.32}$$

where $\bar{\sigma}$ *is defined in* (2.30). *Note that this assumption does not mean that* (2.24)-(2.25) *is monotone.*

Let

$$\mathcal{H}(x):=\partial A(x)+\sum_{\ell=1}^{\nu}p_\ell\partial B_\ell\hat{y}(x,\xi_\ell).$$

We can show that, by Proposition 2.3, $\mathcal{H}$ is a linear Newton approximation scheme of H. By Lemma 2.2, one particular element of $\mathcal{H}(x)$ can be calculated by

$$V_A-\mathbf{B}_\nu\mathbf{U}_\nu(x)\in\mathcal{H}(x),\quad V_A\in\partial A(x), \tag{2.33}$$

where $\mathbf{U}_\nu(x)=((U(x,\xi_1)N_1)^T,\ldots,(U(x,\xi_\nu)N_\nu)^T)^T$ with $U(x,\xi_\ell)$ defined in (2.27).

Now, we discuss the implementation of PSNA for solving (2.1).

- **Step 1** calculates $\hat{y}(x^k, \xi_\ell)$ for each subproblem LCP$(N_\ell x^k + q_\ell, M_\ell)$ in parallel.

- In **Step 2**, by (2.33), an element of $W^k \in \mathcal{H}(x^k)$ can be computed by

$$W^k = V_A^k - \sum_{\ell=1}^{\nu} p_\ell B_\ell U(x^k, \xi_\ell) N_\ell, \quad V_A^k \in \partial A(x^k).$$

  Subproblem (2.14) is strongly monotone and has a unique solution since $W^k + \epsilon_k I$ is positive definite.

- In **Step 3**, the projection iteration (2.15) is implementable. Note that the execution of (2.15) requires the knowledge of H, in which its function value at x^k can be calculated by solving all the second stage LCP$(N_\ell x + q_\ell, M_\ell)$, $\ell = 1, \ldots, \nu$ and letting $H(x^k) = A(x^k) + \mathbf{B}_\nu \hat{\mathbf{y}}_\nu(x^k)$. The projection onto D is equivalent to solving a strongly convex optimization problem.

By the same argument as in Theorems 2.5 and 2.6, we can prove the global convergence and superlinear convergence rate of PSNA for solving (2.24)-(2.25).

We study the superlinear convergence rate of PSNA for $D = [l, u]$, where $l \in \{\mathbb{R} \cup -\infty\}^n$ and $u \in \{\mathbb{R} \cup \infty\}^n$ with $l < u$. In this case, $\hat{Q}(x)$ is reduced to

$$\hat{Q}(x) = \mathrm{mid}(x - l, x - u, H(x)),$$

where mid is a function such that

$$\mathrm{mid}(l, u, x)_i = \Pi_{[l_i, u_i]}(x_i) = \begin{cases} l_i & \text{if } l_i > x_i \\ x_i & \text{if } l_i \le x_i \le u_i \\ u_i & \text{if } x_i > u_i. \end{cases}$$

The following lemma is useful.

Lemma 2.3. *Suppose that all $W \in \mathcal{H}(x)$ are P-matrices. Then, there exists a neighborhood of x such that for any $\bar{x}$ in this neighborhood, all $\bar{W} \in \mathcal{H}(\bar{x})$ are P-matrices. Moreover, there exists a positive constant β such that $\|(I - \Lambda + \Lambda \bar{W})^{-1}\| \le \beta$ for any diagonal matrix Λ with diagonal entries on $[0, 1]$.*

Proof. Since all $W \in \mathcal{H}(x)$ are P-matrices and thus nonsingular, by the same argument in [36, Proposition 3.1], there exists a neighborhood $\mathcal{B}(x)$ of x such that for any $\bar{x} \in \mathcal{B}(x)$, all $\bar{W} \in \mathcal{H}(\bar{x})$ are nonsingular and $\|\bar{W}^{-1}\| \leq \beta$. On the other hand, [22, Theorem 4.3] claims that $\bar{W}$ is a P-matrix if and only if $I - \Lambda + \Lambda\bar{W}$ is nonsingular for any diagonal matrix Λ with $\Lambda_{ii} \in [0, 1]$.

Assume that the conclusion is not true. Then, by the above discussion, there exists a sequence $x^k \to x$, $W^k \in \mathcal{H}(x^k)$ such that either all W^k are nonsingular but not P-matrices or $\|(I - \Lambda_k + \Lambda_k W^k)^{-1}\| \to \infty$ for some Λ_k. Since $\mathcal{H}$ is bounded in a neighbourhood of x, taking a subsequence if necessary, we assume that $\lim_{k\to\infty} W^k \to \tilde{W}$, where $\tilde{W}$ is not a P-matrix. By the closedness of $\mathcal{H}$ at x, it follows that $\tilde{W} \in \mathcal{H}(x)$, which is a contradiction. $\square$

The following theorem establishes the superlinear convergence rate of PSNA under weaker assumptions; that is, the positive definiteness for the elements of $\mathcal{H}(x^*)$ is relaxed to the P-matrix.

Theorem 2.8. *Suppose that Assumptions 2.7(i) and 2.8 hold, the level set $\mathcal{L}_0$ is bounded, and $D = [l, u]$; or Assumptions 2.7(ii) and 2.8 hold, and $D = [l, u]$ is a bounded box. Assume that x^* is an accumulation point of sequence $\{x^k\}$ generated by PSNA, and all $W^* \in \mathcal{H}(x^*)$ are P-matrices. Then, $\{x^k\}$ converges to x^* superlinearly.*

Proof. By Theorem 2.5, there exists a subsequence $\mathcal{K}_0 \subseteq \mathcal{K}$ such that

$$\lim_{k\to\infty, k\in\mathcal{K}_0} x^k = x^* \text{ with } x^* \text{ being a solution.}$$

Since all $W^* \in \mathcal{H}(x^*)$ are P-matrices, by Lemma 2.3, there exists a neighborhood of x^*, denoted by $\mathcal{B}(x^*)$, such that for any $x \in \mathcal{B}(x^*)$, any $W \in \mathcal{H}(x)$ is a P-matrix. When $k \in \mathcal{K}_0$ is sufficiently large, we have $x^k \in \mathcal{B}(x^*)$. Then, all $W^k \in \mathcal{H}(x^k)$ are P-matrices. Hence, (2.14) has a unique solution $\hat{x}^{k+1}$ for any $\epsilon_k > 0$; that is

$$-H(x^k) - W^k_{\epsilon_k}(\hat{x}^{k+1} - x^k) \in \mathcal{N}_{[l,u]}(\hat{x}^{k+1}), \quad \text{with } W^k_{\epsilon_k} = W^k + \epsilon_k I,$$

which can be rewritten as

$$\tilde{Q}(\hat{x}^{k+1}) := \mathrm{mid}\left(\hat{x}^{k+1} - l, \hat{x}^{k+1} - u, H(x^k) + W^k_{\epsilon_k}(\hat{x}^{k+1} - x^k)\right) = 0.$$

Similarly, since x^* is a solution, we have

$$\hat{Q}(x^*) = \mathrm{mid}\,(x^* - l, x^* - u, H(x^*)) = 0.$$

From [11, Lemma 2.1], there exists a diagonal matrix Λ_k with diagonal entries on $[0, 1]$ such that

$$\begin{aligned} 0 &= \tilde{Q}(\hat{x}^{k+1}) - \hat{Q}(x^*) \\ &= (I - \Lambda_k)(\hat{x}^{k+1} - x^*) + \Lambda_k[H(x^k) + W^k_{\epsilon_k}(\hat{x}^{k+1} - x^k) - H(x^*)] \\ &= (I - \Lambda_k)(\hat{x}^{k+1} - x^*) + \Lambda_k[H(x^k) + W^k_{\epsilon_k}(\hat{x}^{k+1} - x^* + x^* - x^k) - H(x^*)]. \end{aligned} \tag{2.34}$$

The matrix $I - \Lambda_k + \Lambda_k W^k_{\epsilon_k}$ is nonsingular since $W^k_{\epsilon_k}$ is a P-matrix. When the function A is semismooth at x^*, arranging (2.34), we get

$$\begin{aligned} \|\hat{x}^{k+1} - x^*\| &= \|(I - \Lambda_k + \Lambda_k W^k_{\epsilon_k})^{-1}\Lambda_k[H(x^k) - H(x^*) - W^k_{\epsilon_k}(x^k - x^*)]\| \\ &\le \|(I - \Lambda_k + \Lambda_k W^k_{\epsilon_k})^{-1}\Lambda_k\|\|H(x^k) - H(x^*) - W^k_{\epsilon_k}(x^k - x^*)\| \\ &\le \|(I - \Lambda_k + \Lambda_k W^k_{\epsilon_k})^{-1}\Lambda_k\|(\|H(x^k) - H(x^*) - W^k(x^k - x^*)\| \\ &\quad + \epsilon_k\|(x^k - x^*)\|) \\ &= o(\|x^k - x^*\|), \end{aligned} \tag{2.35}$$

where the the last equality is due to (2.2), Lemma 2.3 and $\epsilon_k \to 0$.

There exists a diagonal matrix $\tilde{\Lambda}_k$ with diagonal entries on $[0, 1]$ such that

$$\begin{aligned} \hat{Q}(x^k) &= \hat{Q}(x^k) - \tilde{Q}(\hat{x}^{k+1}) \\ &= (I - \tilde{\Lambda}_k)(x^k - \hat{x}^{k+1}) + \tilde{\Lambda}_k W^k_{\epsilon_k}(x_k - \hat{x}_{k+1}) \\ &= (I - \tilde{\Lambda}_k + \tilde{\Lambda}_k W^k_{\epsilon_k})(x_k - \hat{x}_{k+1}), \end{aligned}$$

which implies that

$$\|\hat{x}^{k+1} - x^k\| \le \|(I - \tilde{\Lambda}_k + \tilde{\Lambda}_k W^k_{\epsilon_k})^{-1}\|\|\hat{Q}(x^k)\| \le \beta\|\hat{Q}(x^k)\|.$$

[19, Proposition 7.4.6] shows that a piecewise semismooth function is also semismooth. Since H is semismooth at x^*, $\hat{Q}(x) = \mathrm{mid}(x - l, x - u, H(x))$ is also semismooth at x^*. By the same argument of Theorem 2.6, we can prove (2.17). This implies that $\hat{x}^{k+1}$ computed from Newton iteration (2.14) is always accepted when x^k is sufficiently close to x^*; that is $x^{k+1} = \hat{x}^{k+1}$. Therefore, (2.35) means that x^k converges to x^* superlinearly. □

Corollary 2.1. *Let D be a polyhedron. The sequence $\{x^k\}$ generated by PSNA globally and superlinearly converges to the unique solution of* (2.1) *if one of the following conditions holds; that is,*

(i) $\Theta(\cdot, \cdot, \xi)$ is strongly monotone on $\mathcal{D} \times \mathbb{R}^m$ for any $\xi \in \Xi_\nu$;

(ii) (2.32) *holds with strict inequality and M_ℓ is a P-matrix for any $\xi_\ell \in \Xi_\nu$;*

(iii) (2.32) *holds with strict inequality, M_ℓ is a Z-matrix for any $\xi_\ell \in \Xi_\nu$,* (2.22)-(2.23) *has the relatively complete recourse and D is bounded.*

2.3 Numerical results

In this section, we conduct numerical experiments to test the efficiency of PSNA for the two-stage SVI (2.24)-(2.25), and compare with PHA.

PSNA is terminated if

$$\mathrm{Res} := \|\hat{Q}(x^k)\| \leq 10^{-6}$$

or

$$\|x^k - x^{k-1}\| \leq 10^{-6}.$$

The starting point $x^0 \in \mathbb{R}^n_+$ is randomly chosen, $\alpha = 0.015$ and $\eta = 0.9$. The regularized parameter ϵ_k is set to $\min\{1, \|\hat{Q}(x^k)\|\}$. All codes were implemented in MATLAB R2016b on a desktop with Intel Core i7-4790 (3.5 GHz) and 32 GB RAM.

Example 2.1. *Monotone two-stage SLCP in* [40].[4]

In this example, the first stage problem is an LCP with $A(x) = \tilde{A}x + c$. *Let* $s = \lceil 3(n+m)/4 \rceil$, *and randomly generate positive numbers* α_i *and vectors* $(a_i^T, b_i^T)^T \in \mathbb{R}^{n+m}$ *for* $i = 1, \dots, s$. *For* $\ell = 1, \dots, \nu$, *randomly create* ν *antisymmetric matrices* $O_\ell \in \mathbb{R}^{(n+m)\times(n+m)}$. *Set*

$$\begin{pmatrix} \tilde{A} & B_\ell \\ N_\ell & M_\ell \end{pmatrix} = \sum_{i=1}^{s} \alpha_i \begin{pmatrix} a_i \\ b_i \end{pmatrix} \begin{pmatrix} a_i^T & b_i^T \end{pmatrix} + \begin{pmatrix} 0 & (O_{12})_\ell \\ (O_{21})_\ell & (O_{22})_\ell \end{pmatrix}.$$

Randomly generate c, *and* q_ℓ *for* $\ell = 1, \dots, \nu$.

Example 2.2. *Nonmonotone two-stage SVI with P-matrix LCP in the second stage.*

In this example, the first stage problem is a box affine VI, while the second stage problem is a P-matrix LCP for any fixed $x \in \mathbb{R}^n$ *and* ξ. *Set* $A(x) = \tilde{A}x + c$. *Generate* $\bar{A} \in \mathbb{R}^{n\times n}$, $\bar{U} \in \mathbb{R}^{n\times n}$, $c \in \mathbb{R}^n$, *and* $B_\ell \in \mathbb{R}^{n\times m}$, $\bar{N}_\ell \in \mathbb{R}^{m\times n}$, $U_\ell \in \mathbb{R}^{m\times m}$, $q_\ell \in \mathbb{R}^m$ *for* $\nu = 1, \dots, \nu$, *with entries uniformly distributed on* $[-5, 5]$, *where* U_ℓ *is strictly upper triangular. Create the diagonal matrix* $\bar{\Lambda} \in \mathbb{R}^{n\times n}$ *with entries uniformly distributed on* $[0, 0.3]$, *and* ν *diagonal matrices* $\Lambda_\ell \in \mathbb{R}^{m\times m}$ *with entries from* $[5, 10]$. *Following Harker and Pang [24], we set*

$$\tilde{A} = \bar{A}^T \bar{A} + \bar{\Lambda} + (\bar{U} - \bar{U}^T).$$

The second stage problem is as follows

$$0 \le y_\ell \perp M_\ell y_\ell + \bar{N}_\ell f(x) + q_\ell \ge 0, \ \ \ell = 1, \dots, \nu,$$

with $M_\ell = \Lambda_\ell + U_\ell$, $f_i(x) = \sin x_i$.

Example 2.3. *Nonmonotone two-stage mixed SLCP with Z-matrix LCP in the second stage.*

[4] For this example, PSNA is applied to solve the regularized problem in which M_ℓ is replaced by $M_\ell + \mu_k I$ for each ℓ with $\mu_k = 10^{-9}$.

All parameters are generated in a same way as Example 2.2 except for the settings of D, M_ℓ, N_ℓ and q_ℓ. The set $D = [0, ne_n]$ is an n-dimensional bounded box. Let $m = 2k$ be even with k being a positive integer. All entries of k-th row and (k+1)-th row of $\bar{N} \in \mathbb{R}^{m\times n}$ are set to 1 and -1, respectively, while all other entries are zero. $\bar{M} \in \mathbb{R}^{m\times m}$ is a tridiagonal matrix with -1, 2, -1 on its superdiagonal, main diagonal and subdiagonal, respectively, except for $\bar{M}_{mm} = \bar{M}_{11} = 1$. $q_k = \tilde{q}$ and $q_{k+1} = -\tilde{q}$ with $\tilde{q}$ uniformly drawn from $[-5, 5]$, and other components of q are zero. Generate an i.i.d. sample $\{\xi_1, \ldots, \xi_\nu\}$ of random variable $\xi \in \mathbb{R}$ following uniformly distribution on $[1, 5]$. Set

$$M_\ell = \xi_\ell \bar{M},\ N_\ell = (\xi_\ell + 1)\bar{N},\ q_\ell = (\xi_\ell + 2)q,\ \ell = 1, \ldots, \nu.$$

It is not hard to verify that the LCP$(N_\ell x + q_\ell, M_\ell)$ is feasible for any $x \in D$ and ξ_ℓ, and hence it admits a unique least-element solution. For example, $y = (y_1, \ldots, y_k, y_{k+1}, \ldots, y_{2k})^T$ with $y_1 = \ldots = y_k = 0$ and $y_{k+1} = \cdots = y_{2k} = [(\xi_\ell + 1)\sum_{i=1}^n x_i + (\xi_\ell + 2)\tilde{q}]/\xi_\ell$ is a feasible point of the LCP$(N_\ell x + q_\ell, M_\ell)$.

Example 2.4. *Nonmonotone and nonsmooth two-stage semi-linear SVI with P-matrix LCP in the second stage.*

In this case, $D = [0, ne_n]$. All other parameters are the same as that of Example 2.2 except for $A(x)$, which is of the following form:

$$A_1(x) = x_1^2 + \sum_{i=2}^{n-1}(x_i x_{i+1}) - \sum_{i=2}^{n} x_i + |x_1 - 1| + 3x_1,$$

$$A_2(x) = x_1(1 - x_3) + x_2^2 + |x_2 - 2| + 3x_2,$$

$$A_i(x) = x_1(1 - x_{i-1} - x_{i+1}) + x_i^2 + |x_i - i| + 3x_i,\ i = 3, \ldots, n-1,$$

$$A_n(x) = x_1(1 - x_{n-1}) + x_n^2 + |x_n - n| + 3x_n,$$

$$A(x) = (A_1(x), \ldots, A_n(x))^T + c.$$

The function A is nonsmooth but semismooth at x with $x_i = i$ and any element of $\partial A(x)$ is positive definite for any $x \in \mathbb{R}^n_+$. We generate $c \in \mathbb{R}^n$ in a way such that there is a solution x^* of 20%, 40%, 60% and 80% components being nonsmooth, respectively; that is, the corresponding component x_i^* equals i. The remaining components are set to 0 or n on a fifty-fifty basis, respectively.

We compared our algorithm with PHA for solving Example 2.1, which is a monotone problem and also tested in [40]. Parameters of PHA are set as given in [40]. Examples 2.2-2.4 cannot be solved by PHA.

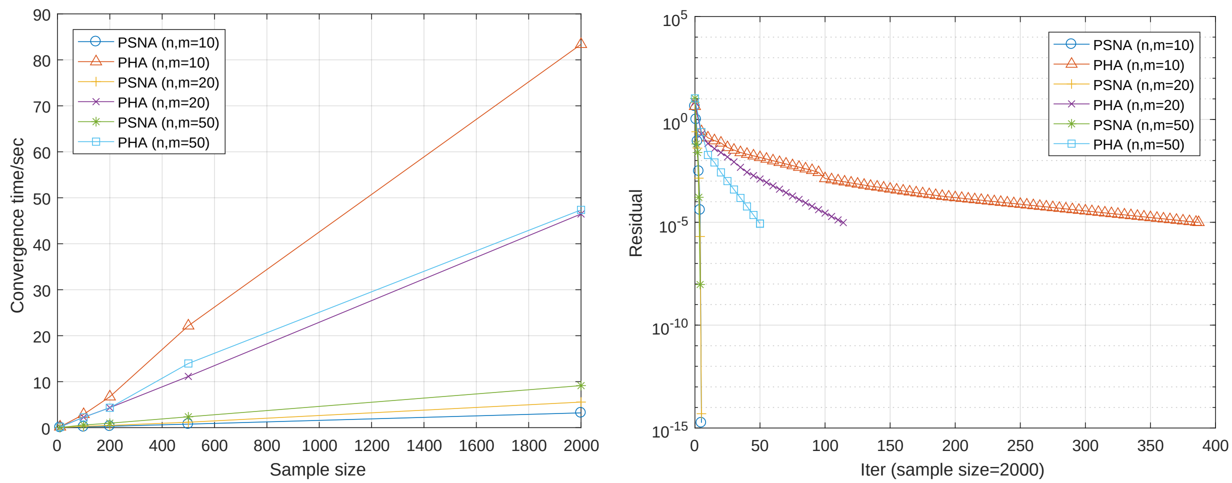
The numerical results for Example 2.1 are reported in Table 2.1 and Figure 2.1, in which the average performance profiles of algorithms are listed based on the results of ten randomly generated problems, such as the average number of iterations, average CPU time, the average solution residual. In Table 2.1, the dimensions of problems are ranging from 110 to 100050. For PSNA, the Newton iteration (2.14) performed are also given, denoted as "Iter/N". One can see that the Newton iteration is always used for most of the problems. Moreover, for PSNA, the number of iterations barely changes for different n, m and ν, while the CPU time increases linearly when n, m and ν become large. Overall, PSNA computes a more accurate solution with less number of iterations and CPU time than PHA. Table 2.1 shows that PSNA is faster than PHA in terms of CPU time. The left figure of Figure 2.1 gives an intuitive comparison of the two algorithms for different n, m when ν increases from 10 to 2000. The right-hand side figure shows the residual history with respect to the iteration number for different n and m. It is clear that PSNA is more efficient than PHA in terms of CPU time as well as the number of iterations.

In Table 2.2, numerical results of PSNA for Example 2.2 are presented. We increase n, m from 10 to 30, ν from 100 to 2000 to study the performance of PSNA. All the problems are successfully solved by PSNA. One can see that the number of iterations barely changes when ν increases and the superlinear convergence rate

Table 2.1: Comparison of PSNA and PHA for Example 2.1

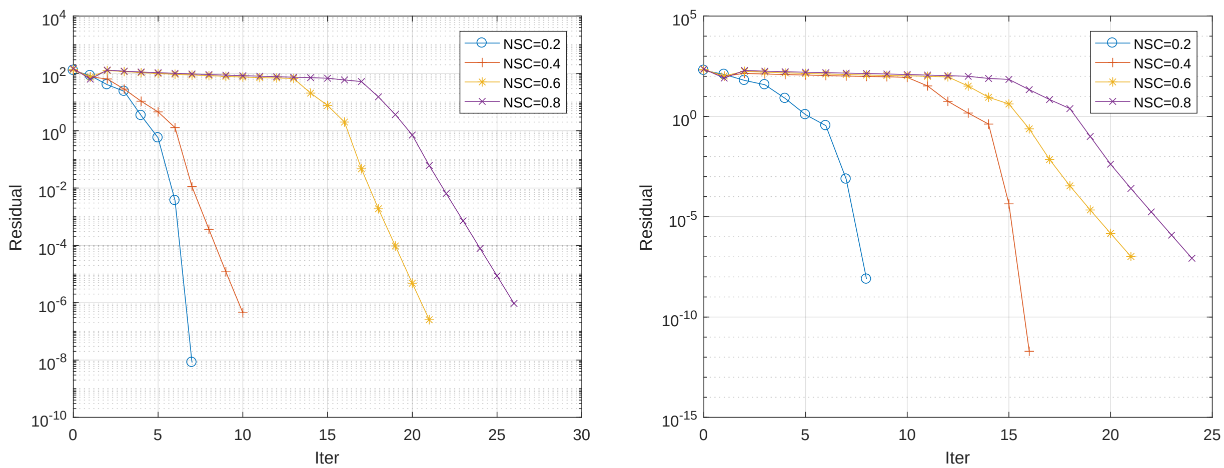
		PSNA				PHA		
n, m	ν	Iter	Iter/N	CPU	Res	Iter	CPU	Res
10	10	7.30	4.10	0.05	2.82e-16	76.20	0.18	9.24e-06
	100	4.60	4.60	0.17	5.87e-16	140.40	2.89	9.74e-06
	200	4.60	4.60	0.31	1.99e-08	171.90	6.72	9.65e-06
	500	4.60	4.60	0.76	1.15e-07	242.10	22.15	9.71e-06
	2000	5.00	5.00	3.22	1.18e-09	225.90	83.38	9.84e-06
20	10	4.00	4.00	0.03	3.60e-16	66.00	0.15	9.04e-06
	100	5.00	5.00	0.29	4.13e-16	99.00	2.26	9.93e-06
	200	4.00	4.00	0.48	8.84e-07	102.00	4.35	9.85e-06
	500	4.00	4.00	1.18	3.39e-15	106.00	11.14	9.85e-06
	2000	5.00	5.00	5.56	5.27e-15	114.00	46.47	9.63e-06
50	10	4.00	4.00	0.04	4.44e-16	38.70	0.18	8.60e-06
	100	4.50	4.50	0.57	6.49e-08	49.30	2.25	9.08e-06
	200	4.20	4.20	0.97	1.42e-07	47.70	4.33	8.95e-06
	500	4.40	4.40	2.36	1.15e-07	70.90	13.96	8.87e-06
	2000	4.20	4.20	9.14	1.27e-07	58.60	47.36	9.38e-06

is observed. Similar results for Example 2.3 are presented in Table 2.3. Table 2.4 and Figure 2.2 show the results of Example 2.4, in which the influence of nonsmooth components (NSC) of the solution is explored, where NSC is equals to the percentage of nonsmooth components A_i of function A at x^*. It can be seen that the NSC of the solution does not affect the superlinear convergence rate of PSNA, although it requires extra projection iterations when NSC is large. These results suggest that PSNA is promising even for solving some nonmonotone problems.



(a) CPU time with increasing sample sizes (b) Residuals

Figure 2.1: Comparison of PSNA and PHA for Example 2.1.



(a) $n = 30, m = 20$ (b) $n = 40, m = 30$

Figure 2.2: Numerical results of PSNA for Example 2.4 with $\nu = 1000$.

Table 2.2: Total iterations, Newton iterations, CPU time/sec for Example 2.2.

Case 1: $l = 0, u = \infty$				
n, m \ ν		500	1000	2000
20 10	Max	3.00 3.00 0.16	3.00 3.00 0.25	3.00 3.00 0.50
	Ave	3.00 3.00 0.13	3.00 3.00 0.24	3.00 3.00 0.50
	Min	3.00 3.00 0.16	3.00 3.00 0.25	3.00 3.00 0.50
30 20	Max	3.00 3.00 0.35	4.00 4.00 0.87	4.00 4.00 1.86
	Ave	3.00 3.00 0.35	3.10 3.10 0.72	3.10 3.10 1.51
	Min	3.00 3.00 0.35	3.00 3.00 0.71	3.00 3.00 1.45
60 50	Max	5.00 5.00 3.16	4.00 4.00 5.14	4.00 4.00 11.65
	Ave	4.10 4.10 2.67	4.00 4.00 5.22	4.00 4.00 11.02
	Min	4.00 4.00 2.54	4.00 4.00 5.14	4.00 4.00 11.65
Case 2: $l = -ne_n, u = ne_n$				
n, m \ ν		500	1000	2000
20 10	Max	4.00 4.00 0.15	4.00 4.00 0.30	3.00 3.00 0.48
	Ave	3.70 3.70 0.14	3.50 3.50 0.28	3.00 3.00 0.50
	Min	3.00 3.00 0.13	3.00 3.00 0.24	3.00 3.00 0.48
30 20	Max	5.00 5.00 0.52	4.00 4.00 0.96	4.00 4.00 1.91
	Ave	4.10 4.10 0.45	4.00 4.00 0.91	3.90 3.90 1.92
	Min	4.00 4.00 0.44	4.00 4.00 0.96	3.00 3.00 1.59
60 50	Max	6.00 6.00 3.85	6.00 6.00 8.21	5.00 5.00 22.43
	Ave	5.60 5.60 3.71	5.60 5.60 7.48	5.00 5.00 18.15
	Min	5.00 5.00 3.36	5.00 5.00 6.74	5.00 5.00 22.43
Case 3: $l_i = -n, u_i = n$ if i is even and $l_i = 0, u_i = \infty$ if i is odd				
n, m \ ν		500	1000	2000
20 10	Max	4.00 4.00 0.24	4.00 4.00 0.37	4.00 4.00 0.81
	Ave	3.20 3.20 0.19	3.20 3.20 0.40	3.10 3.10 0.70
	Min	3.00 3.00 0.16	3.00 3.00 0.37	3.00 3.00 0.61
30 20	Max	4.00 4.00 0.49	4.00 4.00 1.03	4.00 4.00 2.41
	Ave	4.00 4.00 0.60	3.80 3.80 1.09	3.40 3.40 2.06
	Min	4.00 4.00 0.49	3.00 3.00 1.07	3.00 3.00 1.87
60 50	Max	5.00 5.00 3.77	5.00 5.00 9.81	5.00 5.00 13.16
	Ave	5.00 5.00 3.95	4.50 4.50 7.38	4.10 4.10 13.51
	Min	5.00 5.00 3.77	4.00 4.00 6.48	4.00 4.00 11.04

Table 2.3: Numerical results of PSNA for Example 2.3.

		PSNA			
n, m	ν	Iter	Iter/N	CPU/sec	Res
20	100	1.40	1.40	0.32	1.66e-09
	500	1.30	1.30	1.46	3.38e-09
	1000	1.50	1.50	3.24	3.34e-09
	2000	1.70	1.70	6.96	4.61e-09
30	100	1.60	1.60	0.78	3.46e-08
	500	1.70	1.70	3.79	4.12e-08
	1000	1.30	1.30	6.31	1.66e-08
	2000	1.50	1.50	12.17	9.25e-09

Table 2.4: Numerical results of PSNA for Example 2.4 with $\nu = 1000$.

			PSNA			
n	m	NSC	Iter	Iter/N	CPU/sec	Res
30	20	0.20	7.60	7.60	1.73	8.75e-08
		0.40	10.40	10.40	2.48	2.18e-07
		0.60	20.00	7.80	6.61	1.35e-07
		0.80	26.70	10.90	8.87	4.46e-07
40	30	0.20	8.10	8.10	4.36	1.05e-09
		0.40	11.50	8.80	7.18	4.93e-08
		0.60	18.70	9.70	13.17	3.18e-07
		0.80	25.30	10.90	20.00	2.97e-07

Chapter 3

A regularized PSNA for the global crude oil market share problem under the COVID-19 pandemic

In this chapter, a regularized PSNA (rPSNA) is developed to find a solution of a two-stage stochastic equilibrium problem which models the global crude oil market share under the COVID-19 pandemic.

We describe the uncertainties in the demand and supply by random variables and provide two types of production decisions. A prior production decision in the first stage is independent of the outcome of random events and the supply decision in the second stage is allowed to depend on the random events in the future. The two-stage stochastic model is proposed to reflect the real market of time delay between crude oil production and supply in the COVID-19 pandemic. By utilizing oil market data in our model, we are able to simulate the actual outputs of oil producers during the pandemic, indicating their rational ways of maximizing own benefits. Thus, the relatively stable market share can be explained and anticipated by the result of an equilibrium reached by all agents competing noncooperatively.

3.1 Two-stage stochastic quadratic games

We consider an oligopolistic market where J agents compete to supply a homogeneous product non-cooperatively to the market in the future. Each agent needs to make a decision on the quantity of production based on the anticipated future market supply demand relation and other agents' decisions under an uncertain environment. The uncertainties are represented by a random variable $\xi : \Xi \to \mathbb{R}^d$ defined in the probability space $(\Xi, \mathcal{A}, P)$ with a support set Ξ and the space $\mathcal{V}$ of measurable functions from Ξ to $\mathbb{R}^J$.

We define the variables and functions for each agent i, $i = 1, \ldots, J$:

$x_i \in \mathbb{R}$, the production quantity;

$\theta_i : \mathbb{R} \to \mathbb{R}$, the cost function of the production;

$v_i(\xi) \in \mathbb{R}$, the supply quantity;

$\varphi_i : \mathbb{R} \times \Xi \to \mathbb{R}$, the cost function for supplying a quantity.

The uncertain market supply is characterized by a stochastic total supply $\tilde{\eta}(\xi) = \sum_{i=1}^{J} v_i(\xi)$ to the market and an inverse demand function $p(\tilde{\eta}(\xi), \xi) : \mathbb{R} \times \Xi \to \mathbb{R}$. For every realization $\xi \in \Xi$, $p(\tilde{\eta}(\xi), \xi)$ can be regarded as either the spot price in the future or future price at the current time.

Each agent aims to maximize its profit and make its decisions in both stages. The first stage is to make optimal decision on production quantity based on the anticipation of future demand, and the second stage is to make optimal decision on supply quantity based on the observation in the future.

For each realization of a random variable ξ and a nonnegative vector $x \in \mathbb{R}^J$, agent i wants to find an optimal decision $v_i(\xi)$ by solving the following problem

$$\begin{aligned} F_i(x, \xi) := \max_{v_i(\xi)} \quad & p(v_i(\xi) + v_{-i}(\xi), \xi) v_i(\xi) - \varphi_i(v_i(\xi), \xi) \\ \text{s.t.} \quad & 0 \le v_i(\xi) \le x_i, \qquad \text{a.e. } \xi \in \Xi. \end{aligned} \tag{3.1}$$

Moreover, agent i has to make an optimal decision before knowing the future events by solving

$$\begin{array}{ll} \max\limits_{x_i} & \mathbb{E}[F_i(x,\xi)] - \theta_i(x_i, x_{i-1}) \\ \text{s.t.} & 0 \leq x_i. \end{array} \tag{3.2}$$

Here x_{-i} and v_{-i} are decision variables of the agents other than the agent i. Problem (3.2) is the first stage, and problem (3.1) is the second stage of the two-stage stochastic games. We call $(x^*, v^*) \in \mathbb{R}^J \times \mathcal{V}$ an optimal solution of the two-stage stochastic games, if for $i = 1, \ldots, J$, $(x_i^*, v_i^*(\xi))$ is the optimal solution of the two-stage optimization problem:

$$\begin{array}{ll} \max\limits_{x_i} & \mathbb{E}[F_i(x_i, x_{-i}^*, \xi)] - \theta_i(x_i, x_{i-1}^*) \\ \text{s.t.} & 0 \leq x_i \end{array} \tag{3.3}$$

and

$$\begin{array}{lll} F_i(x_i, x_{-i}^*, \xi) := \max\limits_{v_i(\xi)} & p(v_i(\xi) + v_{-i}^*(\xi), \xi) v_i(\xi) - \varphi_i(v_i(\xi), \xi) & \\ \text{s.t.} & 0 \leq v_i(\xi) \leq x_i, \quad \text{a.e. } \xi \in \Xi. \end{array} \tag{3.4}$$

In this chapter, we consider the case where the objective functions in (3.1)-(3.2) are quadratic concave in the following forms:

$$\theta_i(x_i, x_{-i}) = \frac{1}{2} c_i x_i^2 + a_i x_i + r_i x_i \sum_{j=1}^{J} x_j, \tag{3.5}$$

$$p(v_i(\xi) + v_{-i}(\xi), \xi) = \alpha(\xi) - \gamma(\xi) \sum_{j=1}^{J} v_j(\xi), \tag{3.6}$$

$$\varphi_i(v_i(\xi), \xi) = \frac{1}{2} h_i(\xi) v_i^2(\xi) + \beta_i(\xi) v_i(\xi), \tag{3.7}$$

where $c_i > 0, \gamma(\xi) > 0, h_i(\xi) > 0$, a_i, r_i, $\alpha(\xi)$ and $\beta_i(\xi)$ are given real numbers.

Expression (3.5) represents the production cost of the oil, which is different from that in [25]. The quadratic term is used to capture the nonlinear cost of the oil production, like those of exploration, oil field construction, etc. It can also be regarded as the extra cost when the production quantity undergoes rapid changes, e.g., sudden production cut, extra cost due to excessive amount of the production. The first term in expression (3.7) is also used to represent similar type of super-linear costs in supply, e.g., oil tube capacity, oil tanker availability, etc. In the supply stage, the produced oil needs to be refined and transported to the destinations. Since the oil is produced prior to the trading time, the storage cost can also make an impact of adjusted price of each agent. In fact, one major reason that caused the WTI price to fall below US$ zero per barrel was the lack of storage facilities, not only at Cushing where the oil is physically transferred but also all over the globe where hundreds of fully loaded oil tankers being stationary at sea with no destinations to go and even oil pipes are taken as temporal storage. The third term in expression (3.5) is to represent the fact that each agent may have chosen different approaches in response to the strategies of the others. The sign of r_i is to show whether agent i decides to go along with the market or go against it. The ideas and rationale will be explained in detail in Subsection 3.3.2 when we apply the model to real market data.

In such setting, the function

$$\begin{aligned} F_i(x,\xi) = \max_{v_i(\xi)} \quad & (\alpha(\xi) - \gamma(\xi)\sum_{j=1}^{J} v_j(\xi))v_i(\xi) - \frac{1}{2}h_i(\xi)v_i^2(\xi) - \beta_i(\xi)v_i(\xi) \\ \text{s.t.} \quad & 0 \le v_i(\xi) \le x_i, \qquad \text{a.e. } \xi \in \Xi \end{aligned}$$

is continuously differentiable with respect to x_i for $x_i > 0$ and

$$\nabla_{x_i} F_i(x,\xi) = s_i(\xi),$$

where $s_i(\xi) \ge 0$ are Lagrange multipliers for the constraints $v_i(\xi) \le x_i$. If $x_i = 0$, then the optimal solution of the optimization problem (3.4) is $v_i(\xi) \equiv 0$ for any $\xi \in \Xi$, and the subdifferential $\partial_{x_i} F_i(x,\xi)$ of $F_i(x,\xi)$ is

$$\{s_i(\xi) \,|\, s_i(\xi) \ge \max(0, \alpha(\xi) - \beta_i(\xi) - \gamma(\xi)e^T v(\xi)), \, \mathbb{E}[s_i(\xi)] \le r_i e^T x + a_i\}.$$

The Karush-Kuhn-Tucker (KKT) conditions of problems (3.1)-(3.2) with the functions defined in (3.5)-(3.7) derive the following two-stage SLCP

$$0 \leq \begin{pmatrix} x \\ v(\xi) \\ s(\xi) \end{pmatrix} \perp \begin{pmatrix} Ax - \mathbb{E}[s(\xi)] + a \\ (\Lambda_2(\xi) + \gamma(\xi)ee^T)v(\xi) + s(\xi) + \rho(\xi) \\ x - v(\xi) \end{pmatrix} \geq 0, \tag{3.8}$$

for a.e. $\xi \in \Xi$, where

$$x = (x_1, \ldots, x_J)^T, \quad a = (a_1, \ldots, a_J)^T, \quad r = (r_1, \ldots, r_J)^T,$$

$$v(\xi) = (v_1(\xi), \ldots, v_J(\xi))^T, \quad s(\xi) = (s_1(\xi), \ldots, s_J(\xi))^T,$$

$$\rho(\xi) = (-\alpha(\xi) + \beta_1(\xi), \ldots, -\alpha(\xi) + \beta_J(\xi))^T,$$

$$A = \Lambda_1 + re^T, \quad \Lambda_1 = \mathrm{diag}(c_1 + r_1, \ldots, c_J + r_J),$$

$$\Lambda_2(\xi) = \mathrm{diag}(h_1(\xi) + \gamma(\xi), \ldots, h_J(\xi) + \gamma(\xi)),$$

and $e \in \mathbb{R}^J$ is the vector with all components being 1.

We assume that the matrix A is positive definite. Then problems (3.1)-(3.2) with the functions defined in (3.5)-(3.7) are equivalent to problem (3.8) in the sense that if $(x^*, v^*(\cdot))$ is a solution of problems (3.1)-(3.2), then there is an $s^*(\cdot)$ such that $(x^*, v^*(\cdot),\ s^*(\cdot))$ is a solution of (3.8); conversely, if $(x^*, v^*(\cdot), s^*(\cdot))$ is a solution of (3.8), then $(x^*, v^*(\cdot))$ is a solution of problems (3.1)-(3.2). A sufficient condition for the matrix $A = \Lambda_1 + re^T$ being positive definite is

$$c_i + 2r_i > \frac{1}{2} \sum_{j \neq i}^{J} |r_j + r_i|, \quad i = 1, \ldots, J. \tag{3.9}$$

Condition (3.9) implies that $\Lambda_1 + \frac{1}{2}(re^T + er^T)$ is a symmetric diagonally dominate matrix with positive diagonal elements and thus a positive definite matrix.

Let

$$M(\xi) = \begin{pmatrix} \Lambda_2(\xi) + \gamma(\xi)ee^T & I \\ -I & 0 \end{pmatrix} \in \mathbb{R}^{2J\times 2J},$$

$$y(\xi) = \begin{pmatrix} v(\xi) \\ s(\xi) \end{pmatrix} \in \mathbb{R}^{2J},\ q(x,\xi) = \begin{pmatrix} \rho(\xi) \\ x \end{pmatrix} \in \mathbb{R}^{2J}. \tag{3.10}$$

Since $M(\xi)$ is a positive semi-definite matrix, $\mathrm{SOL}(q(x,\xi), M(\xi))$ is a convex set [17]. A vector $\hat{y}(x,\xi)$ is called a least-norm solution of the $\mathrm{LCP}(q(x,\xi), M(\xi))$ if it is the solution of the optimization problem

$$\min \|y\|^2 \quad \text{s.t.}\ y \in \mathrm{SOL}(q(x,\xi), M(\xi)).$$

The following lemma gives the form of the least-norm solution of the $\mathrm{LCP}(q(x,\xi), M(\xi))$ and shows that it is the unique solution of the $\mathrm{LCP}(q(x,\xi), M(\xi))$ when $x > 0$.

Lemma 3.1. *For any $x \geq 0$ and $\xi \in \Xi$, the* $\mathrm{LCP}(q(x,\xi), M(\xi))$ *has the least-norm solution $\hat{y}(x,\xi) = (\hat{v}(x,\xi), \hat{s}(x,\xi))^T$ with*

$$\hat{v}(x,\xi) = \Pi_{[0,x]}(\hat{v}(x,\xi) - \rho(\xi) - (\Lambda_2(\xi) + \gamma(\xi)ee^T)\hat{v}(x,\xi)), \tag{3.11}$$

$$\hat{s}(x,\xi) = \max(0, -\rho(\xi) - (\Lambda_2(\xi) + \gamma(\xi)ee^T)\hat{v}(x,\xi)), \tag{3.12}$$

where $\Pi_{[0,x]}(z)$ is the projection from z to the set $[0,x]$. Moreover, the least-norm solution is the unique solution of the $\mathrm{LCP}(q(x,\xi), M(\xi))$ *if $x > 0$.*

Proof. It is easy to see that $y(\xi) = (x, |\rho(\xi)|)^T \geq 0$ and $M(\xi)y(\xi) + q(x,\xi) \geq 0$; that is, $(x, |\rho(\xi)|)^T$ is a feasible solution of the $\mathrm{LCP}(q(x,\xi), M(\xi))$. Hence from the feasibility and the positive semi-definiteness of $M(\xi)$, the $\mathrm{LCP}(q(x,\xi), M(\xi))$ has at least one solution [17, Theorem 3.1.2].

The $\mathrm{LCP}(q(x,\xi), M(\xi))$ is the first-order optimality condition of the strongly convex quadratic program

$$\begin{aligned} \min\ & \bar{F}_x(v(\xi),\xi) := \frac{1}{2}v(\xi)^T(\Lambda_2(\xi) + \gamma(\xi)ee^T)v(\xi) + \rho(\xi)^T v(\xi) \\ \text{s.t.}\ & 0 \leq v(\xi) \leq x. \end{aligned} \tag{3.13}$$

Because of the strong convexity, the first-order optimality condition is necessary and sufficient for the unique optimal solution $\hat{v}(x,\xi)$ of the above problem, which is a fixed point of the fixed point problem

$$\hat{v}(x,\xi) = \Pi_{[0,x]}(\hat{v}(x,\xi) - \rho(\xi) - (\Lambda_2(\xi) + \gamma(\xi)ee^T)\hat{v}(x,\xi)).$$

Hence, $\hat{v}(x,\xi)$ has the form of (3.11).

By the definition of the projection, we can easily obtain

$$(\rho(\xi) + (\Lambda_2(\xi) + \gamma(\xi)ee^T)\hat{v}(x,\xi))_i$$

$$\begin{cases} = 0 & \text{if } (\hat{v}(x,\xi) - \rho(\xi) - (\Lambda_2(\xi) + \gamma(\xi)ee^T)\hat{v}(x,\xi))_i \in (0, x_i), \\ \leq 0 & \text{if } (\hat{v}(x,\xi) - \rho(\xi) - (\Lambda_2(\xi) + \gamma(\xi)ee^T)\hat{v}(x,\xi))_i \geq x_i, \\ \geq 0 & \text{otherwise.} \end{cases}$$

Hence, the multiplier $\hat{s}(x,\xi)$ in (3.12) with $\hat{v}(x,\xi)$ in (3.11) is a solution of the LCP($q(x,\xi)$, $M(\xi)$), where for the case $\hat{v}_i(x,\xi) < x_i$ it is from $\hat{s}_i(x,\xi) = (\rho(\xi) + (\Lambda_2(\xi)+\gamma(\xi)ee^T)\hat{v}(x,\xi))_i = 0$ and $\hat{s}_i(x,\xi)(x_i - \hat{v}_i(x,\xi)) = 0$, and for the case $\hat{v}_i(x,\xi) = x_i$, it is from $\hat{v}_i(x,\xi)(\hat{s}_i(x,\xi) + (\rho(\xi) + (\Lambda_2(\xi) + \gamma(\xi)ee^T)\hat{v}(x,\xi))_i) = 0$.

If $\hat{v}_i(x,\xi) = x_i > 0$, $\hat{s}_i(x,\xi)$ in (3.12) is uniquely defined. If $\hat{v}_i(x,\xi) = x_i = 0$, then $\hat{s}_i(x,\xi) = \max(0, -(\rho(\xi) + (\Lambda_2(\xi) + \gamma(\xi)ee^T)\hat{v}(x,\xi))_i) = 0$. Hence $\hat{y}(x,\xi) = (\hat{v}(x,\xi), \hat{s}(x,\xi))^T$ defined in (3.11)-(3.12) is the least-norm solution of the LCP($q(x,\xi)$, $M(\xi)$) and the unique solution of LCP($q(x,\xi), M(\xi)$) if $x > 0$. We complete the proof. $\square$

Lemma 3.1 shows that for a fixed vector $x \geq 0$, the least-norm solution of the LCP($q(x,\xi), M(\xi)$) is uniquely defined by $\Lambda_2(\xi)$, $\rho(\xi)$ and $\gamma(\xi)$. These random data may have noise in real applications. Now we consider the perturbation bound of the solution regarding noise in $\Lambda_2(\xi)$, $\rho(\xi)$ and $\gamma(\xi)$.

Let us consider the following LCP($\bar{q}, \bar{M}$) with

$$\bar{M} = \begin{pmatrix} \bar{\Lambda}_2 + \bar{\gamma}ee^T & I \\ -I & 0 \end{pmatrix}, \quad \bar{q} = \begin{pmatrix} \bar{\rho} \\ x \end{pmatrix}.$$

From Lemma 3.1, the LCP$(\bar{q}, \bar{M})$ with $x \geq 0$ has the least-norm solution $\bar{y} = (\bar{u}, \bar{t})^T$ in the following form

$$\bar{u} = \Pi_{[0,x]}(\bar{u} - \bar{\rho} - (\bar{\Lambda}_2 + \bar{\gamma} ee^T)\bar{u}) \quad \text{and} \quad \bar{t} = \max(0, -\bar{\rho} - (\bar{\Lambda}_2 + \bar{\gamma} ee^T)\bar{u}).$$

Now we consider the distance $\|\hat{v}(x,\xi) - \bar{u}\|$ and $\|\hat{s}(x,\xi) - \bar{t}\|$.

Theorem 3.1. *Suppose that $x \geq 0$ and let $\Gamma > 0$ such that $\Gamma\gamma(\xi) \geq 1$. Then the least-norm solution of the LCP$(q(x,\xi), M(\xi))$ is continuous with respect to $\Lambda_2(\xi)$, $\rho(\xi)$ and $\gamma(\xi)$. Moreover, we have the following perturbation error bound*

$$\|\hat{v}(x,\xi) - \bar{u}\| \leq \Gamma(\|\rho(\xi) - \bar{\rho}\| + \|x\|\|\Lambda_2(\xi) - \bar{\Lambda}_2\| + J\|x\|\|\gamma(\xi) - \bar{\gamma}\|). \tag{3.14}$$

Proof. We first prove (3.14). Let

$$w(\xi) = (\Lambda(\xi) + \gamma(\xi) ee^T)\hat{v}(x,\xi) + \rho(\xi) \quad \text{and} \quad \bar{w} = (\bar{\Lambda}_2 + \bar{\gamma} ee^T)\bar{u} + \bar{\rho}.$$

Then we have

$$\hat{v}(x,\xi) = \Pi_{[0,x]}(\hat{v}(x,\xi) - w(\xi)) \quad \text{and} \quad \bar{u} = \Pi_{[0,x]}(\bar{u} - \bar{w}),$$

which implies

$$\text{mid}(\hat{v}(x,\xi), \hat{v}(x,\xi) - x, w(\xi)) = 0 \quad \text{and} \quad \text{mid}(\bar{u}, \bar{u} - x, \bar{w}) = 0,$$

where "mid" is the componentwise median operator. Following the proof of [11, Lemma 2.1], there is a diagonal matrix $\bar{D}$ =diag$(d_1, \ldots, d_J)$ with $0 \leq d_i \leq 1$ such that

$$0 = (I - \bar{D})(\hat{v}(x,\xi) - \bar{u}) + \bar{D}(w(\xi) - \bar{w}).$$

Hence we obtain

$$(I - \bar{D} + \bar{D}(\Lambda_2(\xi) + \gamma(\xi) ee^T))(\hat{v}(x,\xi) - \bar{u}) = -\bar{D}(\rho(\xi) - \bar{\rho} + (\Lambda_2(\xi) - \bar{\Lambda}_2 + (\gamma(\xi) - \bar{\gamma}) ee^T)\bar{u}). \tag{3.15}$$

Since $\Lambda_2(\xi) + \gamma(\xi)ee^T$ is symmetric positive definite, by [13, Theorem 2.7], we have

$$\begin{aligned}\max_{d\in[0,1]^J} \|(I - \bar{D} + \bar{D}(\Lambda_2(\xi) + \gamma(\xi)ee^T))^{-1}\bar{D}\| &= \|(\Lambda_2(\xi) + \gamma(\xi)ee^T))^{-1}\| \\ &\leq \gamma(\xi)^{-1}\|(I + ee^T)^{-1}\|.\end{aligned}$$

Therefore, using $\|(I+ee^T)^{-1}\| = 1$, $\|ee^T\| = J$, $\gamma(\xi)^{-1} \leq \Gamma$ and $0 \leq \bar{u} \leq x$, we obtain (3.14) from (3.15).

Next we show the continuity of the last J-components of the least norm solution of the LCP$(q(x,\xi), M(\xi))$. Without loss of generality, assume that $\sigma = \{i \in \{1,\dots,J\} \,|\, \hat{v}_i(x,\xi) < x_i\} \neq \emptyset$. From (3.14), for any $0 < \epsilon < \min_{i\in\sigma} x_i - \hat{v}_i(x,\xi)$, there is $\delta > 0$ such that if $\|\rho(\xi)-\bar{\rho}\| + \|\Lambda_2(\xi)-\bar{\Lambda}_2\| + \|\gamma(\xi)-\bar{\gamma}\| < \delta$, then $\|\hat{v}(x,\xi)-\bar{u}\| < \epsilon$. This implies for $i \in \sigma$, $x_i - \bar{u}_i \geq x_i - \hat{v}_i(x,\xi) - |\hat{v}_i(x,\xi) - \bar{u}_i| > 0$. Hence, we have $\bar{t}_i = \hat{s}_i(x,\xi) = 0$.

For $i \notin \sigma$, that is, $\hat{v}_i(x,\xi) = x_i$, we have

$$\hat{s}_i(x,\xi) = -(\rho(\xi) + (\Lambda_2(\xi) + \gamma(\xi)ee^T)\hat{v}(x,\xi))_i \geq 0.$$

If $\bar{t}_i = 0$, then $-(\bar{\rho} + (\bar{\Lambda}_2 + \bar{\gamma}ee^T)\bar{u})_i \leq 0$, otherwise $\bar{t}_i = -(\bar{\rho} + (\bar{\Lambda}_2 + \bar{\gamma}ee^T)\bar{u})_i \geq 0$. Hence, using (3.14), we have

$$\begin{aligned}&|\hat{s}_i(x,\xi) - \bar{t}_i| \\ \leq\; & |(\rho(\xi) + (\Lambda_2(\xi) + \gamma(\xi)ee^T)\hat{v}(x,\xi))_i - (\bar{\rho} + (\bar{\Lambda}_2 + \bar{\gamma}ee^T)\bar{u})_i| \\ \leq\; & \|\rho(\xi) - \bar{\rho}\| + (\|\Lambda_2(\xi)\| + \gamma(\xi)J)\|\hat{v}(x,\xi) - \bar{u}\| + \|x\|(J\|\gamma(\xi) - \bar{\gamma}\| + \|\Lambda_2(\xi) - \bar{\Lambda}_2\|) \\ \leq\; & (L + (\|\Lambda_2(\xi)\| + \gamma(\xi)J)\Gamma)(\|\rho(\xi) - \bar{\rho}\| + \|\Lambda_2(\xi) - \bar{\Lambda}_2\| + \|\gamma(\xi) - \bar{\gamma}\|)),\end{aligned}$$

where $L \geq \max\{1, J\|x\|\}$. Hence, the solution of the LCP$(q(x,\xi), M(\xi))$ is continuous with respect to $\Lambda_2(\xi)$, $\rho(\xi)$ and $\gamma(\xi)$. We complete the proof. $\square$

3.2 A regularized projection semismooth Newton algorithm (rPSNA)

In this section, we propose rPSNA to solve the SAA discretization problem of (3.8). Given an i.i.d. sample set $\Xi_\nu = \{\xi_1, \ldots, \xi_\nu\}$ of ξ with $\nu > 0$, the probability for each realization is $1/\nu$. For simplicity, we set $B = (0, I) \in \mathbb{R}^{J \times 2J}$, $\varrho(\xi) = (\rho(\xi), 0)^T \in \mathbb{R}^{2J}$, $n = J + 2J\nu$ and

$$\mathbf{M} = \begin{pmatrix} A & -\frac{1}{\nu}B & \ldots & -\frac{1}{\nu}B \\ B^T & M_1 & & \\ \vdots & & \ddots & \\ B^T & & & M_\nu \end{pmatrix} \in \mathbb{R}^{n \times n}, \quad \mathbf{q} = \begin{pmatrix} a \\ \varrho_1 \\ \vdots \\ \varrho_\nu \end{pmatrix} \in \mathbb{R}^n,$$

where $M_\ell = M(\xi_\ell)$ and $\varrho_\ell = \varrho(\xi_\ell)$, $\ell = 1, \ldots, \nu$. Thus, the SAA discretization problem of (3.8) reads

$$0 \leq \mathbf{z} \perp \mathbf{M}\mathbf{z} + \mathbf{q} \geq 0, \tag{3.16}$$

where $\mathbf{z} = (x, y_1, \ldots, y_\nu)^T$ with $y_\ell = (v_\ell, s_\ell)^T$. In such setting, the problem (3.16) is a deterministic LCP$(\mathbf{q}, \mathbf{M})$. We first study its solvability.

Theorem 3.2. *The problem* (3.16) *has at least one solution* $\mathbf{z}^* = (x^*, y_1^*, \ldots, y_\nu^*)^T$*, and has at most one solution with* $x^* > 0$*.*

Proof. Let $\Lambda = \text{diag}(\nu, 1, \ldots, 1) \in \mathbb{R}^{n \times n}$ with the first J diagonal elements of Λ being ν. It is easy to verify that $\mathbf{z}^*$ is a solution of the LCP$(\Lambda\mathbf{q}, \Lambda\mathbf{M})$ if and only if $\mathbf{z}^*$ is a solution of the LCP$(\mathbf{q}, \mathbf{M})$.

Since $z^T(\Lambda\mathbf{M} + (\Lambda\mathbf{M})^T)z \geq 0$ for any $z \in \mathbb{R}^n$, the matrix $\Lambda\mathbf{M}$ is positive semi-definite. Moreover, $\mathbf{z} = (x, x, |\rho_1|, \ldots, x, |\rho_\nu)|)^T \in \mathbb{R}^n$ with $x_i > \max\{0, (\frac{1}{\nu}\sum_{\ell=1}^{\nu} |\rho_\ell| - a)_i / c_i\}$, $i = 1, \ldots, J$ is a feasible solution of the LCP$(\Lambda\mathbf{q}, \Lambda\mathbf{M})$. Hence the feasibility implies that the LCP$(\Lambda\mathbf{q}, \Lambda\mathbf{M})$ has at least one solution $\mathbf{z}^*$ [17]. If $\mathbf{z}^*$ has the component $x^* > 0$, then by Lemma 3.1, the component $(y_\ell^*, \ldots, y_\nu^*)$ of $\mathbf{z}^*$ is uniquely dependent on $x^* > 0$. Moreover, we have

$$Ax^* + \frac{1}{\nu}\sum_{i=\ell}^{\nu} \bar{D}_\ell(\rho_\ell + ((\Lambda_2)_\ell + \gamma_\ell ee^T)x^*) + a = 0,$$

where $\bar{D}_\ell$ is a diagonal matrix with diagonal elements being 0 or 1. Since A and $(\Lambda_2)_\ell + \gamma_\ell ee^T$ are positive definite, x^* is the unique solution of this system of equations. Hence, the LCP$(\Lambda\mathbf{q}, \Lambda\mathbf{M})$ has at most one solution with $x^* > 0$. □

By Lemma 3.1, we know that the LCP$(\mathbf{q}, \mathbf{M})$ has the relatively complete recourse. Since M_ℓ is only positive semidefinite, the second stage problem may have multiple solutions for some fixed x and ξ. In such case, we always choose the unique least-norm solution $\hat{y}(x, \xi)$. By substituting $\hat{y}(x, \xi)$ into the first stage problem, we get

$$0 \leq x \perp H(x) \geq 0, \tag{3.17}$$

where $H(x) := Ax - \frac{1}{\nu}\sum_{\ell=1}^{\nu} B\hat{y}(x, \xi_\ell) + a$. Its corresponding residual function is written as

$$\hat{Q}(x) = \min(x, Ax - \frac{1}{\nu}\sum_{\ell=1}^{\nu} B\hat{y}(x, \xi_\ell) + a). \tag{3.18}$$

Note that $\hat{y}(x, \xi_\ell)$ is not guaranteed to be a Lipschitz continuous function as pointed out in [14], and thus H and $\hat{Q}$ are not necessarily Lipschitz continuous. To overcome such difficulties, we propose rPSNA to solve (3.17). Since for each fixed x and ξ the second stage problem is only a monotone problem, we add a regularized term $\mu_k I$ (with $\mu_k \to 0$) to the second stage LCP in (3.8) to force the whole problem to be strongly monotone. By LCP$(\mathbf{q}, \mathbf{M}^{\mu_k})$, we denote the regularized problem with μ_k, where

$$\mathbf{M}^{\mu_k} = \begin{pmatrix} A & -\frac{1}{\nu}B & \cdots & -\frac{1}{\nu}B \\ B^T & M_1^{\mu_k} & & \\ \vdots & & \ddots & \\ B^T & & & M_\nu^{\mu_k} \end{pmatrix}$$

with

$$M_\ell^{\mu_k} = \begin{pmatrix} \Lambda_2(\xi_\ell) + \gamma(\xi_\ell)ee^T + \mu_k I & I \\ -I & \mu_k I \end{pmatrix}, \quad \ell = 1, \dots, \nu.$$

With the regularized parameter μ_k, for each fixed x and ξ, the second stage problem becomes a strongly monotone LCP, and there exists a unique solution $\hat{y}_{\mu_k}(x,\xi)$ that is Lipschitz continuous with respect to x for any ξ. Substituting $\hat{y}_{\mu_k}$ into the corresponding first stage problem, we can get the single-stage reformulation of the LCP($\mathbf{q}, \mathbf{M}^{\mu_k}$) as follows:

$$0 \le x \perp H_{\mu_k}(x) \ge 0, \tag{3.19}$$

where

$$H_{\mu_k}(x) := Ax - \frac{1}{\nu}\sum_{\ell=1}^{\nu} B\hat{y}_{\mu_k}(x,\xi_\ell) + a.$$

By Propositions 2.4 and Lemma 2.2, we can show that H_{μ_k} is strongly monotone and Lipschitz continuous on $\mathbb{R}^J$ for each fixed $\mu_k > 0$. In addition, the residual function for (3.19) is also Lipschitz continuous, which reads

$$\hat{Q}_{\mu_k}(x) = \min(x, Ax - \frac{1}{\nu}\sum_{\ell=1}^{\nu} B\hat{y}_{\mu_k}(x,\xi_\ell) + a).$$

Therefore, PSNA can be applied to solve (3.19) for each $\mu_k > 0$.

By [17, Theorem 5.6.2 (b)], we know that

$$\lim_{k\to\infty} \hat{y}_{\mu_k}(x,\xi) = \hat{y}(x,\xi), \quad \text{for any } x \ge 0,\ \xi \in \Xi_\nu.$$

Then, we have for any $x \ge 0$,

$$\lim_{k\to\infty} H_{\mu_k}(x) = H(x), \tag{3.20}$$

$$\lim_{k\to\infty} \hat{Q}_{\mu_k}(x) = \hat{Q}(x). \tag{3.21}$$

We formally present rPSNA as follows.

Algorithm 3.1. Regularized Projection Semismooth Newton Algorithm (rPSNA)

Step 0: *Choose an initial point* $x^0 \in \mathbb{R}^n_+$, $\tau \in (0,1)$ *and initial regularized parameter* $\mu_0 > 0$. *Set* $k = 0$, $x^{k,0} = x^0$.

Step 1: *Calculate* x^k *which solves the regularized single-stage problem* (3.19) *with regularized parameter* μ_k *by PSNA with starting point* $x^{k,0}$. *If* $\|\hat{Q}(x^k)\| = 0$, *stop. Otherwise, go to Step 2.*

Step 2: *Let* $\mu_{k+1} = \min(\|\hat{Q}(x^k)\|, \tau)\mu_k$, $x^{k+1,0} = x^k$. *Set* $k := k+1$; *go back to Step 1.*

Remark: In Step 1, we use a warm-start strategy when solving the regularized single-stage problem (3.19). This not only improves the numerical efficiency of the algorithm but also avoids some computational difficulties that will discuss latter.

The following proposition establishes the strong monotonicity and the solvability of the single-stage problem (3.17).

Proposition 3.3. *The single-stage problem* (3.17) *is strongly monotone on* $\mathbb{R}^J_+$ *and admits a unique solution* x^*. *Then* $(x^*, \hat{y}(x^*, \xi_1), \ldots, \hat{y}(x^*, \xi_\nu))$ *is a solution of the* LCP$(\mathbf{q}, \mathbf{M})$.

Proof. For any fixed $\mu_k > 0$ and ξ_ℓ, it holds

$$\begin{pmatrix} u^T & v^T \end{pmatrix} \begin{pmatrix} A & -B \\ B^T & M_\ell^{\mu_k} \end{pmatrix} \begin{pmatrix} u \\ v \end{pmatrix} \geq \lambda_{\min}\|u\|^2 + \mu_k\|v\|^2, \ \forall\ (u,v) \in \mathbb{R}^J \times \mathbb{R}^{2J},$$

where $\lambda_{\min}$ denotes the minimum eigenvalue of $(A + A^T)/2$. Then, by a similar argument as in [10, Lemma 2.1], we can show that for any fixed $\mu_k > 0$,

$$u^T[A + B(I - \Lambda_{\mathcal{J}} + \Lambda_{\mathcal{J}} M_\ell^{\mu_k})^{-1}\Lambda_{\mathcal{J}} B^T]u \geq \lambda_{\min}\|u\|^2, \quad \forall\ u \in \mathbb{R}^J,$$

where $\mathcal{J} \subseteq [J]$ with $[J] := \{1, \ldots, J\}$, and $\Lambda_{\mathcal{J}} \in \mathbb{R}^{J\times J}$ is a diagonal matrix with $(\Lambda_{\mathcal{J}})_{jj} = 1$ if $j \in \mathcal{J}$ and $(\Lambda_{\mathcal{J}})_{jj} = 0$ otherwise. By [8, Theorem 3.9], we know that

H_{μ_k} is strongly monotone on $\mathbb{R}^J_+$ such that

$$(x - x')^T(H_{\mu_k}(x) - H_{\mu_k}(x')) \geq \lambda_{\min}\|x - x'\|^2, \quad \forall\, x, x' \in \mathbb{R}^J_+.$$

By (3.20), taking the limit for the above inequality, we get that

$$(x - x')^T(H(x) - H(x')) \geq \lambda_{\min}\|x - x'\|^2,$$

which implies that H is strongly monotone on $\mathbb{R}^J_+$. Therefore, (3.17) admits a unique solution x^* and $(x^*, \hat{y}(x^*, \xi_1), \ldots, \hat{y}(x^*, \xi_1))$ is a solution of the LCP$(\mathbf{q}, \mathbf{M})$. $\square$

The following proposition follows from Lemma 2.2.

Proposition 3.4. *For any fixed $\mu > 0$, the unique solution of the regularized second stage problem* LCP$(\varrho_\ell, M^\mu_\ell)$ *for ξ_ℓ reads*

$$\hat{y}_\mu(x, \xi_\ell) = -U_\mu(x, \xi_\ell)(B^T x + \varrho_\ell),$$

where $U_\mu(x, \xi_\ell) = (I - \Lambda(x, \xi_\ell) + \Lambda(x, \xi_\ell)M^\mu_\ell)^{-1}\Lambda(x, \xi_\ell)$ with $\Lambda(x, \xi_\ell)$ being a diagonal matrix satisfying $\Lambda_{ii}(x, \xi_\ell) = 1$ if $(M^\mu_\ell \hat{y}_\mu(x, \xi_\ell) + B^T x + \varrho_\ell)_i < (\hat{y}_\mu(x, \xi_\ell))_i$ and $\Lambda_{ii}(x, \xi_\ell) = 0$ otherwise. $\hat{y}_\mu(\cdot, \xi_\ell)$ is strongly semismooth on $\mathbb{R}^J$ and globally Lipschitz continuous with Lipschitz constant $L_{\hat{y}_\mu} = \|B\|\max_{\mathcal{J}\subseteq[J]}\|(M^\mu_\ell)^{-1}_{\mathcal{J}\mathcal{J}}\|$. In addition, we have

$$-U_\mu(x, \xi_\ell)B^T \in \partial \hat{y}_\mu(x, \xi_\ell).$$

From the above proposition, it is not hard to see that the Lipschitz constant of H_μ is

$$L_{H_\mu} = \|A\| + \frac{1}{\nu}\|B\|^2 \sum_{\ell=1}^{\nu} \max_{\mathcal{J}\subseteq[J]}\|(M^\mu_\ell)^{-1}_{\mathcal{J}\mathcal{J}}\|.$$

H_μ and $\hat{Q}_\mu$ are strongly semimsooth on $\mathbb{R}^J$ by the strong semismoothness of $\hat{y}_\mu(\cdot, \xi_\ell)$ for every ξ_ℓ. The linear Newton approximation scheme $\mathcal{H}_\mu : \mathbb{R}^J_+ \rightrightarrows \mathbb{R}^{J\times J}$ reads

$$\mathcal{H}_\mu(x) = A - \frac{1}{\nu} B \sum_{\ell=1}^{\nu} \partial_x \hat{y}_\mu(x, \xi_\ell).$$

One element of $\mathcal{H}_\mu$ at x can be calculated as

$$A + \frac{1}{\nu} B \sum_{\ell=1}^{\nu} U_\mu(x, \xi_\ell) B^T.$$

Actually, by the proof in Proposition 3.3, it is not hard to see that for any $x \in \mathbb{R}^J_+$, every element in $\mathcal{H}_\mu(x)$ is positive definite. Moreover, since H_μ is strongly monotone on $\mathbb{R}^J_+$, by Lemma 9.1.3 and Corollary 9.1.28 in [19], we can deduce that the level set

$$\mathcal{L}_0^\mu = \{x \in \mathbb{R}^J_+ : \|\hat{Q}_\mu(x)\| \leq \|\hat{Q}_\mu(x^0)\|\} \quad \text{is bounded for any fixed } \mu > 0.$$

Then, we have the following convergence theorem about PSNA applying to the regularized problem (3.19).

Theorem 3.5. *For any fixed $\mu > 0$, PSNA applying to solve the regularized problem* (3.19) *is globally convergent and the convergence rate is superlinear.*

Proof. It follows from Theorem 2.6. We omit the details. □

The above theorem shows that the Step 1 of rPSNA is always implementable for any fixed $\mu > 0$. Next, we study the global convergence of rPSNA.

Theorem 3.6. *The sequence $\{x^k\}$ generated by rPSNA for solving* (3.17) *globally converges to the unique solution x^*.*

Proof. We first show that the sequences $\{\hat{y}_{\mu_k}(x^k, \xi_\ell)\}$ and $\{\hat{y}(x^k, \xi_\ell)\}$ are bounded for each ξ_ℓ. By the proof in Lemma 3.1, for any $x^k \geq 0$, the following optimization problem

$$\min_{v(\xi_\ell)} \bar{F}_{x^k}(v(\xi_\ell), \xi_\ell), \qquad \text{s.t. } 0 \leq v(\xi_\ell) \leq x^k,$$

admits a unique solution which is bounded by the strong convexity of the objective function. By (3.12), it is known that $\hat{s}(x^k, \xi_\ell)$ is bounded. Thus, $\{\hat{y}(x^k, \xi_\ell)\}$ is bounded. Since x^k is the unique solution of (3.19), we have

$$\min(x^k, Ax^k - \sum_{\ell=1}^{\nu} B\hat{y}_{\mu_k}(x^k, \xi_\ell) + a) = 0, \tag{3.22}$$

in which $\hat{y}_{\mu_k}(x^k, \xi_\ell)$ is the unique solution of the LCP$(q(x^k, \xi_\ell), M_\ell^{\mu_k})$, i.e.,

$$0 \leq \hat{y}_{\mu_k}(x^k, \xi_\ell) \perp M_\ell^{\mu_k} \hat{y}_{\mu_k}(x^k, \xi_\ell) + q(x^k, \xi_\ell) \geq 0. \tag{3.23}$$

In addition, $\hat{y}(x^k, \xi_\ell)$ is the unique least-norm solution of LCP$(q(x^k, \xi_\ell), M_\ell)$; that is,

$$0 \leq \hat{y}(x^k, \xi_\ell) \perp M_\ell \hat{y}(x^k, \xi_\ell) + q(x^k, \xi_\ell) \geq 0.$$

By the last two relations and the positive semi-definiteness of M_ℓ, we have

$$\begin{aligned} 0 &\geq (\hat{y}_{\mu_k}(x^k, \xi_\ell) - \hat{y}(x^k, \xi_\ell))^T (M_\ell(\hat{y}_{\mu_k}(x^k, \xi_\ell) - \hat{y}(x^k, \xi_\ell)) + \mu_k \hat{y}_{\mu_k}(x^k, \xi_\ell)) \\ &\geq \mu_k (\hat{y}_{\mu_k}(x^k, \xi_\ell) - \hat{y}(x^k, \xi_\ell))^T \hat{y}_{\mu_k}(x^k, \xi_\ell), \end{aligned}$$

which implies that

$$\|\hat{y}_{\mu_k}(x^k, \xi_\ell)\| \leq \|\hat{y}(x^k, \xi_\ell)\|. \tag{3.24}$$

Thus, $\{\hat{y}_{\mu_k}(x^k, \xi_\ell)\}$ is also bounded.

Next, we show that the sequence $\{x^k\}$ is bounded. Since x^* is the unique solution of (3.17), we have

$$\min(x^*, Ax^* - \sum_{\ell=1}^{\nu} B\hat{y}(x^*, \xi_\ell) + a) = 0. \tag{3.25}$$

By [13], it follows from (3.22) and (3.25) that there exists a diagonal matrix Λ_k with diagonal entries on $[0, 1]$ such that

$$0 = (I - \Lambda_k)(x^k - x^*) + \Lambda_k \left(A(x^k - x^*) - \sum_{\ell=1}^{\nu} (B\hat{y}_{\mu_k}(x^k, \xi_\ell) - B\hat{y}(x^*, \xi_\ell)) \right),$$

which implies that

$$\|x^k - x^*\| \leq \|(I - \Lambda_k + \Lambda_k A)^{-1} \Lambda_k\| \| \sum_{\ell=1}^{\nu} (B\hat{y}_{\mu_k}(x^k, \xi_\ell) - B\hat{y}(x^*, \xi_\ell))\|.$$

The above relation implies that $\{x^k\}$ is bounded since the righthand side is bounded. Then, for any accumulation point $\bar{x}$ of $\{x^k\}$, let $\mathcal{K}$ be the subsequence such that $\lim_{k\in\mathcal{K},k\to\infty} x^k = \bar{x}$. (3.23) is equivalently written as

$$\min(\hat{y}_{\mu_k}(x^k,\xi_\ell), M_\ell^{\mu_k}\hat{y}_{\mu_k}(x^k,\xi_\ell) + q(x^k,\xi_\ell)) = 0.$$

Taking the limit for the above equation in the subsequence $\mathcal{K}$, we have

$$\min(\bar{y}(\bar{x},\xi_\ell), M_\ell\bar{y}(\bar{x},\xi_\ell) + q(\bar{x},\xi_\ell)) = 0,$$

where without loss of generality let $\lim_{k\in\mathcal{K},k\to\infty} \hat{y}_{\mu_k}(x^k,\xi_\ell) = \bar{y}(\bar{x},\xi_\ell)$ (taking a subsequence of $\mathcal{K}$ if necessary). The last relation and (3.24) imply that

$$\lim_{k\in\mathcal{K},k\to\infty} \hat{y}_{\mu_k}(x^k,\xi_\ell) = \bar{y}(\bar{x},\xi_\ell) = \hat{y}(\bar{x},\xi_\ell).$$

Thus, by the above relation and taking the limit for (3.22) in $\mathcal{K}$, it yields that

$$\min(\bar{x}, A\bar{x} - \sum_{\ell=1}^{\nu} B\hat{y}(\bar{x},\xi_\ell) + a) = 0,$$

which implies that $\bar{x}$ solves (3.17). Then every accumulation point of $\{x^k\}$ is a solution of (3.17). Since the single-stage problem (3.17) has a unique solution x^*, we deduce that $\lim_{k\to\infty} x^k \to x^*$. $\square$

We have to point out that Theorem 3.6 is implementable in theory, but one may encounter some difficulties in practical numerical computations. Indeed, by Proposition 3.4, the Lipschitz constant $L_{H_{\mu_k}}$ of H_{μ_k} can be arbitrarily large as μ_k decreases to zero, which means that the step size of projection iteration (2.15) in PSNA will approach zero. In practical numerical implementations, the projection iteration (2.15) in PSNA will stagnate, which is undesirable. Another difficulty is that the computation of $\mathcal{H}_{\mu_k}$ is not well-defined since $I - \Lambda(x,\xi_\ell) + \Lambda(x,\xi_\ell)M_\ell^\mu$ may tend to singular as $\mu_k \to 0$.

To overcome the difficulties discussed above, we need to assume that $\hat{y}(\cdot, \xi_\ell)$ is Lipschitz continuous around the unique solution x^* of (3.17) for all ℓ, which also implies the Lipschitz continuity of H around x^*. Although $\hat{y}(\cdot, \xi_\ell)$ is not necessarily Lipschitz continuous on $\mathbb{R}^J_+$, it is reasonable to expect that it is Lipschitz continuous around some points. A sufficient condition to guarantee the local Lipschitz continuity of $\hat{y}(\cdot, \xi_\ell)$ around the solution is $x^* > 0$. Then, by Lemma 3.1, we know that the second stage problem has a unique solution $\hat{y}(x, \xi_\ell)$ around x^*. By [30, Theorem 3.2], it is known that $\hat{y}(\cdot, \xi_\ell)$ is also Lipschitz continuous around x^*. Denote by $\mathcal{B}(x^*, \delta)$ the open neighborhood on which each $\hat{y}(\cdot, \xi_\ell)$ is Lipschitz continuous. Denote the Lipschitz constant of H around x^* by $\bar{L}_H$. Since $x^k \to x^*$ by Theorem 3.6 and $\lim_{k\to\infty}$ $H_{\mu_k}(x) = H(x)$ for any $x \in \mathbb{R}^J_+$, there exists a finite positive integer $\bar{K}$ such that for all $k \geq \bar{K}$, $x^{k,0}$, $x^k \in \mathrm{cl}\mathcal{B}(x^*, \frac{1}{8}\delta)$ and H_{μ_k} is Lipschitz continuous on $\mathrm{cl}\mathcal{B}(x^*, \frac{1}{2}\delta)$ with Lipschitz constant $\bar{L}_{H_{\mu_k}}$ satisfying

$$|\bar{L}_{H_{\mu_k}} - \bar{L}_H| \leq \frac{1}{2}\bar{L}_H,$$

which implies that H_{μ_k} is Lipschitz continuous on $\mathrm{cl}\mathcal{B}(x^*, \frac{1}{2}\delta)$ with bounded Lipschitz constant for all $k \geq \bar{K}$. Then, the projection iteration (2.15) of PSNA will not stagnate and the computation of $\mathcal{H}_{\mu_k}$ does not encounter difficulties over $\mathrm{cl}\mathcal{B}(x^*, \frac{1}{2}\delta)$ as $\mu_k \to 0$. Indeed, for any $k \geq \bar{K}$, if we choose $\alpha_k = \frac{1}{2}\bar{L}^{-1}_{H_{\mu_k}}$, the sequence $\{x^{k,j}\}_{j=0}^\infty$ generated by the projection iteration (2.15) in PSNA is contained in $\mathrm{cl}\mathcal{B}(x^*, \frac{1}{8}\delta)$. More specifically, let $x^{k,j} \in \mathrm{cl}\mathcal{B}(x^*, \frac{1}{8}\delta)$. Since

$$\|\pi(x^{k,j}) - x^k\| = \|\pi(x^{k,j}) - \pi(x^k)\| \leq \|x^{k,j} - x^k\| + \alpha_k \bar{L}_{H_{\mu_k}} \|x^{k,j} - x^k\|$$

with $\pi(x) = \Pi_{\mathbb{R}^J_+}(x - \alpha_k H_{\mu_k}(x))$, it implies that $\pi(x^{k,j}) \in \mathrm{cl}\mathcal{B}(x^*, \frac{1}{2}\delta)$. By [19, Lemma 12.1.10], it is not hard to see that $x^{k,j+1} \in \mathrm{cl}\mathcal{B}(x^*, \frac{1}{8}\delta)$. Note that x^k is also the unique solution of $\mathrm{VI}(\mathrm{cl}\mathcal{B}(x^*, \frac{1}{8}\delta), H_{\mu_k})$. By the above discussion, the projection iteration (2.15) with starting point $x^{k,0} \in \text{cl}\mathcal{B}(x^*, \frac{1}{8}\delta)$ to solve (3.19) is equivalent to solving VI$(\text{cl}\mathcal{B}(x^*, \frac{1}{8}\delta), H_{\mu_k})$, which is globally convergent such that

$$\lim_{j\to\infty} x^{k,j} = x^k.$$

As $x^{k,j}$ is sufficiently close to x^k, the Newton iteration (2.14) is applicable and the superlinear convergence rate occurs.

3.3 Numerical experiments

In this section, we conduct numerical experiments to test the efficiency of rPSNA. First, we compare rPSNA with PHA for solving (3.16) with randomly generated data. Next, we apply rPSNA to solve the crude oil market share problem under the impact of the COVID-19 pandemic, in which the related parameters of (3.16) are determined using the real data from reliable sources.

3.3.1 Randomly generated problems

PHA can solve the LCP$(\mathbf{q}, \mathbf{M})$, since it is monotone in the sense of (2.4). The details of PHA for the two-stage SLCP are given as follows.

Algorithm 3.2. Progressive Hedging Algorithm (PHA) [40]

Step 0. *Given a point $x^0 \in \mathbb{R}^J$, let $x_\ell^0 = x^0 \in \mathbb{R}^J, y_\ell^0 \in \mathbb{R}^{2J}$ and $w_\ell^0 \in \mathbb{R}^J$, for $\ell = 1, \ldots, \nu$, such that $\frac{1}{\nu}\Sigma_{\ell=1}^{\nu} w_\ell^0 = 0$. Set the initial point $z^0 = (x^0, y_1^0, \ldots, y_\nu^0)^T$. Choose a step size $t > 0$. Set $k = 0$.*

Step 1. *For $\ell = 1, \ldots, \nu$, find $(\hat{x}_\ell^k, \hat{y}_\ell^k)$ that solves the LCP*

$$\begin{aligned} &0 \le x_\ell \perp A x_\ell - B y_\ell + a + w_\ell^k + t(x_\ell - x_\ell^k) \ge 0, \\ &0 \le y_\ell \perp B^T x_\ell + M_\ell y_\ell + \varrho_\ell + t(y_\ell - y_\ell^k) \ge 0. \end{aligned} \tag{3.26}$$

Let $\bar{x}^{k+1} = \frac{1}{\nu}\sum_{\ell=1}^{\nu} \hat{x}_\ell^k$, and for $\ell = 1, \ldots, \nu$, update

$$x_\ell^{k+1} = \bar{x}^{k+1},\ y_\ell^{k+1} = \hat{y}_\ell^k,\ w_\ell^{k+1} = w_\ell^k + t(\hat{x}_\ell^k - x_\ell^{k+1}),$$

to get point $z^{k+1} = (\bar{x}^{k+1}, y_1^{k+1}, \ldots, y_\nu^{k+1})^T$.

Step 2. *Set* $k := k+1$; *go back to Step 1.*

We first use randomly generated problems to compare the performance of rPSNA and PHA. The related parameters of problem (3.16) are generated as follows.

- c, a are uniformly drawn from $[e_J, 5e_J]$ and $[e_J, 10e_J]$, respectively. $r = 0.5e_J$.
- γ_ℓ, h_ℓ and ϱ_ℓ are uniformly distributed on $[0,1]$, $[e_J, 5e_J]$ and $[-e_{2J}, -10e_{2J}]$, respectively.

For rPSNA, the regularized parameter is set to $\mu_k \equiv 5 \times 10^{-9}$, the step size for the projection iteration (2.15) is $\alpha_k \equiv 0.05$, $\epsilon_k \equiv 0$ and $\eta = 0.9$. The step size t of PHA is set to $t = \sqrt{3J}$ according to the suggestion in [40]. We terminate the two algorithms when one of the following three criteria is met; that is, the number of iterations reaches 1000, or $\|\mathbf{z}^k - \mathbf{z}^{k-1}\| \le 10^{-5}$, or

$$\text{Res} := \|\min(\mathbf{M}\mathbf{z}^k + \mathbf{q}, \mathbf{z}^k)\| \le 10^{-5}.$$

We choose $J = 5, 10, 20$, and increase the sample size ν from 10 to 2000. The dimension of the corresponding LCP$(\mathbf{q}, \mathbf{M})$ ranges from 105 to 80020. For each J and ν, we randomly generated 10 problems following the descriptions above. PHA and rPSNA were used to solve these 10 problems. The results reported in Table 3.1 are the average number of iterations, CPU time (seconds), and residuals. Table 3.1 and Figure 3.1 show that rPSNA sucessfully solved the LCP$(\mathbf{q}, \mathbf{M})$ for all J and ν efficiently. In addition, for rPSNA, it is observed that the number of iterations remained almost unchanged when J and ν increased, and the CPU time was linearly increasing as ν increased. Although PHA can solve the problems with small J and ν, it failed to solve the problems with large J and ν within 1000 iterations. Figure 3.1(b) shows that the convergence rate of PHA was slow when iterates were close to

Table 3.1: Comparison of rPSNA and PHA

J	$\nu, J(2\nu+1)$	rPSNA Iter	Iter/N	CPU/sec	Res	PHA Iter	CPU/sec	Res
	10 105	4.80	2.50	0.03	1.83e-08	230.90	0.28	9.83e-06
	100 1005	4.50	3.00	0.13	6.14e-08	273.30	3.18	9.84e-06
5	500 5005	6.20	3.50	0.96	5.36e-07	295.40	16.74	9.91e-06
	1000 10005	5.90	3.40	1.79	1.08e-06	311.10	36.56	9.87e-06
	2000 20005	5.00	3.30	2.75	1.44e-06	298.80	71.87	9.85e-06
	10 210	5.00	2.50	0.05	2.66e-08	377.40	0.54	9.92e-06
	100 2010	7.40	3.20	0.48	8.44e-08	472.30	6.25	9.95e-06
10	500 10010	7.50	3.40	2.30	8.95e-07	504.10	33.19	9.92e-06
	1000 20010	7.50	3.60	4.53	8.76e-07	547.00	69.33	9.93e-06
	2000 40010	7.20	3.30	8.89	1.78e-06	534.40	142.99	9.93e-06
	10 420	10.90	3.00	0.29	3.56e-08	630.00	1.14	9.93e-06
	100 4020	10.00	3.00	2.25	1.12e-07	696.00	12.41	9.99e-06
20	500 20020	9.90	4.00	10.56	2.71e-07	1000.00	82.78	1.17e-05
	1000 40020	11.00	3.10	28.25	1.33e-06	995.00	168.98	9.97e-06
	2000 80020	10.20	3.50	45.59	1.20e-06	1000.00	347.06	1.41e-05

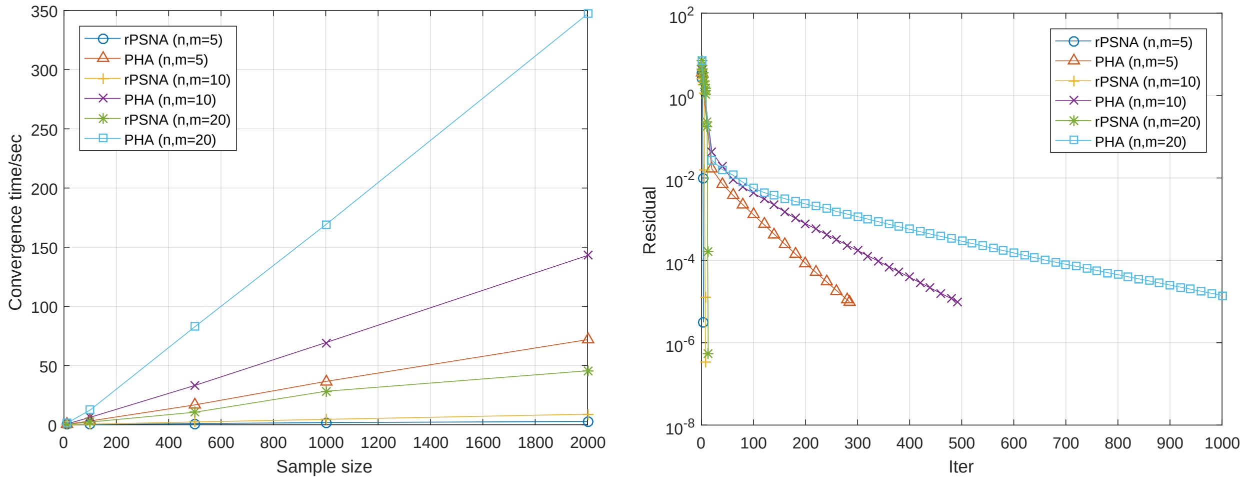
the solution. Overall, it is clear that rPSNA is more efficient than PHA in terms of the number of iterations as well as the CPU time.

3.3.2 Impacts of COVID-19 on oil market share

In this subsection, we use the market data of oil price, demand and market shares of 14 major oil producers to demonstrate the predicability of the two-stage SLCP model with adaptive parameters in the cost functions. Moreover, the model and the simulation results rationalize decisions made by major producers during the COVID-19 pandemic. The proposed rPSNA is used to calculate a solution of the model.

Modelling the oil market share problem as a two-stage stochastic game

We treat the oligopoly market of oil as a two-stage stochastic game where oil producers compete for profit by deciding the optimal production at the first stage of the game. In particular, we consider 15 clusters of oil producers in the following list as 15 non-cooperative agents in the game.



(a) CPU time with increasing sample sizes (b) Residuals for $\nu = 2000$

Figure 3.1: Comparison of rPSNA and PHA.

1 Saudi Arabia, 2 Russia, 3 USA, 4 Iraq, 5 China, 6 Canada, 7 United Arab Emirates (UAE), 8 Iran, 9 Kuwait, 10 Nigeria, 11 Mexico, 12 UK, 13 Venezuela, 14 Indonesia, 15 other.

Note that we treat OPEC+ members as individual agent rather than a cluster of cooperative producers. This is due to the fact that during oil shocks, clusters have less restrictive power over its members in deciding production quantities. As been represented in our model, leading producers with a cluster e.g., OPEC may choose different strategies over other cluster members.

Producer i makes a decision on its oil production quantity x_i, based on its predicted future characteristics of the oil market at a later time. We suppose that the trading occurs at the second stage where producer i supplies part if not all of its produced quantity in the first stage to generate revenue. We assume the spot price of the trading is uncertain at the time of production decision, and mainly depends on the demand supply relation at the second stage. To be more precise, supply quantity $v_i(\xi)$ of producer i is uncertain as being future event at the time of production decision.

In an oligopolistic market, the major suppliers are often considered the price setters, mainly through their control over market supply quantity. To simulate the effect of supply on price, we adapt a simple supply demand relation and express the inverse demand function as:

$$p(\tilde{\eta}(\xi), \xi) = \alpha(\xi) - \gamma(\xi) \sum_{j=1}^{J} v_j(\xi),$$

where $\alpha(\xi)$ is the benchmark price and $\gamma(\xi) > 0$ represents the negative effect on price if there is excessive supply of oil with respect to the observed demand at the second stage.

The benchmark price $\alpha(\xi)$ can be regarded as the intrinsic value of oil if the supply and demand matches. In the case when there exists excessive supply in the market, the price tends to decrease up to a scaling factor $\gamma(\xi)$, which can be learned from the market data. Alternatively, one can simply treat $\alpha(\xi)$ as the highest price in data set and the discount in price is due to excessive supply and is proportional to total supply in the market. Under deterministic setting, the inverse demand function may be over simplified but it is not the case in our data driven approach. Because the parameters in our inverse demand function are adaptive to data set, the model is more than capable of learning the complicated mixture responses of the market. The important feature is to ensure that the price drops when the oil supply surpasses the actual demand in the market.

In conventional oil production, producers need to explore oil fields before constructing the extraction site. After the oil is extracted, refinement and shipment require both time and labour not to mention the investment that involves. Yet, the nature of the oil production has been evolving with the technological advance which enables, e.g., extraction from oil sands. There exist fundamental differences in energy infrastructures between traditional producers and oil sands producers, e.g., US, Canada [28], and it is expected to be reflected at the first stage in choosing production strategies.

Here, we assume that the costs in both stages are quadratic as expressed in (3.5) and (3.6), where the parameters $c_i, a_i, h_i(\xi)$ and $\gamma(\xi)$ are to be learned from market data. We also include an extra term $(r_i \sum_{j=1}^{J} x_j)x_i$ in the first stage to represent the strategic concern of producer i in response to global production quantity $\sum_{j=1}^{J} x_j$. Note that we have no restriction on the sign of r_i and if it is positive it represents the fact that producer i is willing to decrease its production when the global production is high. Typical of such agents are often price setters, since the action would give a boost to the oil price and may be more profitable despite the production cuts. In reality, Russia's refusal to production cuts agreement triggered huge volatility of the market during the COVID-19 pandemic. To express it with our model, it means that Russia adapted its value of r_i to be negative given the market prediction back in March 2020. That is to say, Russia made a "squeeze" strategy and decided that if the price drop caused by over supplying is under control, the maintained market share is potentially more important for long term profitability.

We will use our numerical results to show that most decisions of producers were reasonable at the time from the prospective of the producers, since they had different tolerance levels on low oil prices, and they all attempted the best actions for their own benefits.

Parameter settings and numerical simulations

The market data used in our study are obtained from the following reliable sources.

(i) Monthly Oil Market Report published by OPEC[1].

[1] https://www.opec.org/opec web/en/publications/338.htm

This data set provides average daily production quantities of major oil-producing countries, and the percentages of their production quantities to the total production quantities are regarded as their market shares respectively.

(ii) Oil Price Dynamics Report[2], weekly by Federal Reserve Bank of New York.

This data set reports the crude oil price fluctuations due to the supply contribution, demand contribution and residual contribution (contribution other than the supply and demand contributions).

(iii) U.S. Energy Information Administration[3], weekly and daily spot price of Brent.

For the model (3.1)-(3.2) with (3.5)-(3.7), parameters in the first stage are taken as follows:

- c_i: This parameter represents the quadratic production cost of producer i. For the traditional producers, this contributes to the cost of exploration, site building and equipment setting up, etc. For oil sand producers, the financial cost can also be regarded as non-linear with respect to production quantity. However, all major producers should have similar scale of values to maintain profitable. In our simulation of in-sample and out-of-sample, we took

$$c_1 = 0.11/R_1,\ c_2 = 0.115/R_2,\ c_3 = 0.095/R_3,\ c_i = 0.1/R_i,\ i = 4, \dots, 15,$$

$$c_1 = 0.11/R_1',\ c_2 = 0.115/R_2',\ c_3 = 0.095/R_3',\ c_i = 0.1/R_i',\ i = 4, \dots, 15,$$

respectively, where R_i and R_i' are market shares of producer i in the current month and previous month for the simulation of 2019, and market shares of January 2020 and December 2019 for the simulation of 2020. The data of market shares is given in Tables 3.3-3.4.

[2] https://www.newyorkfed.org/research/policy/oil_price_dynamics_report

[3] https://www.eia.gov

Table 3.2: Values of r for January to July in 2020

2020	Jan	Feb	Mar	Apr	May	Jun	Jul
Saudi Arabia	0	0.01	0	-0.022	0	0.02	0.01
Russia	-0.01	0	0	-0.008	0.01	0.01	0.015
USA	-0.01	0	-0.02	-0.04	0	0	0
Iraq	0	0	0	-0.01	0	0	0.005
China	0	0	0.01	-0.01	0	0	0
Canada	0	0	-0.02	-0.03	0	0	0
UAE	0	0	-0.02	-0.05	0	0	0.005
Iran	0	0	-0.02	-0.045	0	0	0
Kuwait	0	0	-0.01	-0.045	0	0	0
Nigeria	0	0	-0.01	-0.08	-0.06	0	0
Mexico	0	0	0	-0.065	-0.045	0	0
UK	0	0	-0.1	-0.16	-0.08	-0.08	-0.05
Venezuela	0	0	-0.1	-0.23	-0.13	-0.03	0
Indonesia	0	0	-0.1	-0.23	-0.13	-0.06	-0.06
other	-0.01	0	0.005	0.005	-0.005	0	0

- a_i: This parameter represents the linear production cost of producer i. It is widely agreed that traditional producers have very low unit cost of oil production. As for the oil sand producers, the unit cost is much higher. In our simulation, we set

$$a_i = c_i,\ i = 1, 2, 4, 5, 7, \ldots, 15,\ \ a_3 = 6c_3,\ a_6 = 2c_6.$$

- r_i: This parameter represents producer i's response to the total of production by all producers. Before the pandemic, it was widely agreed that supply should be kept in accordance to the demand but little preference was taken till Russia's refusal to further production cuts. In our simulations, $r_i = 0$ for all producers in 2019, and r_i were given in Table 3.2 for different producers in 2020.

These are basic production cost parameters restricted by technological advance and complicated operations, and we do not expect them to change over short periods of time for all producers. For the purpose of forecasting current year production, these parameters are revised monthly taken based on the market share of the month before.

The stochastic parameters are the risk-adjusted spot price $\rho_i(\xi) = \alpha(\xi) - \beta_i(\xi)$ and $\Lambda_2(\xi) = \mathrm{diag}(h_i(\xi) + \gamma(\xi))$, where $\alpha(\xi)$ is the benchmark price, $\gamma(\xi)$ is the stochastic supply discount, and $h_i(\xi)$ and $\beta_i(\xi)$ are the supply cost coefficients.

For our experiments, we randomly choose $\zeta \in [0.05, 0.1]$, and let $h_i(\xi) = \beta_i(\xi) = \zeta \times a_i$ represent 5% to 10% of the unit production cost. The data (ii) gives the crude oil price fluctuations due to different factors of contributions, namely contributions of the demand ΔDemand, supply ΔSupply and residual ΔResidual. Then, the price fluctuation Δprice is expressed as follows:

$$\Delta\text{price} = \Delta\text{Demand} + \Delta\text{Supply} + \Delta\text{Residual}.$$

These contributions ΔDemand, ΔSupply and ΔResidual over certain period of time are uncertain. We assume that it can be described by random variable ξ with unknown distribution, written as $d(\xi)$, $\tilde{s}(\xi)$ and $\tilde{r}(\xi)$. For the purpose of numerical tests, we generate their empirical distributions from historical data and use them as an approximation to the unknown distributions of contributions. Recall that in our model the price is given by

$$p(\tilde{\eta}(\xi), \xi) = \alpha(\xi) - \gamma(\xi)\tilde{\eta}(\xi).$$

Then, for any realization of $\alpha^k(\xi_\ell)$ of the k-th[4] day, it corresponds to

$$\alpha^k(\xi_\ell) = p_0^k(1 + d^k(\xi_\ell) + \tilde{r}^k(\xi_\ell)),$$

where p_0^k is the known price of the prior day given in the data set (iii), and $d^k(\xi_\ell)$ and $\tilde{r}^k(\xi_\ell)$ are realizations taken from empirical distributions of $d(\xi)$ and $\tilde{r}(\xi)$, respectively. Then,

$$p^k(\tilde{\eta}^k(\xi_\ell), \xi_\ell) = \alpha^k(\xi_\ell) - \gamma^k(\xi_\ell)\tilde{\eta}^k(\xi_\ell),$$

[4] The superscript k is used to denote the order in time

where $p^k(\tilde{\eta}^k(\xi_\ell), \xi_\ell)$ is the known price of the k-th day given in the data set (iii) and $\tilde{\eta}^k(\xi_\ell)$ is the total supply of the k-th day taken from a uniformly distributed interval between 99% and 101% of the real daily supply given in the data set (i). It follows that, we can generate a set of data of stochastic supply discount $\gamma^k(\xi_\ell)$

$$\gamma^k(\xi_\ell) = \frac{|p^k(\tilde{\eta}^k(\xi_\ell), \xi_\ell) - \alpha^k(\xi_\ell)|}{\tilde{\eta}^k(\xi_\ell)},$$

where the absolute value $|\cdot|$ ensures that increasement in quantity has a negative influence on price. We choose the sample size $\nu = 800$ of random variable ξ for both in-sample and out-of-sample numerical simulations.

For long-term prediction (yearly market shares prediction), we refer interested readers to [25] for more details. Here, we focus on short-term prediction, namely the monthly in-sample and out-of-sample market shares. Table 3.3 and 3.4 give average of daily market shares of each month for producers from January 2019 to July 2020. Figures 3.2-3.4 display results for the recovered monthly market shares from January 2019 to July 2020. For each month, the first column is the real market share, while the second and third columns are the in-sample and out-sample recovered results, respectively. They show that our two-stage SLCP model recovers and predicts the short-term real market shares from January 2019 to July 2020 very well. Although global oil demand has been hit hard by COVID-19 and oil price has fell to historically low, our results show that a Nash equilibrium for the global oil market share during the COVID-19 pandemic can be expected.

Of particular interests are what had happened in March and April among Russia, Saudi Arabia and U.S.A. In particular, the Brent spot price fell from around \$70 per barrel to about \$50 since the identification of the pandemic and was believed to decrease further. If no actions of change were taken by the major producers, what could have happened is that the high-cost producers would be forced to cut

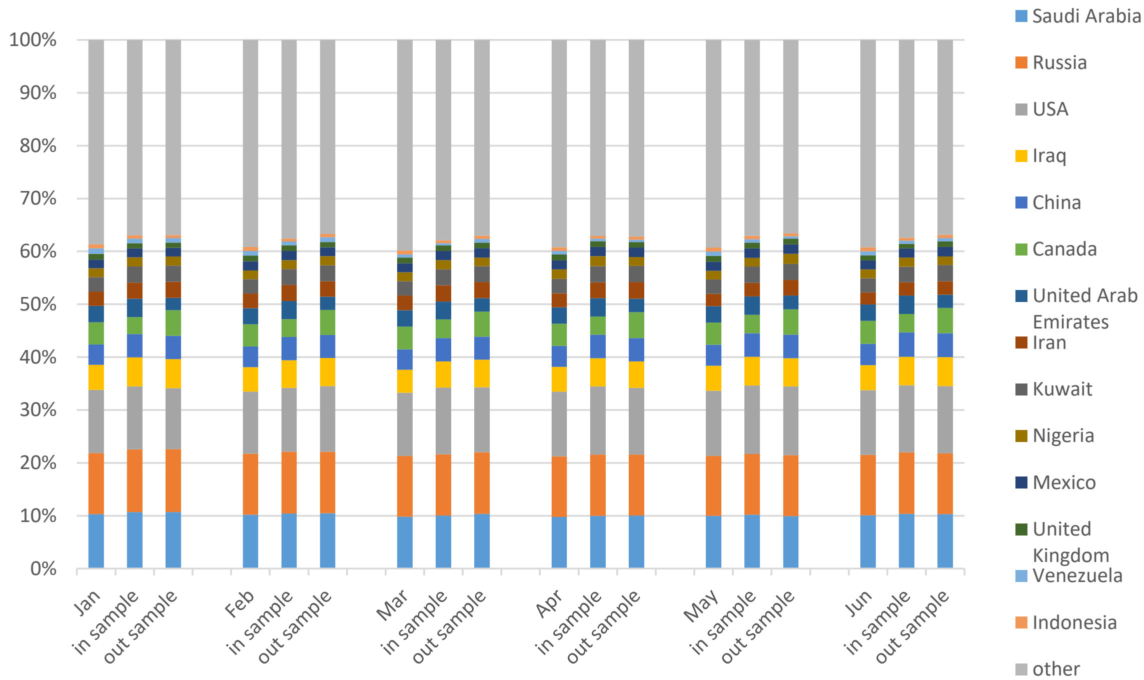


Figure 3.2: Real, in sample and out sample monthly market shares of Jan to Jun in 2019

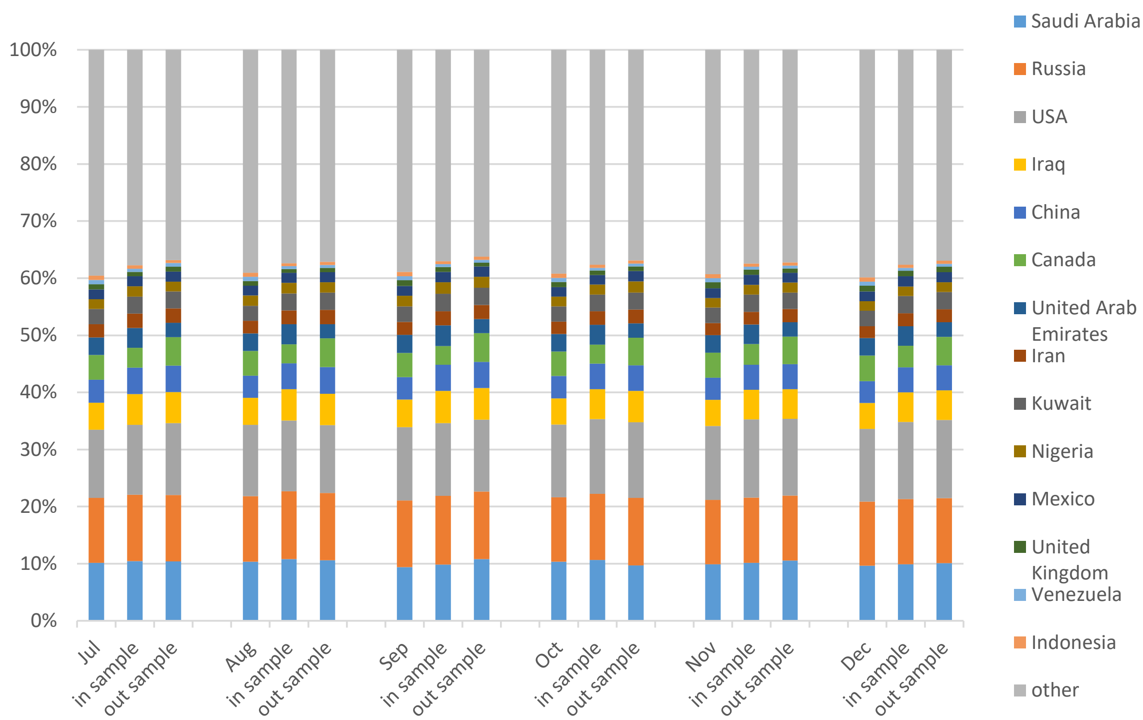


Figure 3.3: Real, in sample and out sample monthly market shares of Jul to Dec in 2019

Table 3.3: Average of daily market shares in each month of 2019

	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
Saudi Arabia	10.31	10.22	9.82	9.77	10.00	10.10	10.12	10.34	9.39	10.36	9.91	9.64
Russia	11.54	11.52	11.46	11.45	11.31	11.42	11.38	11.47	11.67	11.29	11.26	11.23
USA	11.95	11.75	11.96	12.27	12.33	12.25	11.98	12.49	12.83	12.74	12.93	12.76
Iraq	4.72	4.61	4.37	4.68	4.75	4.72	4.70	4.73	4.84	4.59	4.58	4.50
China	3.86	3.91	3.88	3.93	3.95	4.04	4.05	3.90	3.94	3.90	3.89	3.85
Canada	4.20	4.19	4.29	4.21	4.18	4.33	4.31	4.30	4.26	4.27	4.39	4.49
UAE	3.09	3.08	3.06	3.09	3.11	3.09	3.11	3.09	3.16	3.09	3.07	3.04
Iran	2.71	2.77	2.80	2.68	2.34	2.29	2.28	2.22	2.24	2.17	2.11	2.11
Kuwait	2.73	2.73	2.73	2.72	2.75	2.68	2.68	2.63	2.73	2.65	2.71	2.71
Nigeria	1.70	1.62	1.66	1.80	1.60	1.65	1.71	1.81	1.86	1.70	1.68	1.64
Mexico	1.63	1.72	1.70	1.70	1.69	1.70	1.69	1.70	1.75	1.66	1.71	1.70
UK	1.08	1.13	1.12	1.13	1.13	0.97	0.91	0.83	0.97	0.93	1.03	1.04
Venezuela	1.04	0.86	0.60	0.66	0.82	0.79	0.76	0.74	0.73	0.70	0.70	0.71
Indonesia	0.77	0.76	0.76	0.71	0.77	0.73	0.75	0.74	0.75	0.74	0.73	0.73
other	38.65	39.15	39.79	39.19	39.27	39.24	39.56	39.03	38.89	39.21	39.29	39.87

Table 3.4: Average of daily market shares in each month from January to July in 2020

	Jan	Feb	Mar	Apr	May	Jun	Jul
Saudi Arabia	9.72	9.75	10.23	11.57	9.44	8.75	9.47
Russia	11.26	11.30	11.33	11.42	10.44	10.34	10.10
USA	12.72	12.85	12.76	12.28	11.23	12.63	12.37
Iraq	4.25	4.53	4.54	4.49	4.59	4.30	4.23
China	3.88	3.87	3.92	3.90	4.32	4.58	4.37
Canada	4.36	4.41	4.42	3.74	3.66	4.23	4.34
UAE	2.98	2.99	3.54	3.88	2.72	2.72	2.74
Iran	2.10	2.05	2.01	1.96	2.17	2.26	2.18
Kuwait	2.66	2.66	2.90	3.13	2.42	2.44	2.43
Nigeria	1.61	1.67	1.88	1.75	1.65	1.74	1.68
Mexico	1.72	1.75	1.77	1.75	1.83	1.94	1.87
UK	0.97	0.95	1.02	1.00	1.05	1.20	1.05
Venezuela	0.73	0.75	0.73	0.71	0.76	0.46	0.44
Indonesia	0.73	0.72	0.72	0.72	0.79	0.82	0.79
other	40.32	39.75	38.24	37.68	42.93	41.59	41.95

production because they are more vulnerable in low price environment. On one hand, if the production cut is significant, the price will boost and the non-limited producers can be more profitable. On the other hand, if the oil price continues to fall, the optimal production decision has little to change for the traditional producers with low unit cost, e.g., Russian and Saudi Arabia. Indeed, with our model of two-stage stochastic game, the estimated market share given no strategic changes can be seen in accordance with the description above. It is, given the estimated market share, the most rational decision for Russia is to refuse the production cut agreement. Saudi Arabia, who is believed to have the lowest unit production cost, followed the strategy of Russia immediately by offering price discount and increased its production. For both countries, the low price environment has little effects in the sense of maintaining their market shares respectively. The same cannot be said for U.S.A., who has high

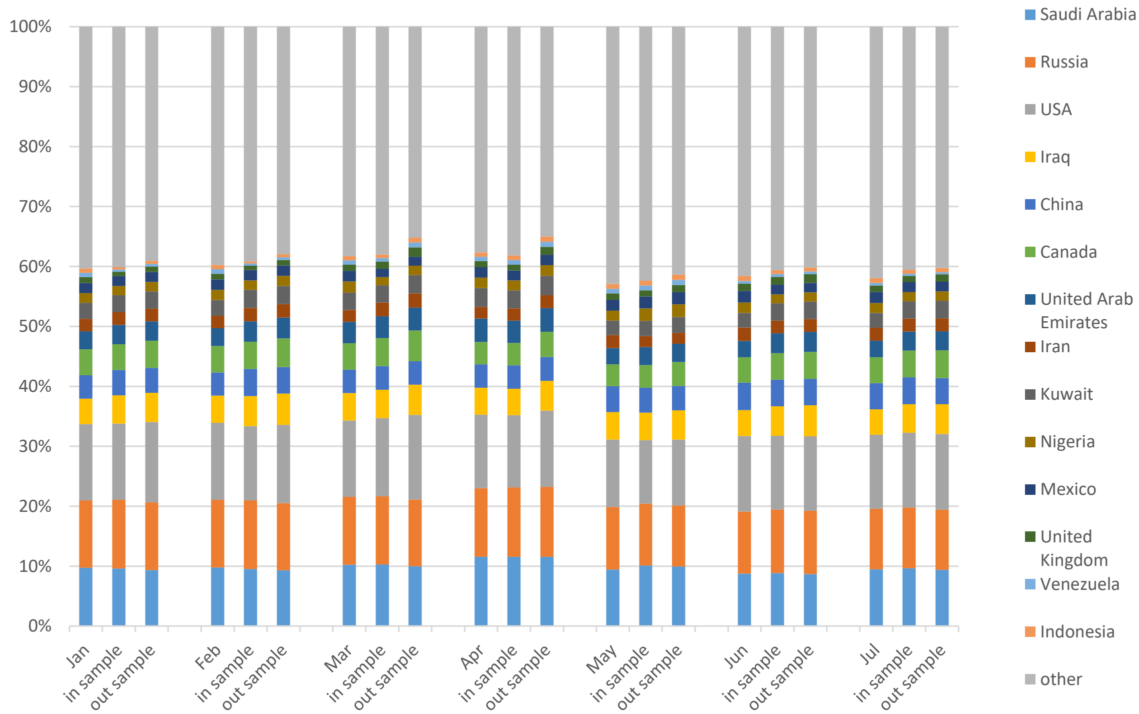


Figure 3.4: Real, in sample and out sample monthly market shares of Jan to July in 2020

cost in unit oil production, and is estimated to loss its market share if not to change its production strategy, presented by choosing non-zero r in our model. It is most apparent from Table 3.2, that all the major producers responded in choosing their strategies to increase production quantities. The reality was counter intuitive at first glance, but those were in fact all rational decisions and can be forecasted by our model.

Chapter 4

Applications of rPSNA to nonmonotone traffic assignment problems

In this chapter, we apply rPSNA developed in Chapter 3 to solve two types of nonmonotone traffic assignment problems. The first is the stochastic traffic assignment which is formulated as a two-stage SVI. The second is the stochastic dynamic traffic assignment which is formulated as a DLSCS. The discretization problem of the latter problem can be viewed as a special two-stage SVI in which the first stage problem is an mLCP and the second stage problem is an LCP.

4.1 Stochastic traffic assignment problems

In this section, we apply the two-stage SVI to formulate the stochastic user equilibrium problem with uncertain demands and capacities, which is an important class of problems in the stochastic traffic assignments The uncertainty for demands and link capacities can be caused by some unpredictable factors, such as adverse weather, road accidents and some other road conditions. The random variable ξ with a finite support set Ξ_ν is used to describe the uncertainty in demands and capacities.

First, we give definitions of notation in the stochastic traffic assignment.

- $\tilde{\mathcal{N}}, \mathcal{P}, \tilde{\mathcal{A}}, \mathcal{W}$: the node set, the path set, the link set and the origin destination (OD) pair set, respectively.

- $\mathcal{P}^w$: the set of paths joining the OD pair w with $\mathcal{P} = \bigcup_{w \in \mathcal{W}} \mathcal{P}^w$.

- $\Upsilon \in \mathbb{R}^{|\tilde{\mathcal{A}}| \times |\mathcal{P}|}$: the link-path incidence matrix where $\Upsilon_{ap} = 1$ if link a is on path p; otherwise, $\Upsilon_{ap} = 0$.

- $\Gamma \in \mathbb{R}^{|\mathcal{W}| \times |\mathcal{P}|}$: the OD-path incidence matrix where $\Gamma_{wp} = 1$ if path p connects OD pair w; otherwise, $\Gamma_{wp} = 0$.

- $h_p(\xi)$: the path travel flow on path p.

- $v_a(\xi)$: the link travel flow on link a, which satisfies $v(\xi) = \Upsilon h(\xi)$.

- $u_w(\xi)$: the minimum travel cost for OD pair w.

- $c_a(\xi)$: the link capacity of link a, which is a positive scalar.

- $d_w(\xi)$: the nonnegative demand function for OD pair w.

- $R_p(h(\xi), \xi)$: the travel cost function through path p.

- $r_a(v(\xi), \xi)$: the travel cost function through link a.

See Figure 4.1 for a simple illustration of the above notation. It is easy to see that $\hat{\mathcal{N}} = \{1, \ldots, 5\}$, $\tilde{\mathcal{A}} = \{1, \ldots, 7\}$, $\mathcal{P} = \{p_1, \ldots, p_6\}$ with $p_1 = 3 \to 7 \to 6$, $p_2 = 3 \to 1$, $p_3 = 4 \to 6$, $p_4 = 3 \to 7 \to 2$, $p_5 = 3 \to 5$ and $p_6 = 4 \to 2$, $\mathcal{W} = \{1 \to 4, 1 \to 5\}$. The link-path and OD-path incidence matrices read

$$\Upsilon = \begin{pmatrix} 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 1 \\ 1 & 1 & 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 & 0 \end{pmatrix}, \quad \Gamma = \begin{pmatrix} 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 \end{pmatrix}.$$

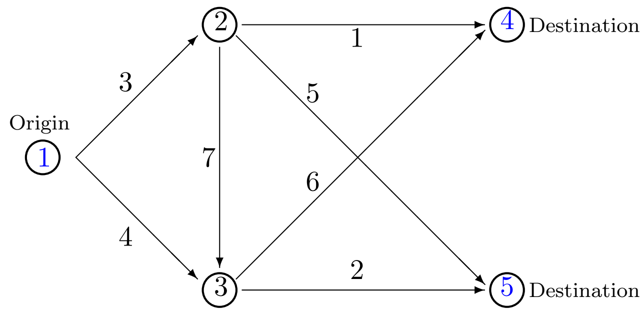


Figure 4.1: A network with 5 nodes, 7 links, 6 paths and 2 OD pairs

The travel demands should be satisfied for any realization of ξ, which is called the flow conservation [37]

$$\sum_{p\in\mathcal{P}^w} h_p(\xi) - d_w(\xi) = 0, \quad \forall\ w \in \mathcal{W}. \tag{4.1}$$

The path travel time function $R : \mathbb{R}^{|\mathcal{P}|} \times \mathbb{R}^d \to \mathbb{R}^{|\mathcal{P}|}$ is

$$R(h(\xi), \xi) = \Upsilon^T r(\Upsilon h(\xi), \xi),$$

where $r : \mathbb{R}^{|\tilde{\mathcal{A}}|} \times \mathbb{R}^d \to \mathbb{R}^{|\tilde{\mathcal{A}}|}$ is the generalized bureau of public road (GBPR) link travel time function as follows:

$$r_a(\Upsilon h(\xi), \xi) = t_a^0 \left(1.0 + 0.15 \left(\frac{v_a(\xi)}{c_a(\xi)}\right)^{n_a}\right),\ a \in \tilde{\mathcal{A}},$$

with t_a^0 and n_a being given positive numbers. Flows choose paths according to the famous Wardrop's principle [37]

$$0 \le h_p(\xi) \perp R_p(h(\xi), \xi) - u_w(\xi) \ge 0, \quad \forall\ w \in \mathcal{W} \text{ and } p \in \mathcal{P}^w. \tag{4.2}$$

The above complementarity condition implies that travel flows always choose the optimal path for their trips. Combining (4.1)-(4.2), it should hold that for any $\xi \in \Xi_\nu$

$$0 \le h(\xi) \perp R(h(\xi), \xi) - \Gamma^T u(\xi) \ge 0, \tag{4.3}$$

$$0 \le u(\xi) \perp \Gamma h(\xi) - d(\xi) \ge 0. \tag{4.4}$$

Let

$$\hat{D}_\xi = \{h \in \mathbb{R}^{|\mathcal{P}|} \mid \Gamma h - d(\xi) = 0, h \geq 0\}.$$

Thus, $\hat{D}_\xi$ is a bounded polyhedral set for any $\xi \in \Xi_\nu$. By [19, Proposition 1.4.8], if $d(\xi) > 0$ for any $\xi \in \Xi_\nu$, (4.3)-(4.4) can be equivalently formulated as an SVI finding $h(\xi) \in \hat{D}_\xi$ such that

$$(h' - h(\xi))^T R(h(\xi), \xi) \geq 0, \quad \forall\, h' \in \hat{D}_\xi, \quad \text{for any } \xi \in \Xi_\nu. \tag{4.5}$$

Let

$$D = \{x \in \mathbb{R}^{|\mathcal{P}|} \mid \Gamma x - \mathbb{E}[d(\xi)] = 0, x \geq 0\},$$

and $\bar{R} : \mathbb{R}^{|\mathcal{P}|} \to \mathbb{R}^{|\mathcal{P}|}$

$$\bar{R}(x) = \mathbb{E}[R(x, \xi)] = \Upsilon^T \mathbb{E}[r(\Upsilon x, \xi)].$$

To solve (4.5) with a fixed ξ, one can minimize the following optimization problem

$$\min_{x \in \hat{D}_\xi} \max\{(x - h(\xi))^T R(x, \xi) \mid h(\xi) \in \hat{D}_\xi\}, \tag{4.6}$$

which can be written as a two-stage optimization problem

$$\begin{aligned} &\min\ x^T R(x, \xi) + Q(x, \xi) \\ &\text{s.t. } x \in \hat{D}_\xi, \\ &\qquad Q(x, \xi) = \max\{-h(\xi)^T R(x, \xi) \mid h(\xi) \in \hat{D}_\xi\}. \end{aligned} \tag{4.7}$$

By duality of linear programming, the function Q can be expressed by

$$Q(x, \xi) = \min\{s(\xi)^T d(\xi) \mid \Gamma^T s(\xi) + R(x, \xi) \geq 0\}.$$

To calculate a here-and-now solution that does not depend on the realization of ξ, we solve the following two-stage stochastic program

$$\begin{aligned} &\min\ x^T \bar{R}(x) + \mathbb{E}[Q(x, \xi)] \\ &\text{s.t. } x \in D, \\ &\qquad Q(x, \xi) = \min\{s(\xi)^T d(\xi) \mid \Gamma^T s(\xi) + R(x, \xi) \geq 0\}, \quad \text{for any } \xi \in \Xi_\nu. \end{aligned} \tag{4.8}$$

Following [6, Example 2.3], we can obtain the first-order optimality condition of (4.8) as follows

$$-\left(\nabla \bar{R}(x)^T x + \bar{R}(x) - \mathbb{E}[\nabla R(x,\xi)^T \lambda(\xi)]\right) \in \mathcal{N}_D(x), \tag{4.9}$$

$$-\left[\begin{pmatrix} 0 & -\Gamma \\ \Gamma^T & 0 \end{pmatrix} y(\xi) + \begin{pmatrix} d(\xi) \\ R(x,\xi) \end{pmatrix}\right] \in \mathcal{N}_C(y(\xi)), \quad \text{any } \xi \in \Xi_\nu, \tag{4.10}$$

where the second stage problem is a mixed LCP with $C = \mathbb{R}^{|\mathcal{W}|} \times \mathbb{R}_+^{|\mathcal{P}|}$, and $y(\xi) = (s(\xi), \lambda(\xi))^T$ with $\lambda(\xi)$ being the multiplier of $\Gamma^T s(\xi) + R(x,\xi) \geq 0$.

Remark 4.1. *Problem* (4.9)-(4.10) *has the relatively complete recourse.*

It is known that $R(x,\xi) > 0$ for any $x \in D$ and $\xi \in \Xi_\nu$. Let $\bar{\lambda}(\xi) \geq 0$ with $\Gamma\bar{\lambda} \geq d(\xi)$ and $\bar{z}(\xi) = 0$. Thus, $(\bar{z}(\xi), \bar{\lambda}(\xi))$ is a feasible solution of the following LCP

$$0 \leq \begin{pmatrix} z(\xi) \\ \lambda(\xi) \end{pmatrix} \perp \begin{pmatrix} 0 & \Gamma \\ -\Gamma^T & 0 \end{pmatrix} \begin{pmatrix} z(\xi) \\ \lambda(\xi) \end{pmatrix} + \begin{pmatrix} -d(\xi) \\ R(x,\xi) \end{pmatrix} \geq 0. \tag{4.11}$$

Then, it is solvable by [17, Theorem 3.1.2].

Let $(z^*(x,\xi), \lambda^*(x,\xi))^T$ be an arbitrary solution of (4.11) *for any fixed $x \in D$ and $\xi \in \Xi_\nu$. If there is $w' \in \mathcal{W}$ such that $(\Gamma\lambda^*(x,\xi) - d(\xi))_{w'} > 0$, by the first complementarity condition in* (4.11)*, we have $z_{w'}(\xi) = 0$. Thus, $(R(x,\xi) - \Gamma^T z(\xi))_p = R_p(x,\xi) > 0$ for any $p \in \mathcal{P}_{w'}$. Then, we have $\lambda_p = 0$ for any $p \in P_{w'}$ by the second complementarity condition in* (4.11)*, which implies that $(\Gamma\lambda(\xi) - d(\xi))_{w'} = -d(\xi)_{w'} \leq 0$. This is a contradiction. Then $(-z^*(x,\xi), \lambda^*(x,\xi))^T$ is a solution of* (4.10) *for any fixed $x \in D$ and $\xi \in \Xi_\nu$.*

For any fixed x and ξ, the problem (4.10) admits a unique least-norm solution function. By substituting the least-norm solution function of (4.10) into the first stage problem (4.9), we can get the single-stage SVI formulation of (4.9)-(4.10). We can find a solution of the original two-stage problem by solving this single-stage problem, since D is a bounded polyhedral set. By the positive semi-definiteness of

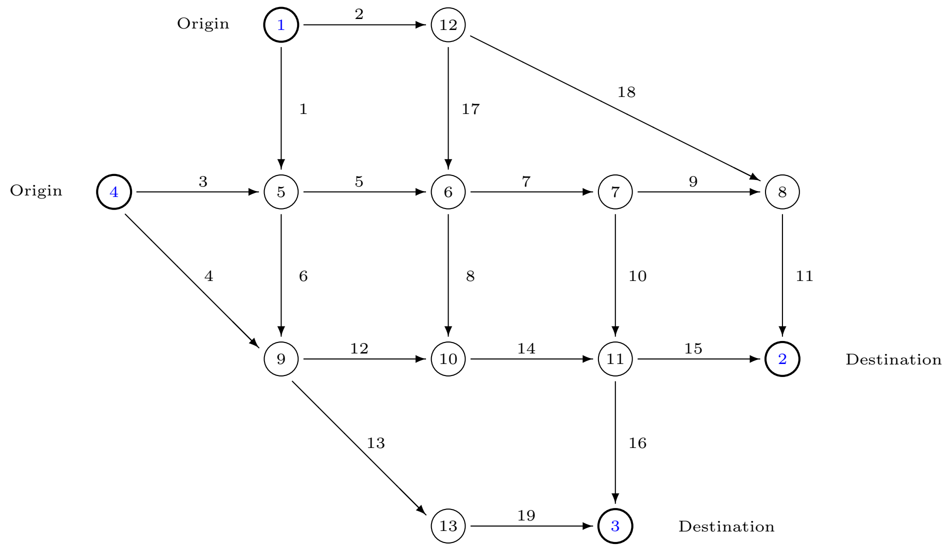


Figure 4.2: Nguyen and Dupuis network with 13 nodes, 19 links, 25 paths and 4 OD pairs.

the coefficient matrix of $y(\xi)$ in (4.10), we can use rPSNA developed in Chapter 3 to solve this problem efficiently.

Nguyen and Dupuis network [12] is used in our test, which has 13 nodes, 19 links, 25 paths and 4 OD pairs. All parameters for Network 4.2 are set as follows. The expected demands for OD pairs are $\mathbb{E}[d(\xi)] = [40, 80, 60, 45]^T$ in which $d(\xi)$ follows a beta distribution as $d(\xi) \sim \underline{\mathrm{d}} + \hat{d} \times beta(5, 1)$, where $\underline{\mathrm{d}} = [30, 70, 50, 35]^T$ and $\hat{d} = [12, 12, 12, 12]^T$. The free-flow travel time for links is $t_0 = 0.1 \times [7, 9, 9, 12, 3, 9, 5, 13, 5, 9, 9, 10, 9, 6, 9, 8, 7, 14, 11]^T$. The capacity $c(\xi)$ has the following three possible realizations with probability $p_1 = 0.5$, $p_2 = p_3 = 0.25$ and

$$c(\xi_1) = 10 \times [8, 3.2, 3.2, 8, 4, 3.2, 8, 2, 2, 2, 4, 4, 8, 6, 4, 4, 1.6, 2.3.8]^T,$$

$$c(\xi_2) = 10 \times [10, 4.4, 1.4, 10, 3, 4.4, 10, 2, 2, 4, 7, 7, 7, 7, 4, 3.5, 2.2, 4.4, 7]^T,$$

$$c(\xi_3) = 10 \times [4, 4, 2, 4, 4, 4, 4, 4, 4, 2, 4, 4, 2, 8, 8, 1, 2, 4, 2]^T.$$

We choose n_a from 1 to 5, and sample size $\nu = 500$ and 2000 to test the efficiency of rPSNA. Note that PHA fails to solve problem (4.9)-(4.10). The settings for rPSNA are $\mu_k \equiv 10^{-12}$, $\epsilon_k \equiv 0$, $\eta = 0.9$ and the step size for the projection iteration (2.15)

Table 4.1: Results of rPSNA for (4.9)-(4.10) with Nguyen and Dupuis network.

		rPSNA			
ν	n_α	Iter	Iter/N	CPU	Res
500	1	1.00	1.00	0.20	8.07e-10
	2	5.00	5.00	0.68	1.11e-07
	3	7.00	6.00	1.13	2.48e-07
	4	12.00	6.00	2.18	6.19e-07
	5	10.00	6.00	1.83	9.99e-07
2000	1	1.00	1.00	0.78	8.03e-10
	2	5.00	5.00	2.75	1.03e-07
	3	7.00	6.00	5.18	1.28e-07
	4	12.00	6.00	9.39	6.20e-07
	5	10.00	6.00	8.21	1.49e-06

is set to $\alpha = 0.1, 0.1, 0.05, 0.05, 0.05$ for $n_a = 1, 2, \dots, 5$, respectively. We terminate the algorithm if the number of iterations reaches 1000, or $\|x^{k+1} - x^k\| \le 10^{-6}$, or the residual is less than 10^{-5}. We reported numerical results of rPSNA for solving (4.9)-(4.10) in Table 4.1. One can see that rPSNA solved all problems with $n_a = 1, \dots, 5$ and $\nu = 500, 2000$ efficiently.

4.2 Stochastic dynamic traffic assignment problems

In this section, we apply rPSNA to solve the stochastic instantaneous dynamic user equilibrium (S-IDUE), a special problem in the stochastic dynamic traffic assignment, which involves some stochastic factors inevitably [37]. An S-IDUE model that considers the randomness of travelers' behavior, a generalization of the deterministic IDUE, can be found in [37]. Here, we focus on the stochasticity of OD demands and link capacities. We first give some notation as follows.

- $\tilde{\mathcal{N}}, \mathcal{P}, \tilde{\mathcal{A}}, \mathcal{W}$: the node set, the path set, the link set and the OD pair set, respectively.

- $\mathcal{P}^w$: the set of paths joining OD pair w with $\mathcal{P} = \bigcup_{w \in \mathcal{W}} \mathcal{P}^w$.

- $d^w(t, \xi)$: the nonnegative time-dependent travel demand profile for OD pair w.

- $c_{ij}(\xi)$: the capacity of link (i, j), which is a positive scalar.

- $q_{ij}(t)$: the expected queue length on link (i, j) at time t.

- $v_{ij}^p(t)$: the expected exit-flow rate on link (i, j) of path p at time t.

- $\underline{\mathrm{v}}_{ij}^p(t, \xi)$: the exit-flow rate on link (i, j) of path p at time t.

- $g_{ij}^p(t, \xi)$: the inflow rate on link (i, j) of path p at time t.

- $T^p(q, t, \xi)$: the instantaneous travel time traversing path p at time t.

- $\eta^w(t, \xi)$: the minimum travel time for OD pair w at time t.

An S-IDUE model is consisted of ODEs and complementarity system. The ODEs describe the dynamic change of the traffic flow on the network, while the complementarity condition is used to illustrate the choice of the optimal path and the constraint of the flow conservation. Here, we use the modified point-queue model for the flow dynamics with zero free-flow travel time [3]. The zero free-flow travel time means that traffic flows reach the end of the link and join the queue if any as soon as they depart from the origin (see [3, 32] for details about the point-queue model). On the other hand, we use the instantaneous travel time as the path choice criterion, which is also used in [4]. Our stochastic model can be viewed as a stochastic generalization of the deterministic model developed in [4]. Another difference is that we build our model based on the path formulation, while the authors in [4] used the link formulation. Both of the path and link formulations have their own advantages. We choose the path formulation since the coefficient matrix in the complementarity condition is positive semidefinite, which guarantees the existence of the weak solution of the model in the sense of Carathéodory [29] and the application of rPSNA.

For any $p \in \mathcal{P}$, we assume that it passes through nodes $i_0 \to i_1 \to \cdots \to i_{n_p}$ with n_p being the number of links for this path, i.e., $p = \{(i_0, i_1), (i_1, i_2), \dots, (i_{n_p-1}, i_{n_p})\}$ with (i_k, i_{k+1}) being a link connecting nodes k and $k+1$. We consider a special setting that different paths have no overlapped links, i.e.,

$$p_1 \cap p_2 = \emptyset, \ \forall \ p_1, p_2 \in \mathcal{P}.$$

The point-queue model to describe the queue dynamics at time t reads

$$\dot{q}_{i_k i_{k+1}}(t) \quad = \quad \mathbb{E}[g^p_{i_k i_{k+1}}(t,\xi) - c_{i_k i_{k+1}}(\xi)], \quad k = 0, 1, \dots, n_p - 1,$$

where $\dot{q}_{i_k i_{k+1}}(t)$ denotes the derivative of $q_{i_k i_{k+1}}$ with respect to t for the link (i_k, i_{k+1}). The meaning of the above formula is that the difference between the inflow rate and the exit-flow rate equals the rate of the queue length change. To guarantee the positiveness of the queue length, we modify it according to [3] as follows

$$\begin{aligned} \dot{q}_{i_k i_{k+1}}(t) &= u_{i_k i_{k+1}}(t) + \mathbb{E}[g^p_{i_k i_{k+1}}(t,\xi) - c_{i_k i_{k+1}}(\xi)], \\ 0 &\leq u_{i_k i_{k+1}}(t) \perp q_{i_k i_{k+1}}(t) \geq 0, \end{aligned} \tag{4.12}$$

where $u_{i_k i_{k+1}}(t)$ is a slack variable. The expected exit-flow rate on link (i_k, i_{k+1}) of $p \in \mathcal{P}$ is

$$v^p_{i_k i_{k+1}}(t) = \mathbb{E}[\underline{\mathrm{v}}^p_{i_k i_{k+1}}(t,\xi)] = \mathbb{E}[g^p_{i_k i_{k+1}}(t,\xi) - \dot{q}_{i_k i_{k+1}}(t)] = \mathbb{E}[c_{i_k i_{k+1}}(\xi) - u_{i_k i_{k+1}}(t)]. \tag{4.13}$$

From (4.12)-(4.13), it is known that $v^p_{i_k i_{k+1}}(t) = \mathbb{E}[c_{i_k i_{k+1}}(\xi)]$ if $q_{i_k i_{k+1}} > 0$. The instantaneous path travel time function $T^p(q,t,\xi)$ is the sum of each link's instantaneous travel time over the path, i.e.,

$$T^p(q,t,\xi) = \sum_{k=0}^{n_p-1} \left(\frac{q_{i_k i_{k+1}}(t)}{c_{i_k i_{k+1}}(\xi)} \right). \tag{4.14}$$

For S-IDUE, flows should follow a path choice strategy $g^p_{i_0 i_1}(t,\xi)$ enabling them to choose a path p with the minimum travel time $T^p(q,t,\xi)$, i.e.,

$$0 \le g^p_{i_0 i_1}(t,\xi) \perp T^p(q,t,\xi) - \eta^w(t,\xi) \ge 0, \qquad \text{for any } t \text{ and a.e. } \xi \in \Xi.$$

Moreover, the flow conservation condition should hold,

$$0 \le \eta^w(t,\xi) \perp \sum_{p \in \mathcal{P}^w} g^p_{i_0 i_1}(t,\xi) - d^w(t,\xi) \ge 0, \qquad \text{for any } t \text{ and a.e. } \xi \in \Xi.$$

Note that that $g^p_{i_0 i_1}(t,\xi)$ is also the path inflow rate of path p since there are no overlapped links between different paths.

Therefore, we have the following S-IDUE model:

$$\begin{cases} \dot{q}_{i_k i_{k+1}}(t) = u_{i_k i_{k+1}}(t) + \mathbb{E}[g^p_{i_k i_{k+1}}(t,\xi) - c_{i_k i_{k+1}}(\xi)], & (i_k, i_{k+1}) \in p,\, p \in \mathcal{P}, \quad (4.15a) \\ 0 \le u_{i_k i_{k+1}}(t) \perp q_{i_k i_{k+1}}(t) \ge 0, & (i_k, i_{k+1}) \in p,\, p \in \mathcal{P}, \quad (4.15b) \\ 0 \le g^p_{i_0 i_1}(t,\xi) \perp T^p(q,t,\xi) - \eta^w(t,\xi) \ge 0, & p \in \mathcal{P}^w, w \in \mathcal{W}, a.e.\ \xi \in \Xi, \quad (4.15c) \\ 0 \le \eta^w(t,\xi) \perp \sum_{p \in \mathcal{P}_w} g^p_{i_0 i_1}(t,\xi) - d^w(t,\xi) \ge 0, & p \in \mathcal{P}^w, w \in \mathcal{W}, a.e.\ \xi \in \Xi, \quad (4.15d) \\ q_{i_k i_{k+1}}(0) = q^0_{i_k i_{k+1}} > 0, & \text{for } (i_k, i_{k+1}) \in p,\, p \in \mathcal{P},\, t \in [0,T], (4.15e) \end{cases}$$

where $T > 0$ is the duration, $g^p_{i_0 i_1}(t,\xi)$ is the inflow rate for the first link of path p (also the path inflow rate for the path p), and the inflow rates for the successor links are defined as

$$g^p_{i_k i_k+1}(t,\xi) = \underline{\mathrm{v}}^p_{i_{k-1} i_k}(t,\xi) = g^p_{i_{k-1} i_k}(t,\xi) - \dot{q}_{i_{k-1} i_k}(t), \quad 1 \le k \le n_p - 1.$$

By the above definition, the expected inflow rate for link $(i_k, i_{k+1}), 1 \le k \le n_p - 1$ is

$$\begin{aligned} \mathbb{E}[g^p_{i_k i_{k+1}}(t,\xi)] = \mathbb{E}[\underline{\mathrm{v}}^p_{i_{k-1} i_k}(t,\xi)] &=: v^p_{i_{k-1} i_k}(t) \\ &= \mathbb{E}[c_{i_{k-1} i_k}(\xi)] - u_{i_{k-1} i_k}(t). \end{aligned} \tag{4.16}$$

To illustrate the model (4.15a)-(4.15e) clearly, we give the following descriptions

$$\begin{cases} \text{(4.15a) : the queue dynamic,} \\ \text{(4.15b) : the nonnegative constraint for queue,} \\ \text{(4.15c) : Wardrop's route choice principle,} \\ \text{(4.15d) : the flow conservation constraint,} \\ \text{(4.15e) : the initial condition.} \end{cases}$$

Let the link set $\tilde{\mathcal{A}} = \{l_1, \dots, l_{|\tilde{\mathcal{A}}|}\}$, the path set $\mathcal{P} = \{p_1, \dots . p_{|\mathcal{P}|}\}$, the OD pair set $\mathcal{W} = \{w_1, \dots, w_{|\mathcal{W}|}\}$. (4.15a)-(4.15e) can be equivalently written in a compact form as follows:

$$\begin{cases} \dot{q}(t) = u(t) + \mathbb{E}[B(g(t,\xi), \eta(t,\xi))^T - c(\xi)], & \text{(4.17a)} \\ 0 \le u(t) \perp q(t) \ge 0, & \text{(4.17b)} \\ 0 \le \begin{pmatrix} g_0(t,\xi) \\ \eta(t,\xi) \end{pmatrix} \perp \begin{pmatrix} 0 & -\Gamma^T \\ \Gamma & 0 \end{pmatrix} \begin{pmatrix} g_0(t,\xi) \\ \eta(t,\xi) \end{pmatrix} & \\ \qquad + \begin{pmatrix} \Upsilon(\xi)^T \\ 0 \end{pmatrix} q(t) + \begin{pmatrix} 0 \\ -d(t,\xi) \end{pmatrix} \ge 0, \ \text{for a.e. } \xi \in \Xi, & \text{(4.17c)} \\ q(0) = q^0 > 0, t \in [0,T], & \text{(4.17d)} \end{cases}$$

where

$$B = [I_{|\tilde{\mathcal{A}}| \times |\tilde{\mathcal{A}}|}, 0_{|\tilde{\mathcal{A}}| \times |\mathcal{W}|}],$$

$$q(t) = (q_{l_1}, q_{l_2}, \dots, q_{l_{|\tilde{\mathcal{A}}|}})^T, u(t) = (u_{l_1}, u_{l_2} \dots, u_{l_{|\tilde{\mathcal{A}}|}})^T,$$

$$g_0(t,\xi) = (g^{p_1}_{i_0 i_1}(t,\xi), g^{p_2}_{i_0 i_1}(t,\xi), \dots, g^{p_{|\mathcal{P}|}}_{i_0 i_1}(t,\xi))^T,$$

$$\eta(t,\xi) = (\eta^{w_1}(t,\xi), \eta^{w_2}(t,\xi), \dots, \eta^{w_{|\mathcal{W}|}}(t,\xi))^T,$$

$$d(t,\xi) = (d^{w_1}(t,\xi), d^{w_2}(t,\xi), \dots, d^{w_{|\mathcal{W}|}}(t,\xi))^T,$$

$\Gamma \in \mathbb{R}^{|\mathcal{W}| \times |\mathcal{P}|}$ is the OD-path incidence matrix and $\Upsilon(\xi) \in \mathbb{R}^{|\tilde{\mathcal{A}}| \times |\mathcal{P}|}$ is a matrix defined as

$$\Upsilon(\xi)_{ij} = \begin{cases} c_{ij}^{-1}(\xi) & \text{if the link } l_i \in p_j \\ 0 & \text{otherwise.} \end{cases}$$

Proposition 4.1. *The LCP* (4.17b)-(4.17c) *is solvable for any fixed* t, ξ *and* $q(t) > 0$. *Moreover, any solution* $(\hat{g}_0(t,\xi), \hat{\eta}(t,\xi),\ \hat{u}(t))^T$ *satisfies* $\sum_{p\in\mathcal{P}^w} \hat{g}_0^p(t,\xi) = d^w(t,\xi)$ *for all* $w \in \mathcal{W}$*; that is, the flow conservation condition is satisfied.*

Proof. Since the coefficient matrix of (4.17b)-(4.17c) is positive semidefinite, we only need to show that the feasible region of the LCP is nonempty by [17, Theorem 3.1.2]. Since $q(t) > 0$, we get by complementarity condition $\bar{u}(t) = 0$. For any $w \in \mathcal{W}$, by taking $\sum_{p\in\mathcal{P}^w} \bar{g}_0^p(t,\xi) \geq d^w(t,\xi)$ with $\bar{g}_0^p(t,\xi) \geq 0$ and $\bar{\eta}^w(t,\xi) = \min_{p\in\mathcal{P}^w} T^p(q(t),t,\xi)$, it is easy to verify that $(\bar{g}_0(t,\xi), \bar{\eta}(t,\xi),\ \bar{u}(t))^T$ is a feasible solution of the LCP (4.17b)-(4.17c) for any fixed t, ξ and $q(t) > 0$, which also implies its solvability.

To show the last assertion, let $(\hat{g}_0(t,\xi), \hat{\eta}(t,\xi),\ \hat{u}(t))^T$ be a solution. Suppose that $\sum_{p\in\mathcal{P}^w} \hat{g}_0^p(t,\xi) > d^{w'}(t,\xi) \geq 0$ for some $w' \in \mathcal{W}$. Then, there exists a $p' \in \mathcal{P}^{w'}$ such that $\hat{g}_0^{p'}(t,\xi) > 0$ and by the complementarity condition we have $\hat{\eta}^{w'}(t,\xi) = 0$. Using the complementarity condition again we get

$$T^{p'}(q(t),t,\xi) = \sum_{k=0}^{n_{p'}-1} q_{i_k i_{k+1}}(t)/c_{i_k i_{k+1}}(\xi) = \hat{\eta}^{w'}(t,\xi) > 0,$$

which is a contradiction. □

The model (4.17a)-(4.17d) is exactly a DLSCS studied in [29]. If we assume that the demand profile $d(t,\xi)$ is Lipschitz continuous with respect to t for a.e. ξ, and it is nonegative and uniformly bounded for a.e. $\xi \in \Xi$ and any $t \in [0,T]$, and $c(\xi) \geq \iota > 0$ for a.e. $\xi \in \Xi$. The boundedness assumption for $d(t,\xi)$ is natural, since demands are always finite in practice. From the above proposition, the solution set of the LCP (4.17b)-(4.17c) is nonempty and bounded with $q_0 > 0$ at $t = 0$ for a.e. ξ. Then, by [29, Theorem 2], there exists $T_0 > 0$ such that the model (4.17a)-(4.17d) admits a weak solution on $[0, T_0]$ in the sense of Carathéodory.

Numerically, we solve the discretization problem of (4.17a)-(4.17b). Give an i.i.d. sample $\{\xi_1, \ldots, \xi_\nu\}$ of ξ, and divide the time interval $[0, T]$ into $\mathcal{K}$ equal subintervals such that $t_{i+1} - t_i = h$ for all $i = 0, \ldots, \mathcal{K} - 1$, i.e.,

$$0 = t_0 < t_1 \ldots < t_K = T.$$

Then, the discretization problem of (4.17a)-(4.17d) reads

$$\begin{cases} -\begin{pmatrix} h^{-1}I & -I \\ I & 0 \end{pmatrix}\begin{pmatrix} q(t_i) \\ u(t_i) \end{pmatrix} + \frac{1}{\nu}\sum_{\ell=1}^{\nu} B\begin{pmatrix} g(t_i,\xi_\ell) \\ \eta(t_i,\xi_\ell) \end{pmatrix} \\ + \begin{pmatrix} h^{-1}q(t_{i-1}) - \nu^{-1}\sum_{\ell=1}^{\nu} c(\xi_\ell) \\ 0 \end{pmatrix} \in \mathcal{N}_{\mathbb{R}^{|\tilde{\mathcal{A}}|}\times\mathbb{R}_+^{|\tilde{\mathcal{A}}|}}((q(t_i)^T, u(t_i)^T)^T), \quad (4.18\text{a}) \\ 0 \le \begin{pmatrix} g_0(t_i,\xi_\ell) \\ \eta(t_i,\xi_\ell) \end{pmatrix} \perp \begin{pmatrix} 0 & -\Gamma^T \\ \Gamma & 0 \end{pmatrix}\begin{pmatrix} g_0(t_i,\xi_\ell) \\ \eta(t_i,\xi_\ell) \end{pmatrix} \\ \qquad + \begin{pmatrix} \Upsilon(\xi_\ell)^T \\ 0 \end{pmatrix} q(t_i) + \begin{pmatrix} 0 \\ -d(t_i,\xi_\ell) \end{pmatrix} \ge 0, \text{ for } \ell = 1,\ldots,\nu, (4.18\text{b}) \\ q(0) = q^0 > 0, \quad i = 1,\ldots,\mathcal{K}. \qquad (4.18\text{c}) \end{cases}$$

For each t_i, (4.18a)-(4.18c) can be viewed as a special two-stage SVI, where the first stage problem is an mLCP and the second stage is a monotone LCP for any fixed $q(t_i)$ and ξ. Then, rPSNA can be applied to solve this problem sequentially from $i = 1$ to $i = \mathcal{K}$.

Next, we give an example to illustrate our model; see the traffic network in Figure 4.3. The original node is the node 1 and the destination is the node 4. The three paths ($p_1 : 1 \to 2 \to 4$), ($p_2 : 1 \to 4$) and ($p_3 : 1 \to 3 \to 4$) do not intersect with each

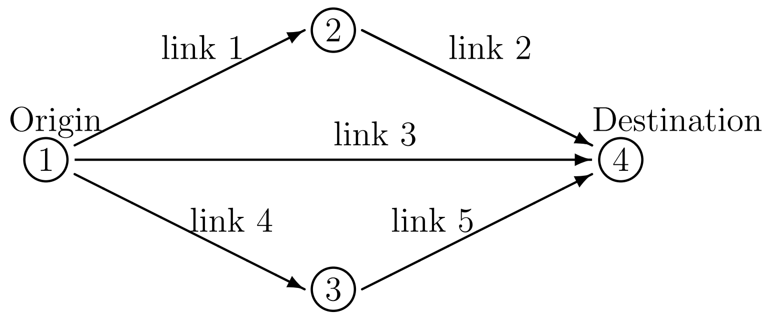


Figure 4.3: A network with 4 nodes, 5 links, 1 OD pair and 3 paths

other. Then, the formulation for this example is as follows:

$$
\begin{aligned}
&\dot{q}(t) = u(t) + \mathbb{E}[B(g(t,\xi), \eta(t,\xi))^T - c(\xi)],\\
&0 \le u(t) \perp q(t) \ge 0,\\
&0 \le \begin{pmatrix} g_0(t,\xi) \\ \eta(t,\xi) \end{pmatrix} \perp \begin{pmatrix} 0 & -e_3^T \\ e_3 & 0 \end{pmatrix} \begin{pmatrix} g_0(t,\xi) \\ \eta(t,\xi) \end{pmatrix}\\
&\qquad + \begin{pmatrix} \Upsilon^T(\xi) \\ 0 \end{pmatrix} q(t) - \begin{pmatrix} 0 \\ d(t,\xi) \end{pmatrix} \ge 0, \quad \text{for a.e. } \xi \in \Xi,\\
&q(0) = q^0, \qquad \forall t \in [0,T],
\end{aligned}
$$

where

$$g(t,\xi) = (g_{12}^{p_1}(t,\xi), g_{24}^{p_1}(t,\xi), g_{14}^{p_2}(t,\xi), g_{13}^{p_3}(t,\xi), g_{34}^{p_3}(t,\xi))^T,$$

$$g_0(t,\xi) = (g_{12}^{p_1}(t,\xi), g_{14}^{p_2}(t,\xi), g_{13}^{p_3}(t,\xi))^T,$$

$$u(t) = (u_{12}(t), u_{24}(t), u_{14}(t), u_{13}(t), u_{34}(t))^T,$$

$$q = (q_{12}(t), q_{24}(t), q_{14}(t), q_{13}(t), q_{34}(t))^T, e_3 = \begin{pmatrix} 1 & 1 & 1 \end{pmatrix},$$

$$c(\xi) = (c_{12}(\xi), c_{24}(\xi), c_{14}(\xi), c_{13}(\xi), c_{34}(\xi))^T,$$

$$\Upsilon(\xi) = \begin{pmatrix} c_{12}^{-1}(\xi) & 0 & 0 \\ c_{24}^{-1}(\xi) & 0 & 0 \\ 0 & c_{14}^{-1}(\xi) & 0 \\ 0 & 0 & c_{13}^{-1}(\xi) \\ 0 & 0 & c_{34}^{-1}(\xi) \end{pmatrix}, \quad B = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \end{pmatrix}.$$

4.2.1 Simulation results

We compare rPSNA developed in Chapter 3 with PHA to solve the discretization problem (4.18a)-(4.18c) with parameters chosen according to the Network 4.3. The settings for rPSNA are that $\mu_k \equiv 10^{-9}$, $\epsilon_k \equiv 0$, $\eta = 0.9$ and the step size for the projection iteration (2.15) is set to $\alpha_k \equiv 0.005$. The step size for PHA is set to $t = 1$. We terminate the algorithms if the number of iterations reaches 1000, or $\|x^{k+1} - x^k\| \leq 10^{-5}$, or the residual of the problem (4.18a)-(4.18c) is less than 10^{-5}. The demand profiles, capacities and initial queue lengths for links are given as follows:

$$\begin{aligned} d^{14}(t,\xi) &= \xi^{(1)} \max(0, 25 - \frac{90}{T^2}(t - T/2)^2), \\ c(\xi) &= \Lambda(\xi)(5, 10, 15, 8, 4)^T, \\ q^0 &= (8, 14, 10, 10, 6)^T, \end{aligned}$$

where $T = 35$ is the duration, $\xi = (\xi^{(1)}, \ldots, \xi^{(5)})$ is a random vector with each component $\xi^{(i)}$ uniformly distributed on $[1, 5]$ and $\Lambda(\xi) = \mathrm{diag}(\sqrt{\xi^{(1)}}, \ldots, \sqrt{\xi^{(5)}})$.

We chose the division of the time interval $\mathcal{K} = 500, 1000$ and the sample size $\nu = 10, 100, 1000$. For each fixed $\mathcal{K}$ and ν, we generated 20 problems with randomly chosen time instant t_i in $[0, 35]$, and $q(t_{i-1})$ is uniformly chosen from $[1, 10]$. The numerical results were reported in Table 4.2, in which the average number of iterations, CPU time and residuals for the two algorithms are provided. In addition, the success rates for the two algorithms for solving these problems are also given. Since (4.18a)-(4.18c) is not monotone in the sense of (2.4) for each t_i, both rPSNA and PHA are not guaranteed to solve it successfully. From Table 4.2, it is clear that rPSNA is promising for solving these problems compared with PHA.

Table 4.2: Comparison between rPSNA and PHA for Network 4.3.

		rPSNA				
$\mathcal{K}$	ν	Iter	Iter/N	CPU/sec	Res	success rate
500	10	2.50	2.25	0.04	1.83e-08	0.80
	100	4.85	4.08	0.57	4.88e-08	0.65
	1000	7.13	4.75	10.73	2.21e-07	0.80
1000	50	1.58	1.58	0.02	1.79e-08	0.60
	410	4.53	3.93	0.54	5.68e-08	0.75
	4010	6.25	5.92	6.02	1.88e-07	0.60

		PHA			
$\mathcal{K}$	ν	Iter	CPU/sec	Res	success rate
500	10	776.50	3.67	5.19e-06	0.10
	100	-	-	-	0.00
	1000	-	-	-	0.00
1000	10	-	-	-	0.00
	100	-	-	-	0.00
	1000	-	-	-	0.00

Chapter 5

Conclusions and future work

In this chapter, we summarize the conclusions of the thesis and discuss some possible future work.

5.1 Conclusions

- In Chapter 2, we have proposed a projection semismooth Newton algorithm (PSNA) for solving the two-stage SVI, which is based on solving its single-stage SVI formulation. The Lipschitz continuity, semismoothness, linear Newton approximation scheme and monotonicity properties of the single-stage SVI are discussed, which are essential for the development of PSNA. PSNA is a hybrid algorithm that combines the projection algorithm with the semismooth Newton algorithm. PSNA has the decomposition property just like PHA, which enables it to solve large-scale problems efficiently. The global convergence and the superlinear convergence rate have been established under suitable assumptions. Preliminary numerical results have illustrated that PSNA outperforms PHA for solving monotone problems. Moreover, PSNA is efficient for solving some nonmonotone problems.

- In Chapter 3, we have developed a regularized PSNA (rPSNA) for finding a solution of a new two-stage stochastic equilibrium model, in which the second stage problem is a monotone LCP for any fixed x and ξ. A regularized term is added to the second stage problem forcing it to be a strongly monotone LCP for any fixed x and ξ. PSNA is applied to solve the regularized model with the regularized parameter decreasing to zero, which we call rPSNA. We have proved that the sequence generated by rPSNA converges to the unique solution of the single-stage problem as the regularized parameter tends to zero. Numerically, rPSNA outperforms PHA for solving this two-stage stochastic equilibrium model with randomly generated data. In addition, by determining the related parameters of the model using the real data from the global crude oil market, the two-stage stochastic model can be used to describe the global crude oil market share under the impact of the COVID-19 pandemic. The proposed rPSNA is efficient in solving this real application problem and the solution obtained is used to explain and predict the global crude oil market share even under the influence of the COVID-19 pandemic.

- In Chapter 4, the proposed rPSNA is further applied to solve nonmonotone problems in the stochastic traffic assignment and stochastic dynamic traffic assignment. Preliminary numerical results have shown that rPSNA performs better than PHA for solving these traffic assignment problems.

5.2 Future work

Several possible future research directions for PSNA are as follows.

- For the general two-stage SVI, we assume the strong regularity condition for the second stage problem, which implies that the second stage problem has a unique solution for any fixed x and ξ. Note that, in the case that the second stage problem is a Z-matrix LCP for fixed x and ξ, the second stage problem may have multiple solutions. We show that the unique least-element solution can be substituted into the first stage problem to calculate a single-stage problem with desirable properties. Then, it is interesting to consider how to choose a specific solution function when the second stage problem is a general VI and has multiple solutions for any fixed x and ξ.

- To guarantee the global convergence of PSNA, we use the projection algorithm to globalize the semismooth Newton algorithm. The monotonicity of the single-stage problem plays an important role in the global convergence of the projection algorithm. It is interesting to consider if the projection algorithm can be replaced by the line search technique to achieve the global convergence. The main difficulty lies in the fact that the single-stage problem is an implicit and nonsmooth function. However, the advantage for using the line search technique is that the monotonicity assumption for the single-stage problem is unnecessary, which extends the corresponding algorithm to solve a broader class of problems.

Bibliography

[1] R. J. Aumann, *Integrals of set-valued functions*, Journal of Mathematical Analysis and Applications, 12 (1965), pp. 1–12.

[2] X. J. Ban, H. X. Liu, M. C. Ferris, and B. Ran, *A link-node complementarity model and solution algorithm for dynamic user equilibria with exact flow propagations*, Transportation Research Part B: Methodological, 42 (2008), pp. 823–842.

[3] X. J. Ban, J.-S. Pang, H. X. Liu, and R. Ma, *Continuous-time point-queue models in dynamic network loading*, Transportation Research Part B: Methodological, 46 (2012), pp. 360–380.

[4] X. J. Ban, J.-S. Pang, H. X. Liu, and R. Ma, *Modeling and solving continuous-time instantaneous dynamic user equilibria: a differential complementarity systems approach*, Transportation Research Part B: Methodological, 46 (2012), pp. 389–408.

[5] X. Chen and M. Fukushima, *Expected residual minimization method for stochastic linear complementarity problems*, Mathematics of Operations Research, 30 (2005), pp. 1022–1038.

[6] X. Chen, T. K. Pong, and R. J.-B. Wets, *Two-stage stochastic variational inequalities: an ERM-solution procedure*, Mathematical Programming, 165 (2017), pp. 71–111.

[7] X. Chen, L. Qi, and D. Sun, *Global and superlinear convergence of the smoothing Newton method and its application to general box constrained variational inequalities*, Mathematics of Computation, 67 (1998), pp. 519–540.

[8] X. Chen, A. Shapiro, and H. Sun, *Convergence analysis of sample average approximation of two-stage stochastic generalized equations*, SIAM Journal on Optimization, 29 (2019), pp. 135–161.

[9] X. Chen, H. Sun, and R. J.-B. Wets, *Regularized mathematical programs with stochastic equilibrium constraints: estimating structural demand models*, SIAM Journal on Optimization, 25 (2015), pp. 53–75.

[10] X. Chen, H. Sun, and H. Xu, *Discrete approximation of two-stage stochastic and distributionally robust linear complementarity problems*, Mathematical Programming, 177 (2019), pp. 255–289.

[11] X. Chen and Z. Wang, *Computational error bounds for a differential linear variational inequality*, IMA Journal of Numerical Analysis, 32 (2012), pp. 957–982.

[12] X. Chen, R. J.-B. Wets, and Y. Zhang, *Stochastic variational inequalities: residual minimization smoothing sample average approximations*, SIAM Journal on Optimization, 22 (2012), pp. 649–673.

[13] X. Chen and S. Xiang, *Perturbation bounds of P-matrix linear complementarity problems*, SIAM Journal on Optimization, 18 (2008), pp. 1250–1265.

[14] X. Chen and S. Xiang, *Newton iterations in implicit time-stepping scheme for differential linear complementarity systems*, Mathematical Programming, 138 (2013), pp. 579–606.

[15] X. Chen and Y. Ye, *On homotopy-smoothing methods for box-constrained variational inequalities*, SIAM Journal on Control and Optimization, 37 (1999), pp. 589–616.

[16] F. H. Clarke, *Optimization and Nonsmooth Analysis*, SIAM Publisher, Philadelphia, 1990.

[17] R. W. Cottle, J.-S. Pang, and R. E. Stone, *The Linear Complementarity Problem*, Academic Press, Boston, MA, 1992.

[18] A. Ehrenmann and Y. Smeers, *Generation capacity expansion in a risky environment: a stochastic equilibrium analysis*, Operations Research, 59 (2011), pp. 1332–1346.

[19] F. Facchinei and J.-S. Pang, *Finite-Dimensional Variational Inequalities and Complementarity Problems*, Springer-Verlag, New York, 2003.

[20] M. C. Ferris and J.-S. Pang, *Engineering and economic applications of complementarity problems*, SIAM Review, 39 (1997), pp. 669–713.

[21] S. A. Gabriel and J. Fuller, *A benders decomposition method for solving stochastic complementarity problems with an application in energy*, Computational Economics, 35 (2010), pp. 301–329.

[22] S. A. Gabriel and J. J. Moré, *Smoothing of mixed complementarity problems*, Complementarity and Variational Problems: State of the Art, 92 (1997), pp. 105–116.

[23] G. Gürkan, A. Yonca Özge, and S. M. Robinson, *Sample-path solution of stochastic variational inequalities*, Mathematical Programming, 84 (1999), pp. 313–333.

[24] P. T. Harker and J.-S. Pang, *A damped Newton method for the linear complementarity problem*, Lectures in Applied Mathematics, 26 (1990), pp. 256–284.

[25] J. Jiang, Y. Shi, X. Wang, and X. Chen, *Regularized two-stage stochastic variational inequalities for Cournot-Nash equilibrium under uncertainty*, Journal of Computational Mathematics, 37 (2019), pp. 813–842.

[26] A. Kannan, U. V. Shanbhag, and H. M. Kim, *Addressing supply-side risk in uncertain power markets: stochastic Nash models, scalable algorithms and error analysis*, Optimization Methods and Software, 28 (2013), pp. 1095–1138.

[27] J. Kyparisis, *Solution differentiability for variational inequalities*, Mathematical Programming, 48 (1990), pp. 285–301.

[28] E. F. Lazzaroni, M. Elsholkami, I. Arbiv, E. Martelli, A. Elkamel, and M. Fowler, *Energy infrastructure modeling for the oil sands industry: current situation*, Applied Energy, 181 (2016), pp. 435–445.

[29] J. Luo, X. Wang, and Y. Zhao, *Convergence of discrete approximation for differential linear stochastic complementarity systems*, Numerical Algorithms, 87 (2021), pp. 223–262.

[30] O. L. Mangasarian and T.-H. Shiau, *Lipschitz continuity of solutions of linear inequalities, programs and complementarity problems*, SIAM Journal on Control and Optimization, 25 (1987), pp. 583–595.

[31] J.-S. Pang, *Newton's method for B-differentiable equations*, Mathematics of Operations Research, 15 (1990), pp. 311–341.

[32] J.-S. Pang, L. Han, G. Ramadurai, and S. Ukkusuri, *A continuous-time linear complementarity system for dynamic user equilibria in single bottleneck traffic flows*, Mathematical Programming, 133 (2012), pp. 437–460.

[33] J.-S. Pang, S. Sen, and U. V. Shanbhag, *Two-stage non-cooperative games with risk-averse players*, Mathematical Programming, 165 (2017), pp. 235–290.

[34] J.-S. Pang and D. E. Stewart, *Differential variational inequalities*, Mathematical Programming, 113 (2008), pp. 345–424.

[35] L. Qi, *Convergence analysis of some algorithms for solving nonsmooth equations*, Mathematics of Operations Research, 18 (1993), pp. 227–244.

[36] L. Qi and J. Sun, *A nonsmooth version of Newton's method*, Mathematical Programming, 58 (1993), pp. 353–367.

[37] B. Ran and D. Boyce, *Modeling Dynamic Transportation Networks: An Intelligent Transportation System Oriented Approach*, Springer Science & Business Media, 2012.

[38] U. Ravat and U. V. Shanbhag, *On the characterization of solution sets of smooth and nonsmooth convex stochastic Nash games*, SIAM Journal on Optimization, 21 (2011), pp. 1168–1199.

[39] S. M. Robinson, *Strongly regular generalized equations*, Mathematics of Operations Research, 5 (1980), pp. 43–62.

[40] R. T. Rockafellar and J. Sun, *Solving monotone stochastic variational inequalities and complementarity problems by progressive hedging*, Mathematical Programming, 174 (2019), pp. 453–471.

[41] R. T. Rockafellar and R. J.-B. Wets, *Scenarios and policy aggregation in optimization under uncertainty*, Mathematics of Operations Research, 16 (1991), pp. 119–147.

[42] R. T. Rockafellar and R. J.-B. Wets, *Stochastic variational inequalities: single-stage to multistage*, Mathematical Programming, 165 (2017), pp. 331–360.

[43] A. Ruszczyński and A. Shapiro, *Optimality and duality in stochastic programming*, Handbooks in Operations Research and Management Science, 10 (2003), pp. 65–139.

[44] A. Shapiro, D. Dentcheva, and A. Ruszczyński, *Lectures on Stochastic Programming: Modeling and Theory*, SIAM, Philadelphia, 2009.

[45] H. Sun, C.-L. Su, and X. Chen, *SAA-regularized methods for multiproduct price optimization under the pure characteristics demand model*, Mathematical Programming, 165 (2017), pp. 361–389.

[46] C. Zhang and X. Chen, *Stochastic nonlinear complementarity problem and applications to traffic equilibrium under uncertainty*, Journal of Optimization Theory and Applications, 137 (2008), pp. 277–295.

[47] C. Zhang, X. Chen, and A. Sumalee, *Robust Wardrops user equilibrium assignment under stochastic demand and supply: expected residual minimization approach*, Transportation Research Part B: Methodological, 45 (2011), pp. 534–552.

[48] M. Zhang, J. Sun, and H. Xu, *Two-stage quadratic games under uncertainty and their solution by progressive hedging algorithms*, SIAM Journal on Optimization, 29 (2019), pp. 1799–1818.