



Two-Stage Stochastic Variational Inequalities: Theory, Algorithms and Applications

Hai-Lin Sun¹ · Xiao-Jun Chen²

Received: 10 February 2019 / Revised: 30 August 2019 / Accepted: 3 September 2019
© The Author(s) 2019

Abstract

The stochastic variational inequality (SVI) provides a unified form of optimality conditions of stochastic optimization and stochastic games which have wide applications in science, engineering, economics and finance. In the recent two decades, one-stage SVI has been studied extensively and widely used in modeling equilibrium problems under uncertainty. Moreover, the recently proposed two-stage SVI and multistage SVI can be applied to the case when the decision makers want to make decisions at different stages in a stochastic environment. The two-stage SVI is a foundation of multistage SVI, which is to find a pair of “here-and-now” solution and “wait-and-see” solution. This paper provides a survey of recent developments in analysis, algorithms and applications of the two-stage SVI.

Keywords Two-stage stochastic variational inequality · Two-stage stochastic complementary problem · Two-stage stochastic games

Mathematics Subject Classification 90C15 · 90C33

Xiao-Jun Chen was partially supported by Hong Kong Research Grant Council PolyU (No.153001/18P). Hai-Lin Sun was partially supported by the National Natural Science Foundation of China (Nos. 11871276 and 11571178).

✉ Xiao-Jun Chen
xiaojun.chen@polyu.edu.hk

Hai-Lin Sun
mathhlsun@163.com

¹ School of Mathematical Sciences, Jiangsu Key Laboratory for NSLSCS, Nanjing Normal University, Nanjing 210023, China

² Department of Applied Mathematics, The Hong Kong Polytechnic University, Hong Kong, China

1 Introduction

The variational inequality (VI) represents the first-order optimality conditions of optimization problems and models equilibrium problems, which plays a key role in optimization and operations research. The definition of VI is as follows.

Definition 1 (*Deterministic VI*) Given a nonempty closed convex set $X \subset \mathbb{R}^n$ and a single-valued mapping $F : X \rightarrow \mathbb{R}^n$, the variational inequality problem, denoted by $VI(X, F)$, is to find an $x^* \in X$ such that

$$0 \in F(x^*) + \mathcal{N}_X(x^*), \quad (1.1)$$

where $\mathcal{N}_X(x^*)$ denotes the normal cone of X at x^* . Furthermore, if X is a cone and X^* is its dual, defined as $X^* := \{h : h^T v \geq 0, \forall v \in X\}$, then the complementarity problem, denoted by $CP(X, F)$, requires an $x^* \in X$ such that $X \ni x^* \perp F(x^*) \in X^*$, where $u \perp v$ means that $u^T v = 0$.

In the case when $X = \mathbb{R}^n$, (1.1) reduces to the system of nonlinear equations

$$0 = F(x^*).$$

When $X = \mathbb{R}_+^n$, (1.1) reduces to the nonlinear complementarity problem (NCP)

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

The deterministic VI has wide applications in the operations research, including convex optimization problems, convex Nash games over continuous strategy sets and economic equilibrium problems. For more details, see [1,2].

However, in many real applications in finance, management, engineering and science, the decision makers have to make sequential decisions in an uncertain environment. In such situations, the deterministic VI may not be suitable. Motivated by those applications, one-stage, two-stage and multistage SVIs arouse the attention of scholars. One-stage SVI has been investigated for many years [3,4]. But the investigation of multistage SVI and even two-stage SVI has just begun [5]. In this paper, starting from one-stage SVI, we will introduce the model, the motivation and recent progress of two-stage SVI (and extend to multistage SVI), including theoretical results, algorithms and applications.

The organization of the paper is as follows. Starting from one-stage SVI, the motivations, the models and the properties of the two-stage SVI (the multistage SVI) are introduced in Sect. 2. In Sect. 3, we introduce several approximation methods and algorithms for different models of two-stage SVI (multistage SVI). Two applications of two-stage SVI are shown in Sect. 4. Section 5 gives several final remarks about the challenge of this research area.

Through out this paper, we use $\xi : \Omega \rightarrow \mathbb{R}^d$ to denote a random vector in probability space (Ω, \mathcal{F}, P) with support set $\mathcal{E} \subset \mathbb{R}^d$. For function $f(\cdot, \cdot) : \mathbb{R}^n \times \mathbb{R}^m \rightarrow \mathbb{R}$, $\partial_x f(x, y)$ denotes the Clarke subdifferential of f w.r.t. x .

2 Two-Stage Stochastic Variational Inequalities: Modeling and Analysis

In this section, we will give the motivation and model of two-stage stochastic variational inequalities and introduce the analysis in the literature. We will start the model of one-stage SVI firstly and then turn to two-stage SVI.

2.1 From One-Stage SVI to Two-Stage SVI

As a stochastic generalization of the deterministic VI under uncertainties, one-stage SVI has been investigated deeply and applied widely. However, a good formulation of a VI in a stochastic environment is not straightforward. We use the following one-stage stochastic game to introduce the motivation of different one-stage SVI formulations firstly.

Example 1 (Motivation of one-stage SVI) We consider a duopoly market where two firms compete to supply a homogeneous product (or service) noncooperatively in future. They need to make a decision on the quantity of production based on the other firm’s decision in an uncertain environment.

The market demand in future is characterized by a random inverse demand function $p(q, \xi(\omega))$, where $p(q, \xi(\omega))$ is the market price and q is the total supply to the market. Specifically, for each realization of the random vector ξ , we obtain a different inverse demand function $p(\cdot, \xi)$. The uncertainty in the inverse demand function is then characterized by the distribution of the random variable ξ . Firm i ’s cost function for producing (supplying) a quantity of y_i in future is $H_i(y_i, \xi)$, $i = 1, 2$ with limit capacity c_i . We assume $H_i(y_i, \xi)$ is twice continuously differentiable, $H'_i(y_i, \xi) \geq 0$ and $H''_i(y_i, \xi) \geq 0$ for $y_i \geq 0$, $p(q, \xi)$ is twice continuously differentiable in q and $p'_q(q, \xi) < 0$ and $p'_q(q, \xi) + qp''_{qq}(q, \xi) \leq 0$ for $q \geq 0$ and $\xi \in \mathcal{E}$.

We further assume each firm aims to maximize their profit and consider two situations: 1. The decision makers can make decision after they observe the further uncertainty; 2. the decision makers make decision before they observe the further uncertainty.

- For situation 1, to maximize each firm’s profit, they need to find $(y_1^*(\cdot), y_2^*(\cdot))$ such that it solves the following problem:

$$\begin{aligned} \max_{y_i(\xi)} \quad & p(y_i(\xi) + y_{-i}^*(\xi), \xi)y_i(\xi) - H_i(y_i(\xi), \xi) \\ \text{s.t.} \quad & 0 \leq y_i(\xi) \leq c_i, \end{aligned} \tag{2.1}$$

for almost every (a.e.) $\xi \in \mathcal{E}$, where y_{-i} is for decision variable of the firm(s) other than i . Moreover, we can write down the Karush–Kuhn–Tucker (KKT) conditions for (2.1) as follows: for a.e. $\xi \in \mathcal{E}$

$$\begin{aligned} 0 &\leq \begin{pmatrix} y_i(\xi) \\ \mu_i(\xi) \end{pmatrix} \\ \perp &\begin{pmatrix} -p(y_i(\xi) + y_{-i}(\xi), \xi) - y_i(\xi)p'_q(y_i(\xi) + y_{-i}(\xi), \xi) + H'_i(y_i(\xi), \xi) + \mu_i(\xi) \\ c_i - y_i(\xi) \end{pmatrix} \geq 0, \end{aligned} \tag{2.2}$$

and (2.2) is a “wait-and-see” model.

- For situation 2, we consider the expected residual minimization (ERM) formulation firstly. To maximize each firm’s profit in ERM formulation, they need to find (y_1^*, y_2^*) such that it solves the following problem:

$$\min_{0 \leq y_1 \leq c_1, y_2 \leq c_2} E[\phi(y, \mu, \xi)], \tag{2.3}$$

where $y = (y_1, y_2), \mu = (\mu_1, \mu_2)$ ϕ is a residual function of one-stage SVI in “wait-and-see” model (2.2), e.g.,

$$\begin{aligned} &\phi(y, \mu, \xi) \\ &:= \left\| \min \left\{ \begin{pmatrix} y_1 \\ \mu_1 \\ y_2 \\ \mu_2 \end{pmatrix}, \begin{pmatrix} -p(y_1 + y_2, \xi) - y_1 p'_q(y_1 + y_2, \xi) + H'_1(y_1, \xi) + \mu_1 \\ c_1 - y_1 \\ -p(y_2 + y_1, \xi) - y_2 p'_q(y_2 + y_1, \xi) + H'_2(y_2, \xi) + \mu_2 \\ c_2 - y_2 \end{pmatrix} \right\} \right\|_2^2 \end{aligned}$$

and (2.3) is a one-stage SVI in ERM formulation.

- We then consider the expected-value (EV) formulation. To maximize each firm’s profit in EV formulation, they need to find (y_1^*, y_2^*) such that it solves the following problem:

$$\begin{aligned} &\max_{y_i} E[p(y_i + y_{-i}^*, \xi)y_i - H_i(y_i, \xi)] \\ &\text{s.t. } 0 \leq y_i \leq c_i, \end{aligned} \tag{2.4}$$

for almost every $\xi \in \mathcal{E}$. Moreover, we can write down the KKT conditions for (2.4) as follows:

$$\begin{aligned} &0 \leq \begin{pmatrix} y_i \\ \mu_i \end{pmatrix} \perp \begin{pmatrix} E[-p(y_i + y_{-i}, \xi) - y_i p'_q(y_i + y_{-i}, \xi) + H'_i(y_i, \xi) + \mu_i] \\ c_i - y_i \end{pmatrix} \\ &\geq 0, \text{ for a.e. } \xi \in \mathcal{E}, \end{aligned} \tag{2.5}$$

and (2.5) is a one-stage SVI in EV formulation.

By the three formulations of the two-stage stochastic game, we can give the definition of three kinds of one-stage SVI. Let $f : \mathbb{R}^n \times \mathcal{E} \rightarrow \mathbb{R}^n$ be a continuous function.

1. “wait-and-see” model: find $x : \mathcal{E} \rightarrow \mathbb{R}^n$ such that

$$0 \in f(x(\xi), \xi) + \mathcal{N}_X(x(\xi), \xi), \text{ for a.e. } \xi \in \mathcal{E}.$$

In this model, for every scenario ξ in the further, we have $x(\xi)$ as the solution and $x(\cdot)$ is a measurable function on \mathcal{E} . Since the solution $x(\xi)$ depends on the further uncertain scenario, we call it “wait-and-see” solution and the model “wait-and-see” model. Although the model can give perfect solution for every scenario in

the further, in many real-world applications when the decision needs to be made before we observe the uncertain scenario, it is unworkable.

2. The expected residual minimization (ERM) formulation [6–10]: Find $x \in \mathbb{R}^n$ such that it is a solution of

$$\min_{x \in X} E[\phi(x, \xi)],$$

where $\phi(\cdot, \xi) : X \rightarrow \mathbb{R}$ is a residual function of the VI($X, f(\cdot, \xi)$) for a.e. $\xi \in \mathcal{E}$, that is,

$$\phi(\cdot, \xi) \geq 0, \quad \text{and} \quad \phi(x, \xi) = 0 \Leftrightarrow 0 \in f(x, \xi) + \mathcal{N}_X(x), \quad \text{for a.e. } \xi \in \mathcal{E}.$$

Different with the “wait-and-see model,” the solutions of the ERM formulation do not depend on the further uncertain scenario and the decision makers can make decisions before they observe further uncertainty. We call the solutions “here-and-now” solutions. Moreover, the ERM formulation can quantify the quality of the “here-and-now” solution. The value of $\phi(x, \xi)$ can be considered as the “loss” due to failure of the equilibrium and hence can measure the quality of the “here-and-now” solution at scenario ξ . Then, the expect value of $\phi(x, \xi)$ measures the quality of the “here-and-now” solution in the sense of expectation. In [11], Chen et al. further investigated the ERM in the case when X depends on random vector ξ .

3. The expected-value (EV) formulation [12–15]: Find $x \in \mathbb{R}^n$ such that

$$0 \in E[f(x, \xi)] + \mathcal{N}_X(x). \quad (2.6)$$

The solutions of EV formulation are also “here-and-now” solutions. Similar as deterministic VI, the EV formulation can be used to represent the first-order optimality conditions of one-stage stochastic optimization and describe one-stage stochastic Nash equilibrium. But the “here-and-now” solution of the EV formulation is made in the sense of expectation, for different scenarios in the further, the “here-and-now” solution may not be a good solution. Moreover, when we set $G(x) := E[f(x, \xi)]$, then the EV formulation is the same as the deterministic VI: $0 \in G(x) + \mathcal{N}_X(x)$. Similar as the ERM formulation, we can also reformulate the EV formulation as a minimization problem

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

where $\theta : X \rightarrow \mathbb{R}_+$ is a residual function of the deterministic VI, that is,

$$\theta(\cdot) \geq 0, \quad \text{and} \quad \theta(x) = 0 \Leftrightarrow 0 \in G(x) + \mathcal{N}_X(x).$$

One popular residual function is the regularized gap function [16] as follows:

$$\theta(x) = \max_{v \in X} \langle x - v, G(x) \rangle - \frac{\alpha}{2} \|x - v\|^2. \quad (2.7)$$

Moreover, let $g : \mathbb{R}^n \times \mathcal{E} \rightarrow \mathbb{R}$ be a Lipschitz continuous function w.r.t. x and $f(x, \xi) \in \partial_x g(x, \xi)$, where ∂ denotes the Clarke subdifferential. Consider one-stage stochastic program

$$\min_{x \in X} E[g(x, \xi)]. \tag{2.8}$$

Then, $E[f(x, \xi)] \in E[\partial_x g(x, \xi)]$ and (2.6) is the first-order necessary optimal condition of (2.8). Moreover, (2.8) has wide applications in a lot of areas. For example, in risk management

$$\min_{x \in X} \text{CVaR}_\alpha(l(x, \xi))$$

with some $\alpha \in (0, 1]$ and the loss function $l : \mathbb{R}^n \times \mathcal{E} \rightarrow \mathbb{R}$ can be reformulated as (2.8), where CVaR denotes ‘‘conditional value of risk’’ [17]. For another example, in machine learning and compressed sensing, the formulation of Lasso [18]

$$\min_{x \in \mathbb{R}^n} E[(a(\xi)^T x - b(\xi))^2] + \lambda \|x\|_1$$

can also be formulated as (2.8), where $a : \mathcal{E} \rightarrow \mathbb{R}^n$ and $b : \mathcal{E} \rightarrow \mathbb{R}^n$ are given functions.

Since the solutions of the EV formulation and ERM formulation are all ‘‘here-and-now’’ solutions, we can call the two formulations ‘‘here-and-now’’ model.

In the case when we consider two-stage decisions under uncertain environment, the mathematical tools of one-stage SVI are not enough. To elicit two-stage SVI, we consider an extension of Example 1:

Example 2 (Motivation of two-stage SVI) Similar as Example 1, we consider a duopoly market where two firms compete to supply a homogeneous product (or service) noncooperatively in future. The only difference with Example 1 is that neither of the firms has an existing capacity, and thus, they must make a decision at the present time on their capacity for future supply of quantities in order to have enough time to build the necessary facilities.

For $i = 1, 2$, firm i ’s cost function for building up capacity x_i is $C_i(x_i)$ and the other settings are the same as in the situation 1 of Example 1. Assuming each firm aims to maximize the expected profit, we can then develop a mathematical model for their decision making: For $i = 1, 2$, find $(x_i^*, y_i^*(\cdot))$ such that it solves the following two-stage stochastic programming problem:

$$\begin{aligned} \max_{x_i, y_i(\cdot)} & E[p(y_i(\xi) + y_{-i}^*(\xi), \xi)y_i(\xi) - H_i(y_i(\xi), \xi)] - C_i(x_i) \\ \text{s.t.} & 0 \leq y_i(\xi) \leq x_i, \text{ for a.e. } \xi \in \mathcal{E}. \end{aligned} \tag{2.9}$$

Note that when we know (or fix) the first-stage decision variables x_i , the problem becomes: for $i = 1, 2$, find y_i^* such that

$$\begin{aligned}
 y_i^*(\xi) \in \arg \max_{y_i} & \quad p(y_i + y_{-i}^*, \xi)y_i - H_i(y_i, \xi) \\
 \text{s.t.} & \quad 0 \leq y_i \leq x_i, \quad \text{for a.e. } \xi \in \mathcal{E},
 \end{aligned}
 \tag{2.10}$$

and this is exactly the “wait-and-see” model in Example 1.

In this situation, we fix $x_i, i = 1, 2$ and handle the second-stage game problem firstly. Then, the second-stage SVI representation of (2.9) is (2.10) and we denote $\bar{y}(x, \xi)$ the solution of (2.10). With the second-stage equilibrium $\bar{y}(x, \xi)$, we are ready to write down the first-stage decision-making problem for player i :

$$\begin{aligned}
 \max_{x_i} & \quad E[v_i(x, \xi)] - C_i(x_i) \\
 \text{s.t.} & \quad x_i \geq 0,
 \end{aligned}
 \tag{2.11}$$

where

$$v_i(x, \xi) := p(\bar{y}_i(x, \xi) + \bar{y}_{-i}(x, \xi), \xi)\bar{y}_i(x, \xi) - H_i(\bar{y}_i(x, \xi), \xi).$$

A 4-tuple $(x_1^*, x_2^*, y_1^*(\cdot), y_2^*(\cdot))$ with $(y_1^*(\cdot), y_2^*(\cdot)) = (\bar{y}_1(x^*, \cdot), \bar{y}_2(x^*, \cdot))$ is called a *two-stage stochastic equilibrium* if (x_i^*, x_{-i}^*) solves (2.11). Then, the first-stage equilibrium (x_i^*, x_{-i}^*) satisfies

$$x_i^* \in \arg \max_{x_i \geq 0} E[v_i(x_i, x_{-i}^*, \xi)] - C_i(x_i), \quad i = 1, 2.
 \tag{2.12}$$

Assuming that $C_i(\cdot)$ is continuously differentiable, we may write down the first-order optimality condition of (2.12):

$$0 \in C'_i(x_i) - \partial_{x_i} E[v_i(x, \xi)] + \mathcal{N}_{[0, \infty)}(x_i), \quad i = 1, 2.
 \tag{2.13}$$

In the case when the inverse demand function $p(q, \xi)$ satisfies the following conditions:

- (i) $p(q, \xi)$ is twice continuously differentiable in q and $p'_q(q, \xi) < 0$ for $q \geq 0$ and $\xi \in \mathcal{E}$;
- (ii) $p'_q(q, \xi) + qp''_{qq}(q, \xi) \leq 0$, for $q \geq 0$ and $\xi \in \mathcal{E}$.

It follows by Ralph and Xu [19, Lemma 5.2] that the KKT system of second-stage problem (2.10) has a unique solution; $v(x, \xi)$ is continuously differentiable w.r.t. x_i for $x_i > 0$ and

$$\nabla_{x_i} v_i(x, \xi) = \frac{L(y_i(\xi), \lambda_i(\xi), \mu_i(\xi), x_i)}{dx_i} = \mu_i(\xi),$$

where

$$\begin{aligned}
 L(y_i(\xi), \lambda_i(\xi), \mu_i(\xi), x_i) & := p(y_1(\xi) + y_2(\xi), \xi)y_i(\xi) \\
 & \quad - H_i(y_i(\xi), \xi) + \lambda_i(\xi)y_i(\xi) - \mu_i(\xi)(y_i(\xi) - x_i).
 \end{aligned}$$

Note that in the case when $x_i = 0, y_i(\xi) \equiv 0$, we have

$$\partial_{x_i} v_i(x, \xi) = \{\mu_i(\xi) : \mu_i(\xi) \geq (-g_i(0, 0, \xi))_+, \forall \xi \in \mathcal{E} \text{ and } E[\mu_i(\xi)] \geq C'_i(x_i)\}.$$

Summarizing the discussions above, we can rewrite (2.13) as

$$0 \leq x_i \perp C'_i(x_i) - E[\mu_i(\xi)] \geq 0, \quad i = 1, 2, \tag{2.14}$$

and derive the following two-stage stochastic linear complementarity problem:

$$\begin{cases} 0 \leq x_i \perp C'_i(x_i) - E[\mu_i(\xi)] \geq 0, & i = 1, 2, \\ 0 \leq \begin{pmatrix} y_i(\xi) \\ \mu_i(\xi) \end{pmatrix} \perp \begin{pmatrix} g_i(y_i(\xi), y_{-i}(\xi), \xi) + \mu_i(\xi) \\ x_i - y_i(\xi) \end{pmatrix} \geq 0, & \text{for a.e. } \xi \in \mathcal{E}, \quad i = 1, 2, \end{cases} \tag{2.15}$$

where

$$g_i(y_i(\xi), y_{-i}(\xi), \xi) = -p(y_i(\xi) + y_{-i}(\xi), \xi) - y_i(\xi) p'_q(y_i(\xi) + y_{-i}(\xi), \xi) + H'_i(y_i(\xi), \xi)$$

and (2.15) can be considered as an example of two-stage SVI.

In what follows, we give the definition of two-stage SVI.

Definition 2 (*Two-stage SVI*) Let ξ be a random vector defined as above and \mathcal{Y} be the measurable function space defined on \mathcal{E} . The two-stage SVI is to find an $(x^*, y^*(\cdot)) \in D \times \mathcal{Y}$ such that

$$0 \in E[\Phi(x, y(\xi), \xi)] + \mathcal{N}_D(x), \tag{2.16}$$

$$0 \in \Psi(x, y(\xi), \xi) + \mathcal{N}_{C(\xi)}(y(\xi)), \quad \text{for a.e. } \xi \in \mathcal{E}, \tag{2.17}$$

where $D \subset \mathbb{R}^n$ and $C(\xi) \subset \mathbb{R}^m$, a.e. $\xi \in \mathcal{E}$ are nonempty closed convex sets, $\Phi : \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^d \rightarrow \mathbb{R}^n$ and $\Psi : \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^d \rightarrow \mathbb{R}^m$, and $\Gamma_1 : \mathbb{R}^n \rightrightarrows \mathbb{R}^n, \Gamma_2 : \mathbb{R}^m \times \mathcal{E} \rightrightarrows \mathbb{R}^m$ are multifunctions (point-to-set mappings). We assume throughout the paper that $y(\cdot) \in \mathcal{Y}$ with \mathcal{Y} being the space of measurable functions from \mathcal{E} to \mathbb{R}^m such that the expected value in (2.16) is well defined.

Moreover, if the sets D and $C(\xi), \xi \in \mathcal{E}$, are closed convex cones, then

$$\mathcal{N}_D(x) = \{x^* \in D^* : x^\top x^* = 0\}, \quad x \in D,$$

where $D^* = \{x^* : x^\top x^* \leq 0, \forall x \in D\}$ is the (negative) dual of cone D . In that case, SVI (2.16)–(2.17) reduces to the following two-stage stochastic cone VI:

$$\begin{aligned} D \ni x \perp E[\Phi(x, y(\xi), \xi)] &\in -D^*, \\ C(\xi) \ni y(\xi) \perp \Psi(x, y(\xi), \xi) &\in -C^*(\xi), \quad \text{for a.e. } \xi \in \mathcal{E}. \end{aligned}$$

In particular when $D := \mathbb{R}_+^n$ with $D^* = -\mathbb{R}_+^n$, and $C(\xi) := \mathbb{R}_+^m$ with $C^*(\xi) = -\mathbb{R}_+^m$ for all $\xi \in \mathcal{E}$, SVI (2.16)–(2.17) reduces to the two-stage stochastic nonlinear complementarity problem (SNCP):

$$\begin{aligned} 0 &\leq x \perp E[\Phi(x, y(\xi), \xi)] \geq 0, \\ 0 &\leq y(\xi) \perp \Psi(x, y(\xi), \xi) \geq 0, \quad \text{for a.e. } \xi \in \mathcal{E}, \end{aligned}$$

which is a generalization of the two-stage stochastic linear complementarity problem (SLCP):

$$0 \leq x \perp Ax + E[B(\xi)y(\xi)] + q_1 \geq 0, \tag{2.18}$$

$$0 \leq y(\xi) \perp N(\xi)x + M(\xi)y(\xi) + q_2(\xi) \geq 0, \quad \text{for a.e. } \xi \in \mathcal{E}, \tag{2.19}$$

where $A \in \mathbb{R}^{n \times n}$, $B : \mathcal{E} \rightarrow \mathbb{R}^{n \times m}$, $N : \mathcal{E} \rightarrow \mathbb{R}^{m \times n}$, $M : \mathcal{E} \rightarrow \mathbb{R}^{m \times m}$, $q_1 \in \mathbb{R}^n$, $q_2 : \mathcal{E} \rightarrow \mathbb{R}^m$. More specifically, when $D := \mathbb{R}^n$ and $C(\xi) := \mathbb{R}^m$ for all $\xi \in \mathcal{E}$, SVI (2.16)–(2.17) reduces to the two-stage stochastic equations:

$$\begin{aligned} E[\Phi(x, y(\xi), \xi)] &= 0, \\ \Psi(x, y(\xi), \xi) &= 0, \quad \text{for a.e. } \xi \in \mathcal{E}. \end{aligned}$$

The two-stage SVI is basically an infinite dimension VI and not easy to handle. In [20], Chen et. al. proposed the two-stage SVI model to deal with random variables in VI and formulate this model as a two-stage stochastic programming with recourse by using an ERM solution procedure. We will introduce the procedure in the next section.

In [20], Chen et. al. considered a two-stage SVI as follows:

$$0 \in G(x) + \mathcal{N}_D(x), \tag{2.20}$$

$$0 \in F(\bar{y}(x, \xi), \xi) + \mathcal{N}_{C(\xi)}(\bar{y}(x, \xi)), \quad \text{for a.e. } \xi \in \mathcal{E}, \tag{2.21}$$

where $G : \mathbb{R}^n \rightarrow \mathbb{R}^n$, $F : \mathbb{R}^n \times \mathcal{E} \rightarrow \mathbb{R}^m$, $\bar{y} : D \times \mathcal{E} \rightarrow \mathbb{R}^m$, D and $C(\xi)$, $\forall \xi \in \mathcal{E}$ are nonempty closed convex subsets of \mathbb{R}^n and \mathbb{R}^m . Formulations (2.20)–(2.21) are special cases of two-stage SVI.

Moreover, Chen et. al. [20] also considered a special case of two-stage SVI (2.20)–(2.21) as follows:

$$0 \in G(x) + \mathcal{N}_D(x), \tag{2.22}$$

$$0 \in F(u_\xi, \xi) + \mathcal{N}_{C(\xi)}(u_\xi), \quad \text{for a.e. } \xi \in \mathcal{E}, \tag{2.23}$$

where the first-stage problem (2.22) is an EV form of the one-stage SVI and the second-stage problem (2.23) is the “wait-and-see” model of the one-stage SVI. In general, the solution sets of the variational inequalities (2.22) and (2.23) can have multiple solutions. Naturally, a “here-and-now” solution x should have minimum total distances to the solution set of (2.23) for almost all observations $\xi \in \mathcal{E}$. Then, the

problem can be written as a mathematical programming with equilibrium constraints (MPEC) [21] as the following:

$$\begin{aligned} \min E[\|u_\xi - x\|^2] \\ \text{s.t. } 0 \in G(x) + \mathcal{N}_D(x), \quad 0 \in F(u_\xi, \xi) + \mathcal{N}_{C(\xi)}(u_\xi), \quad \text{for a.e. } \xi \in \mathcal{E}. \end{aligned} \tag{2.24}$$

2.2 Multistage SVI

In [22], Rockafellar and Wets first proposed the formwork of the multistage SVI. The multistage SVI is an extension of the two-stage SVI and can deal with multistage problems of optimization and equilibrium in a stochastic setting which involves actions that respond to increasing levels of information. Moreover, Rockafellar and Sun [23, 24] extended the progressive hedging method to solve multistage SVIs.

We begin with two-stage SVI setting in (2.16)–(2.17) and explain the equivalence between the formulations in [22] and in the previous section. Let $\mathcal{L}_{n+m} = \mathcal{L}_n \times \mathcal{L}_m$, where \mathcal{L}_n and \mathcal{L}_m are function spaces of all functions $x(\cdot) : \mathcal{E} \rightarrow \mathbb{R}^n$ and $y(\cdot) : \mathcal{E} \rightarrow \mathbb{R}^m$. Let

$$\mathcal{C} := \{(x(\cdot), y(\cdot)) \in \mathcal{L}_{n+m} \mid x(\xi) \in D, y(\xi) \in C(\xi) \text{ for a.e. } \xi \in \mathcal{E}\}$$

and

$$\mathcal{N} := \{(x(\cdot), y(\cdot)) \in \mathcal{L}_{n+m} \mid x(\xi) \text{ is the same for all } \xi \in \mathcal{E}\}$$

be nonanticipativity subspaces of \mathcal{L}_{n+m} . Consider a mixed case of nonanticipativity: The solution pair (x, y) with two-stage variables x (here-and-now) and y (wait-and-see) is in the function space \mathcal{L}_{n+m} with $x(\cdot) : \mathcal{E} \rightarrow \mathbb{R}^n$ and $y(\cdot) : \mathcal{E} \rightarrow \mathbb{R}^m$ and $(x, y) \in \mathcal{C} \cap \mathcal{N}$. Then, the basic form of the two-stage SVI introduced in [22] is

$$0 \in \mathcal{F}(x(\cdot), y(\cdot)) + N_{\mathcal{C} \cap \mathcal{N}}(x(\cdot), y(\cdot)), \tag{2.25}$$

where $\mathcal{F} : \mathcal{L}_{n+m} \rightarrow \mathcal{L}_{n+m}$ is a continuous mapping such that for $(x(\cdot), y(\cdot)) \in \mathcal{L}_{n+m}$, $\mathcal{F}(x(\cdot), y(\cdot))$ is the function in \mathcal{L}_{n+m} that takes $\xi \in \mathcal{E}$ to $(\Phi(x(\xi), y(\xi), \xi), \Psi(x(\xi), y(\xi), \xi)) \in \mathbb{R}^{n+m}$.

Moreover, let

$$\begin{aligned} \mathcal{M} = \{w(\cdot) = (w_1(\cdot), w_2(\cdot)) \in \mathcal{L}_{n+m} \mid w_1(\cdot) \in \mathcal{L}_n, \\ w_2(\cdot) \in \mathcal{L}_m, E[w_1(\xi)] = 0, w_2(\xi) = 0, \text{ a.e. } \xi \in \mathcal{E}\}. \end{aligned}$$

It is easy to verify that \mathcal{M} is the orthogonal complement of \mathcal{N} and can be considered as the space of nonanticipativity multipliers. Then, the extensive form of the two-stage SVI in [22] is:

$(x(\cdot), y(\cdot)) \in \mathcal{N}$ and there exists $w(\cdot) \in \mathcal{M}$ such that

$$-\begin{pmatrix} \Phi(x(\xi), y(\xi), \xi) \\ \Psi(x(\xi), y(\xi), \xi) \end{pmatrix} - \begin{pmatrix} w_1(\xi) \\ w_2(\xi) \end{pmatrix} \in \begin{pmatrix} N_D(x(\xi)) \\ N_{C(\xi)}(y(\xi)) \end{pmatrix}, \quad \text{a.e. } \xi \in \mathcal{E}. \tag{2.26}$$

Under the constraints qualification (CQ):

$$\text{there exists some } \hat{x}(\cdot) \in \mathcal{N} \text{ such that } \hat{x}(\xi) \in \text{ri}C(\xi) \text{ a.e. } \xi \in \mathcal{E}, \tag{2.27}$$

the equivalence between (2.25) and (2.26) has been proved in [22, Theorem 3.2] for multistage SVI. Now we consider the equivalence between (2.26) and (2.16)–(2.17).

Under (2.27), suppose $(x^*(\xi), y^*(\xi), w_1^*(\xi))$ ($w_2(\xi) \equiv 0$) is a solution of (2.26) with $x^*(\xi) = x^*$ for all $\xi \in \mathcal{E}$ and $E[w_1^*(\xi)] = 0$. Then, taking expectation at first line of (2.26), we have (2.16)–(2.17). Conversely, suppose $(x^*, y^*(\xi))$ is a solution of (2.16)–(2.17). By (2.16), $-E[\Phi(x^*, y^*(\xi), \xi)] \in N_D(x^*)$, there exists $w_1^*(\xi)$ such that $-\Phi(x^*, y^*(\xi), \xi) - w_1^*(\xi) \in N_D(x^*)$ and $E[w_1^*(\xi)] = 0$, which implies $(x^*(\xi), y^*(\xi), w_1^*(\xi))$ with $x^*(\xi) = x^*$ and $E[w_1^*(\xi)] = 0$ is a solution of (2.26). Then, (2.16)–(2.17) and (2.26) are equivalence.

Then, we consider the multistage SVI. We adopt an N -stage pattern

$$x_1, \xi_1, x_2, \xi_2, \dots, x_N, \xi_N,$$

where $x_k \in \mathbb{R}^{n_k}$, $n = n_1 + \dots + n_N$ and $\xi_k \in \mathcal{E}_k$, x_k is the decision to be taken at the k th stage before we observe the information ξ_i , $i \geq k$ and ξ_k stands for the information revealed after that decision, but before the next. Let

$$\xi = (\xi_1, \dots, \xi_N) \in \mathcal{E} = \mathcal{E}_1 \times \dots \times \mathcal{E}_N$$

and

$$\tilde{x}(\xi) = (x_1, x_2(\xi_1), x_3(\xi_1, \xi_2), \dots, x_N(\xi_1, \dots, \xi_{N-1})) \in \mathbb{R}^n = \mathbb{R}^{n_1} \times \dots \times \mathbb{R}^{n_N}.$$

Similar as in two-stage SVI case, we use $x(\cdot) : \xi \rightarrow (x_1(\xi), \dots, x_N(\xi))$ in \mathcal{L}_n to replace $\tilde{x}(\xi)$ and restrict it to nonanticipativity subspace of \mathcal{L}_n :

$$\mathcal{N} = \{x(\cdot) = (x_1(\cdot), \dots, x_N(\cdot)) \mid x_k(\xi) \text{ does not depend on } \xi_k, \dots, \xi_N\}. \tag{2.28}$$

Corresponding nonanticipativity multipliers will again come from a subspace \mathcal{M} of \mathcal{L}_n as follows

$$\mathcal{M} = \{w(\cdot) = (w_1(\cdot), \dots, w_N(\cdot)) \mid E_{\xi_k, \dots, \xi_N}[w_k(\xi_1, \dots, \xi_{k-1}, \xi_k, \dots, \xi_N)] = 0\}, \tag{2.29}$$

where the expectation is the conditional expectation knowing the initial components ξ_1, \dots, ξ_{k-1} . Note also that \mathcal{N} and \mathcal{M} are orthogonal complement of each other, that is, $\mathcal{N}^\perp = \mathcal{M}$. Moreover, let $C(\xi)$ be a nonempty closed convex subset of \mathbb{R}^n , a.e. $\xi \in \mathcal{E}$,

$$C = \{x(\cdot) \in \mathcal{L}_n \mid x(\xi) \in C(\xi), \text{ a.e. } \xi \in \mathcal{E}\}$$

be a nonempty closed convex subset of \mathcal{L}_n , $F(x, \xi) = (F_1(x, \xi), \dots, F_N(x, \xi))$ be a continuous vector valued function w.r.t. $x \in \mathbb{R}^n$ with $F_k(x, \xi) \in \mathbb{R}^{n_k}$ and \mathcal{F} from \mathcal{L}_n to \mathcal{L}_n such that

$$\mathcal{F}(x(\cdot)) : \xi \rightarrow F(x(\xi), \xi) = (F_1(x(\xi), \xi), \dots, F_N(x(\xi), \xi)).$$

Then, the basic and extensive forms of multistage SVI are as follows:

Definition 3 The basic form of multistage SVI is to find $x(\cdot) \in \mathcal{L}_n$ such that

$$0 \in \mathcal{F}(x(\cdot)) + \mathcal{N}_{\mathcal{C} \cap \mathcal{N}}(x(\cdot)), \tag{2.30}$$

where the extensive form of multistage SVI is $x(\cdot) \in \mathcal{N}$ and there exists $w(\cdot) \in \mathcal{M}$ such that

$$0 \in F(x(\xi), \xi) + w(\xi) + \mathcal{N}_{C(\xi)}(x(\xi)), \text{ a.e. } \xi \in \Xi. \tag{2.31}$$

[22, Theorem 3.2] gives the proof of equivalence between (2.30) and (2.31) under the CQ (2.27). Moreover, if

$$C(\xi) = C_1 \times C_2(\xi_1) \times \dots \times C_N(\xi_1, \dots, \xi_{N-1}),$$

(2.31) can be written as

$$0 \in E_{\xi_1, \dots, \xi_N} [F_1(x_1, x_2(\xi_1), \dots, x_N(\xi_1, \dots, \xi_{N-1}), \xi)] + \mathcal{N}_{C_1}(x_1), \tag{2.32}$$

$$0 \in E_{\xi_2, \dots, \xi_N} [F_2(x_1, x_2(\xi_1), \dots, x_N(\xi_1, \dots, \xi_{N-1}), \xi)] + \mathcal{N}_{C_2(\xi_1)}(x_2(\xi_1)), \text{ for a.e. } \xi_1 \in \Xi_1, \tag{2.33}$$

$$0 \in E_{\xi_3, \dots, \xi_N} [F_3(x_1, x_2(\xi_1), \dots, x_N(\xi_1, \dots, \xi_{N-1}), \xi)] + \mathcal{N}_{C_3(\xi_1, \xi_2)}(x_3(\xi_1, \xi_2)), \tag{2.34}$$

$$\text{for a.e. } (\xi_1, \xi_2) \in \Xi_1 \times \Xi_2,$$

...

$$0 \in E_{\xi_N} [F_N(x_1, x_2(\xi_1), \dots, x_N(\xi_1, \dots, \xi_{N-1}), \xi)] + \mathcal{N}_{C_N(\xi_1, \dots, \xi_{N-1})}(x_N(\xi_1, \dots, \xi_{N-1})), \text{ for a.e. } (\xi_1, \dots, \xi_{N-1}) \in \Xi_1 \times \dots \times \Xi_{N-1}. \tag{2.35}$$

The equivalence between (2.31) and (2.32)–(2.35) is proved in [22, Theorem 3.4]

Moreover, if, in (2.30), the sets $C(\xi)$ are specified by

$$x \in C(\xi) \iff x \in B(\xi) \text{ and } f_i(x, \xi) \begin{cases} \leq 0, & \text{for } i = 1, \dots, r, \\ = 0, & \text{for } i = r + 1, \dots, m, \end{cases}$$

where $B(\xi)$ is a nonempty closed convex set and $f_i(x, \xi)$ is differentiable and convex in x for $i = 1, \dots, r$ and affine in x for $i = r + 1, \dots, m$, then under the CQ

there exists $\hat{x}(\cdot) \in \mathcal{N}$ such that $\hat{x}(\xi) \in \text{ri}B(\xi)$ and

$$f_i(\hat{x}(\xi), \xi) \begin{cases} < 0, & i \leq r, \\ = 0, & i > r, \end{cases} \text{ a.e. } \xi \in \Xi. \tag{2.36}$$

The representation of $N_C(x(\cdot))$ ([22, Theorem 3.7]) is

$$\begin{aligned} &\exists y(\cdot) \in \mathcal{L}_m, z(\cdot) \in \mathcal{L}_n, \text{ such that} \\ &v(\xi) = \sum_{i=1}^m y_i(\xi) \nabla_x f_i(x(\xi), \xi) + z(\xi) \text{ with} \\ &f(x(\xi), \xi) \in N_Y(y(\xi)), z(\xi) \in N_{B(\xi)}(x(\xi)), \end{aligned} \tag{2.37}$$

where $Y = [0, +\infty]^r \times (-\infty, \infty)^{m-r}$ and the Lagrangian basic form of the associated SVI is

$$\begin{aligned} &x(\cdot) \in \mathcal{N}, y(\cdot) \in \mathcal{L}_m, \text{ and } \exists w(\cdot) \in \mathcal{M} \text{ such that} \\ &0 \in \left(F(x(\xi), \xi) + \sum_{i=1}^s y_i(\xi) \nabla_x f_i(x(\xi), \xi) - f(x(\xi), \xi) \right) \\ &\quad + (w(\xi), 0) + \mathcal{N}_{B(\xi) \times Y}(x(\xi), y(\xi)). \end{aligned} \tag{2.38}$$

3 Algorithms and Approximation Methods

In this section, we consider the algorithms and approximation methods for two-stage and multistage SVI. We first introduced the ERM formulation of (2.20)–(2.21) as follows.

3.1 ERM Solution Procedure

As we introduced in Sect. 2.1, the ERM formulation is one of three important formulations of one-stage SVI. Chen et. al. [11] first introduced second-stage recourse variables into SVI, that is, find x and $u(x, \xi)$ such that

$$0 \in f(u(x, \xi), \xi) + \mathcal{N}_{C(\xi)}(u(x, \xi)).$$

In [20], Chen et. al. extended the ERM formulation from one-stage SVI to two-stage SVI (2.20)–(2.21). For this propose, they need an extension of residual function as follows.

Definition 4 (SVI-residual function) Given a closed, convex set $D \in \mathbb{R}^n$ and the random vector ξ , let us consider the following collection of VI (SVI):

$$\begin{aligned} &\text{find } \bar{x} \in \mathbb{R}^n, u : D \times X \rightarrow \mathbb{R}^m, \text{ P-measurable in } \xi \text{ such that} \\ &0 \in F(\bar{u}(x, \xi), \xi) + \mathcal{N}_{C(\xi)}(\bar{u}(x, \xi)). \end{aligned}$$

A function $r : \mathbb{R}^m \times \mathcal{E} \rightarrow \mathbb{R}$ is a residual function for these inclusions if the following conditions are satisfied:

1. $r(u, \xi) \geq 0$ for all $u \in C(\xi)$, a.e. $\xi \in \mathcal{E}$;

2. For any $u : C \times \mathcal{E} \rightarrow \mathbb{R}^n$, it holds that

$$0 \in F(\bar{u}(x, \xi), \xi) + \mathcal{N}_{C(\xi)}(\bar{u}(x, \xi)) \\ \Leftrightarrow r(\bar{u}(x, \xi), \xi) = 0 \text{ and } \bar{u}(x, \xi) \in C(\xi), \text{ for a.e. } \xi \in \mathcal{E}.$$

One popular SVI-residual function which used in [20] is the regularized gap function [16]

$$r(u, \xi) := \max_{z \in C(\xi)} \left\{ \langle u - z, F(u, \xi) \rangle - \frac{\alpha}{2} \|u - z\|^2 \right\}. \tag{3.1}$$

The use of residual functions θ defined in (2.7) and r above for two-stage SVI (2.20)-(2.21) leads us to seeking a solution of the stochastic program

$$\min_{x \in X} \theta(x) + \lambda E[r(\bar{u}(x, \xi), \xi) + Q(x, \xi)], \\ \text{where } \bar{u}(x, \xi) = x + Wu_\xi^*, \quad Q(x, \xi) = \frac{1}{2} \langle u_\xi^*, Hu_\xi^* \rangle, \text{ for a.e. } \xi \in \mathcal{E}, \\ u_\xi^* = \arg \min \left\{ \frac{1}{2} \langle u, Hu \rangle \mid x + Wu \in C(\xi) \right\}. \tag{3.2}$$

Assumption 1 Assume (i) W has full row rank and (ii) $C(\xi) \subseteq K$, a compact convex set for all ξ .

Theorem 1 Suppose Assumption 1 holds and r is a residual function defined in Definition 4. Then for any $x \in D$ and a.e. $\xi \in \mathcal{E}$, the function $r(\bar{y}(x, \xi), \xi) + Q(x, \xi)$ in (3.2) is finite, nonnegative with

$$v(\bar{y}(x, \xi), \xi) = \text{prj}_{C(\xi)}(\bar{y}(x, \xi) - \frac{1}{\alpha} F(\bar{y}(x, \xi), \xi))$$

as the unique maximizer of the maximization problem in (3.1).

Theorem 1 means that problem (3.2) is a two-stage stochastic program with complete recourse. However, the objective function of problem (3.2) involves minimizers of constrained quadratic programs for $\xi \in \mathcal{E}$ and is not necessarily differentiable even when the sample is finite.

Assumption 2 The functions $F(\cdot, \xi)$ and $G(\cdot)$ are continuously differentiable for all $\xi \in \mathcal{E}$. Moreover, for any compact set $Y \subset \mathbb{R}^m$, there are functions $d, \rho : \mathcal{E} \rightarrow \mathbb{R}_+$ such that $\|F(y, \xi)\| \leq d_\xi$ and $\|\nabla F(y, \xi)\| \leq \rho_\xi$ for all $y \in Y$, where $d \in L_1^\infty$ and $\rho \in L_1^1$.

Lemma 1 Suppose Assumptions 1–2 hold. Then for a.e. ξ , $r(\cdot, \xi)$ is continuously differentiable and its gradient is given by

$$\nabla_y r(y, \xi) = F(y, \xi) - (\nabla_y F(y, \xi) - \alpha I)(v(y, \xi) - y).$$

Moreover, for any measurable $y(\xi) \in_{a.e.} C(\xi)$, both $\xi \rightarrow r(y(\xi), \xi)$ and $\xi \rightarrow \nabla_y r(y(\xi), \xi)$ are not only measurable but actually summable uniformly in $y(\xi)$. In particular, this means that the objective function in (3.2) is well defined at any $x \in D$, and the optimal value of (3.2) is finite.

By the idea of the L-shaped algorithm [25], they lead to the following problem, whose objective is smooth when the sample is finite:

$$\begin{aligned} \min_{x \in D} \theta(x) + \lambda E \left[r(y(\xi), \xi) + \frac{1}{2} \langle y(\xi) - x, B(y(\xi) - x) \rangle \right] \\ \text{s.t. } y(\xi) \in_{a.e.} C(\xi), \end{aligned} \quad (3.3)$$

and equivalent to

$$\min_{x \in D} \theta(x) + \lambda E \left[r(y(\xi), \xi) + \frac{1}{2} \langle y(\xi) - x, B(y(\xi) - x) \rangle + \delta_{C(\xi)}(y(\xi)) \right], \quad (3.4)$$

where $B = (WH^{-1}W^T)^{-1}$, $\delta_{C(\xi)}(\cdot)$ is the indicator function of the set $C(\xi)$ which is zero in the inside of the set and is infinity otherwise. It is not hard to see that the optimal value of (3.4) is smaller than that of (3.2) since fewer restrictions are imposed on $y(\xi)$. Hence, it follows from Lemma 1 that the optimal value of (3.4) is also finite. Moreover, by the interchange of minimization and integration [20, Lemma 3.6] and [26, Theorem 14.60], problem (3.4) is equivalent to

$$\begin{aligned} \min_{x \in D} \phi(x) := \theta(x) + \lambda E[\psi(x, \xi)] \\ \text{s.t. } \psi(x, \xi) := \min_{y(\xi) \in C(\xi)} r(y(\xi), \xi) + \frac{1}{2} \langle y(\xi) - x, B(y(\xi) - x) \rangle. \end{aligned} \quad (3.5)$$

Theorem 2 *Suppose Assumptions 1–2 hold. Then, problems (3.2) and (3.4) are solvable. Let v_1 and v_2 be the optimal values of (3.2) and (3.4), respectively. Then, $v_1 \geq v_2$. Moreover, if for any $x \in D$ and $x + Wy, x + Wz \in C(\xi)$, we have*

$$|r(x + Wy, \xi) - r(x + Wz, \xi)| \leq \frac{1}{2} |\langle y, Hy \rangle - \langle z, Hz \rangle|, \text{ for a.e. } \xi \in \mathcal{E},$$

then the two problems have the same optimal value.

Note that since problem (3.4) is equivalent to (3.5), we can replace (3.4) by (3.5) in Theorem 2.

Under Assumptions 1–2, similar as Lemma 1, Chen et. al. [20] proved the smoothness of the objective function in the second-stage problem of (3.5) ([20, Proposition 3.8]). Moreover, they also considered the convexity of (3.5) ([20, Proposition 3.9 and Corollary 3.11]).

Let $\{\xi_1, \dots, \xi_N\}$ be the independent and identically distributed (i.i.d.) samples of ξ and

$$\begin{aligned} G_N(x) &= \frac{1}{N} \sum_{i=1}^N f(x, \xi_i), \\ \theta_N &= \max_{v \in C} \langle x - v, G_N(x) \rangle - \frac{\alpha}{2} \|x - v\|^2, \\ \psi(x, \xi_i) &= \min_{y \in C(\xi)} r(\xi, y) + \frac{1}{2} \langle y - x, B(y - x) \rangle. \end{aligned}$$

Then, the SAA problem of (3.5) is

$$\min_{x \in D} \phi_N(x) := \theta_N(x) + \frac{\lambda}{N} \sum_{i=1}^N \psi(x, \xi_i). \quad (3.6)$$

In [20], Chen et. al. gave the following convergence result.

Theorem 3 (Convergence theorem) *Suppose Assumptions 1–2 hold. Then, ϕ_N converges to ϕ a.s.-uniformly on the compact set \bar{D} such that $S, S^* \subseteq \bar{D}$. Let $\{x_N\}$ be a sequence of minimizers of problem (3.6) generated by iid samples. Then, $\{x_N\}$ is \mathcal{P} -a.e. bounded and any accumulation point x^* of $\{x_N\}$ as $v \rightarrow \infty$ is \mathcal{P} -a.e. a solution of (3.5).*

Chen et. al. [20] also considered an ERM formulation to solve the MPEC reformulation (2.24) of the two-stage SVI (2.22)–(2.23) as follows:

$$\begin{aligned} \min \quad & \frac{1}{\lambda\rho} \theta(x) + \frac{1}{\rho} E[r(u_\xi, \xi)] + E[\|u_\xi - x\|^2] \\ \text{s.t.} \quad & x \in D, u_\xi \in C(\xi), \text{ for a.e. } \xi \in \mathcal{E}, \end{aligned} \quad (3.7)$$

where $\theta(\cdot)$ and $r(\cdot, \xi)$ are residual functions of $0 \in G(x) + \mathcal{N}_D(x)$ and $0 \in_{a.e.} F(u_\xi, \xi) + \mathcal{N}_{C(\xi)}(u_\xi)$, respectively. They then propose a nonconvex Douglas–Rachford splitting method [27,28] to solve problem (3.7). The detailed model and the algorithm are given in [20, Section 5].

3.2 Progressive Hedging Algorithm for SVI

In the case when the random vectors follow a discrete distribution, for solving two-stage SVI (2.16)–(2.17) and multistage SVI (2.31), Rockafellar and Sun [23,24] extended the progressive hedging algorithm (PHA) [29] for multistage stochastic optimization problem to multistage SVI and multistage stochastic Lagrangian variational inequalities. With the notation in Sect. 2.2, the progressive hedging algorithm for SVI is as follows:

The convergence of Algorithm 1 is given in [23] as follows:

Theorem 4 [23, Theorem 2] *As long as the (monotone) variational inequality (2.30) has at least one solution, the sequence of pairs $(x^v(\cdot), w^v(\cdot))$ generated by Algorithm 1 will converge to pair $(\bar{x}(\cdot), \bar{w}(\cdot))$ satisfying (2.31) and thus furnish $\bar{x}(\cdot)$ as a solution to (2.30). The decrease will surely be at a linear rate if, in particular, the sets $C(\xi)$ are polyhedral and functions $F(\cdot, \xi)$ are affine.*

Algorithm 1 Progressive Hedging Algorithm

Initialization: Set iteration $\nu = 0$, $x^0 \in \mathcal{N}$ and $w^0 \in \mathcal{M}$, where \mathcal{N} and \mathcal{M} are defined in (2.28) and (2.29).

Step 1: Determining $\hat{x}^\nu(\cdot) \in \mathcal{L}$ by solving a separate problem for each scenario ξ to get $\hat{x}^\nu(\xi)$:

$$\hat{x}^\nu(\xi) = \text{the unique } x(\xi) \text{ such that} \tag{3.8}$$

$$-F(x(\xi), \xi) - w^\nu(\xi) - r[x(\xi) - x^\nu(\xi)] \in N_{C(\xi)}(x(\xi)).$$

Step 2: Set

$$x^{\nu+1}(\cdot) = P_{\mathcal{N}}(\hat{x}^\nu(\cdot)) \text{ and } w^{\nu+1}(\cdot) = w^\nu(\cdot) + rP_{\mathcal{M}}(\hat{x}^\nu(\cdot)),$$

where $P_{\mathcal{N}}$ and $P_{\mathcal{M}}$ denote the projection mapping.

Step 3: $\nu = \nu + 1$, go to Step 1.

Moreover, in [24], Rockefallar and Sun considered the multistage Lagrangian stochastic variational inequality (LSVI) problem. With the notation in Sect. 2.2, consider an N stage problem. Let $X(\xi) \subset \mathbb{R}^n$ and $Y(\xi) \subset \mathbb{R}^m$ be a pair of nonempty closed convex sets and

$$\mathcal{X} = \{x(\cdot) \in \mathcal{L}_n | x(\xi) \in X(\xi), \forall \xi\}, \quad \mathcal{Y} = \{y(\cdot) \in \mathcal{L}_m | y(\xi) \in Y(\xi), \forall \xi\},$$

where, similar as $x(\cdot), y(\cdot) : \xi \rightarrow y(\xi) = (y_1(\xi), \dots, y_N(\xi)) \in \mathbb{R}^{m_1} \times \dots \times \mathbb{R}^{m_N} = \mathbb{R}^m$. Likewise $\mathcal{N}_n \subset \mathcal{L}_n$ and its complement \mathcal{M}_n , we define the nonanticipativity subspace $\mathcal{N}_m \subset \mathcal{L}_m$ and its complement \mathcal{M}_m . Then, we introduce continuously differentiable functions

$L(\cdot, \cdot, \xi)$ on $X(\xi) \times Y(\xi)$ such that $L(x, y, \xi)$ is convex in x and concave in y

and define

$$\begin{aligned} \Lambda(x(\cdot), y(\cdot)) &= E_\xi[L(x(\xi), y(\xi), \xi)] \\ &= \sum_{\xi \in \mathcal{E}} \pi(\xi)L(x(\xi), y(\xi), \xi) \text{ for } x(\cdot) \in \mathcal{X}, y(\cdot) \in \mathcal{Y} \end{aligned}$$

and

$$\mathcal{F}(x(\cdot), y(\cdot)) = (\nabla_{x(\cdot)}\Lambda(\bar{x}(\cdot), \bar{y}(\cdot)), -\nabla_{y(\cdot)}\Lambda(x(\cdot), y(\cdot)))$$

arises from component mappings

$$F(x, y, \xi) = (-\nabla_x L(x, y, \xi), \nabla_y L(x, y, \xi)).$$

Then, the multistage LSVI is as follows

$$\begin{aligned} \text{find } x(\cdot) \in \mathcal{N}_n, y(\cdot) \in \mathcal{N}_m, \text{ for which } \exists \bar{w} \in \mathcal{M}_n, \bar{z}(\cdot) \in \mathcal{M}_m, \text{ such that} \\ -\nabla_x \Lambda(\bar{x}, \bar{y}) - \bar{w}(\cdot) \in N_X(\bar{x}), \nabla_y \Lambda(\bar{x}, \bar{y}) + \bar{z}(\cdot) \in N_Y(\bar{y}), \end{aligned} \tag{3.9}$$

which is equivalent to having, for all scenarios $\xi \in \Xi$,

$$\begin{aligned} -\nabla_x L(\bar{x}(\xi), \bar{y}(\xi), \xi) - \bar{w}(\xi) &\in N_{X(\xi)}(\bar{x}(\xi)), \\ \nabla_y L(\bar{x}(\xi), \bar{y}(\xi), \xi) + \bar{z}(\xi) &\in N_{Y(\xi)}(\bar{y}(\xi)). \end{aligned}$$

The related progressive hedging algorithm is given in [24] as follows.

Algorithm 2 Progressive Hedging Algorithm for Lagrangian Stochastic Variational Inequalities

Initialization: Set iteration $v = 0$, $x^0 \in \mathcal{N}_n$, $y^0 \in \mathcal{N}_m$, $w^0 \in \mathcal{M}_n$ and $z^0 \in \mathcal{M}_m$.

Step 1: Determining $(\hat{x}^v(\cdot), \hat{y}^v) \in \mathcal{L}_n \times \mathcal{L}_m$ by solving a separate problem for each scenario ξ to get $(\hat{x}^v(\xi), \hat{y}^v(\xi))$:

$$\begin{aligned} -\nabla_x L^v(\hat{x}^v(\xi), \hat{y}^v(\xi), \xi) &\in \mathcal{N}_{X(\xi)}(\hat{x}^v(\xi)), \\ \nabla_y L^v(\hat{x}^v(\xi), \hat{y}^v(\xi), \xi) &\in \mathcal{N}_{Y(\xi)}(\hat{y}^v(\xi)), \end{aligned} \tag{3.10}$$

where

$$L^v(x, y, \xi) = L(x, y, \xi) + w^v(\xi)x + \frac{r}{2}\|x - x^v(\xi)\|^2 - z^v(\xi)y - \frac{r}{2}\|y - y^v(\xi)\|^2.$$

Step 2: Set

$$x^{v+1}(\cdot) = P_{\mathcal{N}_n}(\hat{x}^v(\cdot)), \quad w^{v+1}(\cdot) = w^v(\cdot) + rP_{\mathcal{M}_n}(\hat{x}^v(\cdot)), \tag{3.11}$$

$$y^{v+1}(\cdot) := P_{\mathcal{N}_m}(\hat{y}^v(\cdot)), \quad z^{v+1}(\cdot) = z^v(\cdot) + rP_{\mathcal{M}_m}(\hat{y}^v(\cdot)). \tag{3.12}$$

Step 3: $v := v + 1$, go to Step 1.

This version of progressive hedging inherits from the one for general stochastic variational inequalities in the preceding section the property that, as long as a solution exists, the sequence of iterates $(x^v(\cdot), y^v(\cdot), w^v(\cdot), z^v(\cdot))$ will converge to a particular solution $(\bar{x}(\cdot), \bar{y}(\cdot), \bar{w}(\cdot), \bar{z}(\cdot))$. They also consider the variant version of the algorithm when the parameters r are different in the x part and y part and apply Algorithm 2 to multistage stochastic optimization problem.

3.3 Discrete Approximation Methods

When random vectors follow a continuous distribution, the PHA cannot be applied to the two-stage SVI. In this case, Chen et al. [30] proposed a discrete approximation method for two-stage SLCP and Chen et al. [31] investigated the sample average approximation (SAA) method for two-stage stochastic generalized equation (SGE).

Discrete approximation for two-stage SLCP For the two-stage SLCP (2.18)–(2.19), Chen et al. [30] firstly investigated the existence and uniqueness of a solution under the assumptions as follows.

Assumption 3 There exists a positive continuous function $\kappa(\xi)$ such that $E[\kappa(\xi)] < +\infty$ and for a.e. ξ ,

$$(z^T, u^T) \begin{pmatrix} A & B(\xi) \\ N(\xi) & M(\xi) \end{pmatrix} \begin{pmatrix} z \\ u \end{pmatrix} \geq \kappa(\xi)(\|z\|^2 + \|u\|^2), \quad \forall z \in \mathbb{R}^n, u \in \mathbb{R}^m. \quad (3.13)$$

Moreover, $E[\|B(\xi)\|] < \infty$, $E[\|M(\xi)\|] < \infty$, $E[\|N(\xi)\|] < \infty$ and $E[\|q(\xi)\|] < \infty$.

Under Assumption 3, some properties of the two-stage SLCP are given in [30] as follows:

Proposition 1 Let Assumption 3 hold. For any given x and $\xi \in \Xi$, let $D(x, \xi)$ be an m -dimensional diagonal matrix with

$$D_{jj}(x, \xi) := \begin{cases} 1, & \text{if } (M(\xi)y(\xi) + N(\xi)x + q_2(\xi))_j \leq y_j(\xi), \\ 0, & \text{otherwise.} \end{cases}$$

Let

$$W(x, \xi) := [I - D(x, \xi)(I - M(\xi))]^{-1}D(x, \xi) \quad (3.14)$$

and

$$J(x, \xi) := \{j : (M(\xi)y(\xi) + N(\xi)x + q_2(\xi))_j \leq y_j(\xi)\}.$$

Then, the following assertions hold.

- (i) The two-stage SLCP (2.18)–(2.19) has a unique solution $(x^*, y^*(\cdot)) \in \mathbb{R}^n \times \mathcal{Y}$.
- (ii) The solution to the second stage of SLCP (2.18)–(2.19) can be written as

$$\bar{y}(x, \xi) = -W(x, \xi)(N(\xi)x + q_2(\xi)) \quad (3.15)$$

and $\bar{y}(\cdot, \xi)$ is globally Lipschitz continuous w.r.t x .

- (iii) The first equation of SLCP (2.18)–(2.19) can be reformulated as

$$\begin{aligned} 0 \leq x \perp & (A - E[B(\xi)W(x, \xi)N(\xi)])x \\ & - E[B(\xi)W(x, \xi)q_2(\xi)] + q_1 \geq 0, \end{aligned} \quad (3.16)$$

where

$$\|(A - E[B(\xi)W(x, \xi)N(\xi)])^{-1}\| \leq \frac{1}{E[\kappa(\xi)]} < +\infty.$$

(iv) Let

$$F(x) := \min \left(x, (A - E[B(\xi)W(x, \xi)N(\xi)])x - E[B(\xi)W(x, \xi)q_2(\xi)] + q_1 \right). \tag{3.17}$$

Then, F is Lipschitz continuous and every matrix V_x in the Clarke generalized Jacobian $\partial F(x)$ (see definition in [32, Section 2.6]) is nonsingular with $\|V_x^{-1}\| \leq \bar{d}$ for some constant $\bar{d} > 0$ which is independent of x .

Besides the existence and uniqueness of the two-stage SVI, the globally Lipschitz continuity and formulation (3.15) of $\bar{y}(\cdot, \xi)$ allow us to substitute $\bar{y}(\cdot, \xi)$ into the first-stage stochastic function $Ax + E[B(\xi)y(\xi)] + q_1$ and rewrite the two-stage SVI as one-stage SVI (3.16). Then, they proposed their discrete approximation method in [30] as follows.

Suppose \mathcal{E} is a compact and convex set. Let $\{\mathcal{E}_i^K\}$ be a partition of the support set \mathcal{E} , where \mathcal{E}_i^K is a compact and convex subset of \mathcal{E} such that

$$\bigcup_{i=1}^K \mathcal{E}_i^K = \mathcal{E}, \quad \text{int}\mathcal{E}_i^K \cap \text{int}\mathcal{E}_j^K = \emptyset, \quad \forall i \neq j, \quad i, j = 1, \dots, K,$$

where $\text{int}S$ denotes the interior of S and K denotes the number of partitions. Note that since \mathcal{E} is assumed to be a compact set, each \mathcal{E}_i^K is also a compact set. Let

$$E_{\mathcal{E}_i^K}[H(\xi)] := \frac{1}{p_i^K} \int_{\xi \in \mathcal{E}_i^K} H(\xi) \mathcal{P}(d\xi) \quad \text{with} \quad p_i^K = \mathcal{P}(\mathcal{E}_i^K) \tag{3.18}$$

for $H(\xi) = M(\xi), N(\xi), B(\xi)$ or $q_2(\xi)$. Let

$$\Delta(\mathcal{E}_i^K) := \max_{\xi_1, \xi_2 \in \mathcal{E}_i^K} \|\xi_1 - \xi_2\| \tag{3.19}$$

denote the diameter of \mathcal{E}_i^K . We require $\max_{i \in \bar{K}} \Delta(\mathcal{E}_i^K) \rightarrow 0$ as $K \rightarrow \infty$. Let $p_i^K := P(\mathcal{E}_i^K)$. Then, a discrete approximation of two-stage SLCP (2.18)–(2.19) is

$$0 \leq x \perp Ax + \sum_{i=1}^K p_i^K E_{\mathcal{E}_i^K}[B(\xi)]y_i + q_1 \geq 0, \tag{3.20}$$

$$0 \leq y_i \perp E_{\mathcal{E}_i^K}[M(\xi)]y_i + E_{\mathcal{E}_i^K}[N(\xi)]x + E_{\mathcal{E}_i^K}[q_2(\xi)] \geq 0, \tag{3.21}$$

$i \in 1, \dots, K.$

Moreover, let (x^K, \mathbf{y}^K) denote the solution of (3.20)–(3.21) whereby we write \mathbf{y}^K for (y_1^K, \dots, y_K^K) . Let

$$y^K(\xi) := \sum_{i=1}^K y_i^K \mathbf{1}_{\mathcal{E}_i^K}(\xi). \tag{3.22}$$

The following theorem states the convergence of $(x^K, y^K(\xi))$ to $(x^*, y^*(\xi))$, the true solution of the two-stage SLCP (2.18)–(2.19), as $K \rightarrow \infty$.

Theorem 5 Under Assumption 3, the following assertions hold.

- (i) The complementarity problem (3.20)–(3.21) has a unique solution (x^K, y^K) .
- (ii) If, in addition, $\max_{i \in \{1, \dots, K\}} \Delta(\mathcal{E}_i^K) \rightarrow 0$, then $\{(x^K, y^K(\cdot))\}$ is bounded on $\mathbb{R}^n \times \mathcal{Y}$, where the boundedness of $y^K(\cdot)$ is in the sense of the norm topology of $\mathcal{L}_1(\mathcal{Y})$.
- (iii) $\{x^K, y^K(\cdot)\}$ converges to the true solution $(x^*, y^*(\cdot))$ of problem (2.18)–(2.19), where the convergence of $\{y^K(\cdot)\} \rightarrow y^*(\cdot)$ is in the sense of the norm topology of $\mathcal{L}_2(\mathcal{Y})$.

To establish the quantitative convergence analysis, we need more assumptions.

Assumption 4 $M(\cdot), N(\cdot), q_2(\cdot)$ and $B(\cdot)$ are Lipschitz continuous over a compact set containing \mathcal{E} with Lipschitz constant L .

Theorem 6 Under Assumptions 3 and 4, there exists a positive number $\gamma \geq 0$ and nonnegative integrably bounded functions $c(\xi)$ and $h(\xi)$ such that

$$\|x^K - x^*\| \leq \gamma E[\|B(\xi)\|c(\xi)]L \max_{i \in \bar{K}} \Delta(\mathcal{E}_i^K) \tag{3.23}$$

and

$$\|y^K(\xi) - y^*(\xi)\| \leq h(\xi)L \max_{i \in \bar{K}} \Delta(\mathcal{E}_i^K), \text{ for a.e. } \xi \in \mathcal{E}. \tag{3.24}$$

Sample average approximation method for two-stage SGE. In Chen et al. [31], the authors considered the following two-stage (SGE):

$$0 \in E[\Phi(x, y(\xi), \xi)] + \Gamma_1(x), \quad x \in X, \tag{3.25}$$

$$0 \in \Psi(x, y(\xi), \xi) + \Gamma_2(y(\xi), \xi), \quad \text{for a.e. } \xi \in \mathcal{E}, \tag{3.26}$$

where $X \subseteq \mathbb{R}^n$ is a nonempty closed convex set, ξ is a random vector with support set $\mathcal{E} \subset \mathbb{R}^d$ defined as above, $\Phi : \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^d \rightarrow \mathbb{R}^n$ and $\Psi : \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^d \rightarrow \mathbb{R}^m$ such that $\Phi(\cdot, \cdot, \xi)$ and $\Psi(\cdot, \cdot, \xi)$ are Lipschitz continuous with Lipschitz modulus $\kappa_\Phi(\xi)$ and $\kappa_\Psi(\xi)$, respectively, $\Gamma_1 : \mathbb{R}^n \rightrightarrows \mathbb{R}^n, \Gamma_2 : \mathbb{R}^m \times \mathcal{E} \rightrightarrows \mathbb{R}^m$ are multifunctions (point-to-set mappings) and $y(\cdot) \in \mathcal{Y}$ with \mathcal{Y} being the space of measurable functions from \mathcal{E} to \mathbb{R}^m such that the expected value in (3.25) is well defined.

Without assuming relatively complete recourse, the authors of [31] studied convergence and the exponential rate of convergence of the SAA

$$0 \in N^{-1} \sum_{j=1}^N \Phi(x, y_j, \xi^j) + \Gamma_1(x), \quad x \in X, \tag{3.27}$$

$$0 \in \Psi(x, y_j, \xi^j) + \Gamma_2(y_j, \xi^j), \quad j = 1, \dots, N \tag{3.28}$$

of the two-stage SGE (2.16)–(2.17) with y_j being a copy of the second-stage vector for $\xi = \xi^j, j = 1, \dots, N$, where N denotes the sample size, ξ^1, \dots, ξ^N is an independent identically distributed (iid) sample of random vector ξ .

To investigate the convergence analysis without assuming *relatively complete recourse*, some notations and assumptions are needed. Denote by \mathcal{X} the set of $x \in X$ such that the second-stage generalized equation (3.26) has a solution. For every $\xi \in \Xi$, by $\bar{\mathcal{X}}(\xi)$, we denote the set of $x \in X$ such that the second-stage problem

$$0 \in \Psi(x, y, \xi) + \Gamma_2(y, \xi) \tag{3.29}$$

has a solution, $\bigcap_{\xi \in \Xi} \bar{\mathcal{X}}(\xi) = \mathcal{X}$ and $\bar{\mathcal{X}}_N := \bigcap_{j=1}^N \bar{\mathcal{X}}(\xi^j)$ the set of x such that problem (3.28) has a solution. The condition of relatively complete recourse means that $X = \bar{\mathcal{X}}_N$.

Assumption 5 For a.e. $\xi \in \Xi$, problem (3.29) has a unique solution for any $x \in \mathcal{X}$.

Assumption 6 For every ξ and $x \in \bar{\mathcal{X}}(\xi)$, there is a neighborhood \mathcal{V} of x and a measurable function $v(\xi)$ such that $\|\hat{y}(x', \xi)\| \leq v(\xi)$ for all $x' \in \mathcal{V} \cap \bar{\mathcal{X}}(\xi)$.

Lemma 2 Suppose that Assumptions 5 and 6 hold, and for a.e. $\xi \in \Xi$ the multifunction $\Gamma_2(\cdot, \xi)$ is closed. Then for a.e. $\xi \in \Xi$, the solution $\hat{y}(x, \xi)$ is a continuous function of $x \in \mathcal{X}$.

By Lemma 2, the first-stage problem (3.25) can be written as the following generalized equation:

$$0 \in \phi(x) + \Gamma_1(x), \quad x \in \mathcal{X}, \tag{3.30}$$

where

$$\hat{\Phi}(x, \xi) := \Phi(x, \hat{y}(x, \xi), \xi) \text{ and } \phi(x) := E[\hat{\Phi}(x, \xi)] \tag{3.31}$$

are continuous functions. Then, consider the SAA problem (3.27)–(3.28). Similar as (3.30), by Lemma 2 the SAA problem (3.27)–(3.28) can be written as

$$0 \in \hat{\phi}_N(x) + \Gamma_1(x), \quad x \in \bar{\mathcal{X}}_N, \tag{3.32}$$

where $\hat{\phi}_N(x) := N^{-1} \sum_{j=1}^N \hat{\Phi}(x, \xi^j)$ with $\hat{\Phi}(x, \xi)$ defined in (3.31). Denote by \mathcal{S}^* the set of solutions of the first-stage problem (3.30) and by $\hat{\mathcal{S}}_N$ the solution set of the SAA problem (3.32).

For $\delta \in (0, 1)$, consider a compact set $\bar{\Xi}_\delta \subset \Xi$ such that $\mathcal{P}(\bar{\Xi}_\delta) \geq 1 - \delta$, and the multifunction $\Delta_\delta : X \rightrightarrows \bar{\Xi}_\delta$ defined as

$$\Delta_\delta(x) := \{\xi \in \bar{\Xi}_\delta : x \in \bar{\mathcal{X}}(\xi)\}. \tag{3.33}$$

Assumption 7 For any $\delta \in (0, 1)$, the multifunction $\Delta_\delta(\cdot)$ is outer semicontinuous.

The following lemma shows that this assumption holds under mild conditions.

Lemma 3 Suppose $\Psi(\cdot, \cdot, \cdot)$ is continuous, $\Gamma_2(\cdot, \cdot)$ is closed and Assumption 6 holds. Then, the multifunction $\Delta_\delta(\cdot)$ is outer semicontinuous.

Then, the almost sure convergence result is given as follows:

Theorem 7 Suppose that: (i) Assumptions 5–7 hold, (ii) the multifunctions $\Gamma_1(\cdot)$ and $\Gamma_2(\cdot, \xi)$, $\xi \in \mathcal{E}$, are closed, (iii) there is a compact subset X' of X such that $\mathcal{S}^* \subset X'$ and w.p.1 for all N large enough the set $\hat{\mathcal{S}}_N$ is nonempty and is contained in X' , (iv) $\|\hat{\Phi}(x, \xi)\|_{x \in \mathcal{X}}$ is dominated by an integrable function, (v) the set \mathcal{X} is nonempty. Let $\vartheta_N := \mathbb{D}(\mathcal{X}_N \cap X', \mathcal{X} \cap X')$. Then, \mathcal{S}^* is nonempty and the following statements hold.

- (a) $\vartheta_N \rightarrow 0$ and $\mathbb{D}(\hat{\mathcal{S}}_N, \mathcal{S}^*) \rightarrow 0$ w.p.1 as $N \rightarrow \infty$.
- (b) In addition, assume that: (vi) for any $\delta > 0$, $\tau > 0$ and a.e. $\omega \in \Omega$, there exist $\gamma > 0$ and N_0 such that for any $x \in \mathcal{X} \cap X' + \gamma \mathcal{B}$ and $N \geq N_0$, there exists $z_x \in \mathcal{X} \cap X'$ such that¹

$$\|z_x - x\| \leq \tau, \quad \Gamma_1(x) \subseteq \Gamma_1(z_x) + \delta \mathcal{B}, \quad \text{and} \quad \|\hat{\phi}_N(z_x) - \hat{\phi}_N(x)\| \leq \delta. \quad (3.34)$$

Then, w.p.1 for N large enough it follows that

$$\mathbb{D}(\hat{\mathcal{S}}_N, \mathcal{S}^*) \leq \tau + R^{-1} \left(\sup_{x \in \mathcal{X} \cap X'} \|\phi(x) - \hat{\phi}_N(x)\| \right), \quad (3.35)$$

where for $\epsilon \geq 0$ and $t \geq 0$,

$$R(\epsilon) := \inf_{x \in \mathcal{X} \cap X', d(x, \mathcal{S}^*) \geq \epsilon} d(0, \phi(x) + \Gamma_1(x)),$$

$$R^{-1}(t) := \inf\{\epsilon \in \mathbb{R}_+ : R(\epsilon) \geq t\}.$$

Note that in the case when $\Gamma_1(\cdot) := \mathcal{N}_D(\cdot)$ with a nonempty polyhedral convex set D , the first and second inequalities of (3.34) hold automatically.

To drive the exponential rate of convergence based on uniform large deviations theorem (cf., [33–35]), more assumptions are needed.

Assumption 8 For a.e. $\xi \in \mathcal{E}$, there exists a unique, parametrically CD-regular [36] solution $\bar{y} = \hat{y}(\bar{x}, \xi)$ of the second-stage generalized equation (2.17) for all $\bar{x} \in \mathcal{X}$.

Assumption 9 The set \mathcal{X} is convex, its interior $\text{int}(\mathcal{X}) \neq \emptyset$, and for a.e. $\xi \in \mathcal{E}$, the generalized equation,

$$0 \in G_{\bar{x}}(y) = \Psi(\bar{x}, \bar{y}, \xi) + J(y - \bar{y}) + \Gamma_2(y, \xi), \quad \text{for which} \quad G_{\bar{x}}(\bar{y}) \ni 0,$$

has a locally Lipschitz continuous solution function at 0 for \bar{y} with Lipschitz constant $\kappa_G(\bar{x}, \xi)$ for any $\bar{x} \in \mathcal{X}$ and there exists a measurable function $\bar{\kappa}_G : \mathcal{E} \rightarrow \mathbb{R}_+$ such that $\kappa_G(\bar{x}, \xi) \leq \bar{\kappa}_G(\xi)$ and $E[\bar{\kappa}_G(\xi)\kappa_\Psi(\xi)] < \infty$.

Let

$$M_x^i(t) := E \left\{ \exp(t[\hat{\Phi}_i(x, \xi) - \phi_i(x)]) \right\}$$

¹ Recall that $\hat{\phi}_N(x) = \hat{\phi}_N(x, \omega)$ are random functions defined on the probability space $(\Omega, \mathcal{F}, \mathcal{P})$.

be the moment generating function of the random variable $\hat{\Phi}_i(x, \xi) - \phi_i(x)$, $i = 1, \dots, n$, and

$$M_\kappa(t) := E \left\{ \exp \left(t \left[\kappa_\Phi(\xi) + \kappa_\Phi(\xi)\kappa(\xi) - E \left[\kappa_\Phi(\xi) + \kappa_\Phi(\xi)\kappa(\xi) \right] \right] \right) \right\}.$$

Assumption 10 For every $x \in \mathcal{X}$ and $i = 1, \dots, n$, the moment generating functions $M_x^i(t)$ and $M_\kappa(t)$ have finite values for all t in a neighborhood of zero.

Theorem 8 Suppose: (i) Assumptions 5, 7–10 hold, (ii) S^* is nonempty and w.p.1 for N large enough, \hat{S}_N are nonempty, (iii) the multifunctions $\Gamma_1(\cdot)$ and $\Gamma_2(\cdot, \xi)$, $\xi \in \mathcal{E}$, are closed and monotone. Then, the following statements hold.

(a) For sufficiently small $\epsilon > 0$, there exist positive constants $\varrho = \varrho(\epsilon)$ and $\zeta = \zeta(\epsilon)$, independent of N , such that

$$P \left\{ \sup_{x \in \mathcal{X}} \|\hat{\phi}_N(x) - \phi(x)\| \geq \epsilon \right\} \leq \varrho(\epsilon)e^{-N\zeta(\epsilon)}. \tag{3.36}$$

(b) Assume in addition: (iv) the condition of part (b) in Theorem 7 holds and w.p.1 for N sufficiently large,

$$S^* \cap \text{cl}(\text{bd}(\mathcal{X}) \cap \text{int}(\bar{\mathcal{X}}_N)) = \emptyset. \tag{3.37}$$

(v) $\phi(\cdot)$ has the following strong monotonicity property for every $x^* \in S^*$:

$$(x - x^*)^\top (\phi(x) - \phi(x^*)) \geq g(\|x - x^*\|), \quad \forall x \in \mathcal{X}, \tag{3.38}$$

where $g : \mathbb{R}_+ \rightarrow \mathbb{R}_+$ is such a function that function $\tau(\tau) := g(\tau)/\tau$ is monotonically increasing for $\tau > 0$.

Then, $S^* = \{x^*\}$ is a singleton, and for any sufficiently small $\epsilon > 0$, there exists N sufficiently large such that

$$P \left\{ \mathbb{D}(\hat{S}_N, S^*) \geq \epsilon \right\} \leq \varrho \left(\tau^{-1}(\epsilon) \right) \exp \left(-N\zeta \left(\tau^{-1}(\epsilon) \right) \right), \tag{3.39}$$

where $\varrho(\cdot)$ and $\zeta(\cdot)$ are defined in (3.36), and $\tau^{-1}(\epsilon) := \inf \{ \tau > 0 : \tau(\tau) \geq \epsilon \}$ is the inverse of $\tau(\tau)$.

Moreover, Chen et al. [31] investigated the convergence properties of the two-stage SGE (2.16)–(2.17) when $\Phi(x, y, \xi)$ and $\Psi(x, y, \xi)$ are continuously differentiable w.r.t. (x, y) for a.e. $\xi \in \mathcal{E}$ and $\Gamma_1(x) := \mathcal{N}_D(x)$ and $\Gamma_2(y) := \mathcal{N}_{\mathbb{R}_+^m}(y)$ with $D \subseteq \mathbb{R}^p$ being a nonempty, polyhedral, convex set. That is, they considered the mixed two-stage SVI-NCP

$$0 \in E[\Phi(x, y(\xi), \xi)] + \mathcal{N}_D(x), \tag{3.40}$$

$$0 \leq y(\xi) \perp \Psi(x, y(\xi), \xi) \geq 0, \quad \text{for a.e. } \xi \in \mathcal{E}, \tag{3.41}$$

and studied convergence analysis of its SAA problem

$$0 \in N^{-1} \sum_{j=1}^N \Phi(x, y(\xi^j), \xi^j) + \mathcal{N}_D(x), \tag{3.42}$$

$$0 \leq y(\xi^j) \perp \Psi(x, y(\xi^j), \xi^j) \geq 0, \quad j = 1, \dots, N. \tag{3.43}$$

They first investigated the properties of the second-stage problems.

Assumption 11 For a.e. $\xi \in \mathcal{E}$ and all $x \in \mathcal{X} \cap D$, $\Psi(x, \cdot, \xi)$ is strongly monotone, that is, there exists a positive-valued measurable $\kappa_y(\xi)$ such that for all $y, u \in \mathbb{R}^m$,

$$\langle \Psi(x, y, \xi) - \Psi(x, u, \xi), y - u \rangle \geq \kappa_y(\xi) \|y - u\|^2$$

with $E[\kappa_y(\xi)] < +\infty$.

Theorem 9 Let $\Psi : \mathbb{R}^n \times \mathbb{R}^m \times \mathcal{E} \rightarrow \mathbb{R}^m$ be Lipschitz continuous and continuously differentiable over $\mathbb{R}^n \times \mathbb{R}^m$ for a.e. $\xi \in \mathcal{E}$. Suppose Assumption 11 holds and $\Phi(x, y, \xi)$ is continuously differentiable w.r.t. (x, y) for a.e. $\xi \in \mathcal{E}$. Then, for a.e. $\xi \in \mathcal{E}$ and $x \in \mathcal{X}$, the following holds.

- (a) The second-stage SNCP (3.41) has a unique solution $\hat{y}(x, \xi)$ which is parametrically CD-regular and the mapping $x \mapsto \hat{y}(x, \xi)$ is Lipschitz continuous over $\mathcal{X} \cap X'$, where X' is a compact subset of \mathbb{R}^n .
- (b) The Clarke Jacobian of $\hat{y}(x, \xi)$ w.r.t. x is as follows

$$\begin{aligned} \partial \hat{y}(x, \xi) &= \text{conv} \left\{ \lim_{z \rightarrow x} \nabla_z \hat{y}(z, \xi) : \nabla_z \hat{y}(z, \xi) \right. \\ &\quad \left. = -[I - J_\alpha(I - M(z, \hat{y}(z, \xi), \xi))]^{-1} D_\alpha L(z, \hat{y}(z, \xi), \xi) \right\}, \end{aligned}$$

where $M(x, y, \xi) = \nabla_y \Psi(x, y, \xi)$, $L(x, \hat{y}(x, \xi), \xi) = \nabla_x \Psi(x, \hat{y}(x, \xi), \xi)$,

$$\alpha = \{i : (\hat{y}(x, \xi))_i > (\Psi(x, \hat{y}(x, \xi), \xi))_i\},$$

J_α is an m -dimensional diagonal matrix and

$$(J_\alpha)_{jj} := \begin{cases} 1, & \text{if } j \in \alpha, \\ 0, & \text{otherwise.} \end{cases} \tag{3.44}$$

Under Assumption 11, the two-stage SVI-NCP can be reformulated as a single-stage SVI with $\hat{\Phi}(x, \xi) = \Phi(x, \hat{y}(x, \xi), \xi)$ and $\phi(x) = E[\hat{\Phi}(x, \xi)]$ as follows

$$0 \in \phi(x) + \mathcal{N}_C(x). \tag{3.45}$$

To investigate the properties of two-stage SVI-NCP, more assumptions are needed. Let

$$\Theta(x, y(\xi), \xi) = \begin{pmatrix} \Phi(x, y(\xi), \xi) \\ \Psi(x, y(\xi), \xi) \end{pmatrix}$$

and $\nabla\Theta(x, y, \xi)$ be the Jacobian of Θ . Then,

$$\nabla\Theta(x, y, \xi) = \begin{pmatrix} A(x, y, \xi) & B(x, y, \xi) \\ L(x, y, \xi) & M(x, y, \xi) \end{pmatrix},$$

where $A(x, y, \xi) = \nabla_x\Phi(x, y, \xi)$, $B(x, y, \xi) = \nabla_y\Phi(x, y, \xi)$, $L(x, y, \xi) = \nabla_x\Psi(x, y, \xi)$ and $M(x, y, \xi) = \nabla_y\Psi(x, y, \xi)$.

Assumption 12 For a.e. $\xi \in \mathcal{E}$, $\Theta(x, y(\xi), \xi)$ is strongly monotone with parameter $\kappa(\xi)$ at $(x, y(\cdot)) \in C \times \mathcal{Y}$, where $E[\kappa(\xi)] < +\infty$.

Theorem 10 Let Sol^* be the solution set of the mixed SVI-NCP (3.40)–(3.41). Suppose (i) there exists a compact set X' such that $\text{Sol}^* \cap X' \times \mathcal{Y}$ is nonempty, (ii) Assumption 12 holds over $\text{Sol}^* \cap X' \times \mathcal{Y}$ and (iii) assume

$$E[\|A(x, \hat{y}(x, \xi), \xi) - B(x, \hat{y}(x, \xi), \xi)M(x, \hat{y}(x, \xi), \xi)^{-1}L(x, \hat{y}(x, \xi), \xi)\|] < +\infty \tag{3.46}$$

over $X \cap X'$. Then,

- (a) For any $(x, y(\cdot)) \in \text{Sol}^*$, every matrix in $\partial\hat{\Phi}(x)$ is positive definite, and $\hat{\Phi}$ and ϕ are strongly monotone at x .
- (b) Any solution $x^* \in \mathcal{S}^* \cap X'$ of SVI (3.45) is CD-regular and an isolate solution.
- (c) Moreover, if replacing conditions (i) and (ii) by (iv) Assumption 12 holds over $\mathbb{R}^n \times \mathcal{Y}$, then SVI (3.45) has a unique solution x^* and the solution is CD-regular.

Here the definition of CD-regular can be found in [36].

Then, the properties and the convergence analysis of SAA problem can be investigated. Define

$$\mathcal{G}_N(x, y(\cdot)) := \begin{pmatrix} N^{-1} \sum_{j=1}^N \Phi(x, y(\xi^j), \xi^j) \\ \Psi(x, y(\xi^1), \xi^1) \\ \vdots \\ \Psi(x, y(\xi^N), \xi^N) \end{pmatrix}.$$

Theorem 11 Suppose Assumption 12 holds over $C \times \mathcal{Y}$ and $\Phi(x, y, \xi)$, and $\Psi(x, y, \xi)$ are continuously differentiable w.r.t. (x, y) for a.e. $\xi \in \mathcal{E}$. Then,

- (a) $\mathcal{G}_N : C \times \mathcal{Y} \rightarrow C \times \mathcal{Y}$ is strongly monotone with $N^{-1} \sum_{j=1}^N \kappa(\xi^j)$ and hemicontinuous.
- (b) The SAA two-stage SVI (3.42)–(3.43) has a unique solution.

Note that function $\hat{\Phi}(x, \xi) = \Phi(x, \hat{y}(x, \xi), \xi)$, where $\hat{y}(x, \xi)$ is a solution of the second-stage problem (3.41). Then, the first stage of SAA problem with second-stage solution can be written as

$$0 \in N^{-1} \sum_{j=1}^N \hat{\Phi}(x, \xi^j) + \mathcal{N}_D(x). \tag{3.47}$$

In what follows, the almost sure and exponential rate of convergence are established in [31] as follows.

Theorem 12 *Suppose conditions (i)–(iii) of Theorem 10 hold. Let x^* be a solution of SVI (3.45) and X' be a compact set such that $x^* \in \text{int}(X')$. Assume there exists $\epsilon > 0$ such that for N sufficiently large,*

$$x^* \notin \text{cl}(\text{bd}(\mathcal{X}) \cap \text{int}(\bar{\mathcal{X}}_N \cap X')). \quad (3.48)$$

Then, there exists a solution \hat{x}_N of the SAA problem (3.47) and a positive scalar δ such that $\|\hat{x}_N - x^\| \rightarrow 0$ as $N \rightarrow \infty$ w.p.1 and for N sufficiently large w.p.1*

$$\|\hat{x}_N - x^*\| \leq \delta \sup_{x \in \mathcal{X} \cap X'} \|\hat{\phi}_N(x) - \phi(x)\|. \quad (3.49)$$

Theorem 13 *Let $X' \subset D$ be a convex compact subset such that $\mathcal{B}_\delta(x^*) \subset X'$. Suppose the conditions in Theorem 12 and Assumption 10 hold. Then for any $\epsilon > 0$, there exist positive constants $\delta > 0$ (independent of ϵ), $\varrho = \varrho(\epsilon)$ and $\zeta = \zeta(\epsilon)$ (independent of N) such that*

$$P \left\{ \sup_{x \in \mathcal{X}} \|\hat{\phi}_N(x) - \phi(x)\| \geq \epsilon \right\} \leq \varrho(\epsilon) e^{-N\zeta(\epsilon)}, \quad (3.50)$$

and

$$P \left\{ \|x_N - x^*\| \geq \epsilon \right\} \leq \varrho(\epsilon/\delta) e^{-N\zeta(\epsilon/\delta)}. \quad (3.51)$$

4 Applications

Two-stage SVI has wide applications in economics, traffic network, electricity markets, supply chain problems, finance, risk management under uncertain environment. One type of two-stage SVI involves making a “here-and-now” decisions at the present time to meet the uncertainty that are revealed at a later time. This is one of a motivation of both two-stage stochastic optimization and two-stage stochastic Nash equilibrium problem (SNEP), such as Example 2. [37] discussed a scenario-based dynamic oligopolistic problem under uncertainty. In electricity market, [15,38–40] considered capacity expansion problem under uncertain environment. [41] investigated the supply-side risk in uncertainty power market. [42,43] discussed two-settlement markets consisting of a deterministic (first-stage) spot market and a stochastic (second-stage) market known as the forward. [44] presented a stochastic complementarity model of equilibrium in an electric power market with uncertainty power demand. [45] presented Nash equilibrium models of perfectly competitive capacity expansion involving risk-averse participants in the presence of discrete state uncertainties and pricing mechanisms of different kinds. [46] modeled a production and supply competition of a homogenous product under uncertainty in an oligopolistic market by a two-stage SVI, and used the model to describe the market share observation in the world market of crude oil. Here we introduce two important applications which form

traffic equilibrium problems and noncooperative multi-agent games under uncertain environment.

Traffic equilibrium problems Recently, Chen et al. [20] considered a traffic equilibrium problem under uncertain environment (uncertain capacity and demand) by following the Ferris–Pang multicommodity formulation [47] which associates a (different) commodity with each destination node $d \in \mathcal{D} \subset \mathcal{G}$. Let $N(\mathcal{G}, \mathcal{A})$ be an oriented network, the random vectors $c_\xi^j := ((c_\xi^j)_1, \dots, (c_\xi^j)_{|\mathcal{A}|})^T$ with $(c_\xi^j)_a$ the maximum flow capacity for each commodity j arc a and demand $d_\xi^j = ((d_\xi^j)_1, (d_\xi^j)_2, \dots, (d_\xi^j)_{|\mathcal{G}|})$ with $(d_\xi^j)_i$ the demand for each commodity j at each node $i, a = 1, \dots, |\mathcal{A}|, j = 1, \dots, |\mathcal{D}|, i = 1, \dots, |\mathcal{G}|$. Let h_ξ^{od} be the demand for each origin(O)-destination(D) pairs. R_ξ^{od} are all (acyclic) routes r connecting o to d with V being the arcs(a)/routes(r) incidence matrix, i.e., $V_{a,r} = 1$ if arc $a \in r$. A route flow $f_\xi = \{f_\xi^r, r \in \cup_{od} R_\xi^{od}\}$ results in an arc flow $u_\xi^a = \langle V_a, f_\xi \rangle$. Then, $Vu_\xi^j = d_\xi^j, u_\xi^j \geq 0, j \in \mathcal{D}$. Let

$$A = \begin{pmatrix} V & & \\ & \ddots & \\ & & V \end{pmatrix} \in \mathbb{R}^{|\mathcal{D}||\mathcal{N}| \times |\mathcal{D}||\mathcal{A}|}, \quad u_\xi = \begin{pmatrix} u_\xi^1 \\ \vdots \\ u_\xi^{|\mathcal{D}|} \end{pmatrix} \in \mathbb{R}^{|\mathcal{D}||\mathcal{A}|},$$

$$b_\xi = \begin{pmatrix} d_\xi^1 \\ \vdots \\ d_\xi^{|\mathcal{D}|} \end{pmatrix} \in \mathbb{R}^{|\mathcal{D}||\mathcal{N}|}, \quad c_\xi = \begin{pmatrix} c_\xi^1 \\ \vdots \\ c_\xi^{|\mathcal{A}|} \end{pmatrix} \in \mathbb{R}^{|\mathcal{D}||\mathcal{A}|}.$$

Then, the flow conservation constraints for each realization ξ can be written as

$$C(\xi) := \{Au_\xi = b_\xi, 0 \leq u_\xi, Pu_\xi \leq c_\xi\},$$

where $P = (I, \dots, I) \in \mathbb{R}^{|\mathcal{A}| \times |\mathcal{D}||\mathcal{A}|}$ and I is the $|\mathcal{A}| \times |\mathcal{A}|$ identity matrix. We then consider the cost of the traffic equilibrium problem. The arc travel time function $h(\xi, \cdot) : \mathcal{R}^{|\mathcal{D}||\mathcal{A}|} \rightarrow \mathcal{R}^{|\mathcal{A}|}$ is a stochastic vector and each of its entries $h_a(u_\xi, \xi)$ is assumed to follow a generalized Bureau of Public Roads (GBPR) function,

$$h_a(\xi, u_\xi) = \left(\eta_a + \tau_a \left(\frac{(Pu_\xi)_a}{(\gamma_\xi)_a} \right)^{n_a} \right), \quad a = 1, \dots, |\mathcal{A}|,$$

where $\eta_a, \tau_a, (\gamma_\xi)^a$ are given positive parameters. Let $F(u_\xi, \xi) = P^T h(\xi, u_\xi)$. Then for $n_a = 1$

$$\nabla_u F(u_\xi, \xi) = P^T \text{diag} \left(\frac{\tau_a}{(\gamma_\xi)_a} \right) P$$

is symmetric positive semi-definite for any $u_\xi \in \mathbb{R}_+^{|\mathcal{D}||\mathcal{A}|} \supseteq C(\xi)$. Then, the stochastic VI formulation for Wardrop’s user equilibrium seeks an equilibrium arc flow $u_\xi \in C(\xi)$ for a known event $\xi \in \mathcal{E}$, such that

$$- F(\xi, u_\xi) \in_{a.s.} \mathcal{N}_{C(\xi)}(u_\xi). \tag{3.1}$$

Moreover, let

$$x = \begin{pmatrix} x^1 \\ \vdots \\ x^{|\mathcal{D}|} \end{pmatrix} \in \mathbb{R}^{|\mathcal{D}||\mathcal{A}|}, \quad \bar{b} = \begin{pmatrix} \bar{d}^1 \\ \vdots \\ \bar{d}^{|\mathcal{D}|} \end{pmatrix} \in \mathbb{R}^{|\mathcal{D}||\mathcal{N}|}, \quad \bar{c}_\xi = \begin{pmatrix} \bar{c}^1 \\ \vdots \\ \bar{c}^{|\mathcal{A}|} \end{pmatrix} \in \mathbb{R}^{|\mathcal{D}||\mathcal{A}|},$$

and

$$D := \{Ax = \bar{b}, 0 \leq x, Px \leq \bar{c}\}, \quad G(x) = P^T \bar{h}(x),$$

where $\bar{d}^i = E[d_\xi^i], i = 1, \dots, |\mathcal{D}|, \bar{c}^a = E[c_\xi^a], a = 1, \dots, |\mathcal{A}|$ and \bar{h} is defined by

$$\bar{h}_a(x) = \eta_a + \tau(Px)_a^{n_a} E[(\gamma_\xi)_a^{-n_a}], \quad a = 1, \dots, |\mathcal{A}|.$$

Then, the deterministic VI formulation for Wardrop’s user equilibrium seeks a forecast arc flows $x \in D$ satisfying

$$- G(x) \in \mathcal{N}_D(x). \tag{3.2}$$

Similar as (2.22)–(2.23), we can consider (3.1)–(3.2) as a two-stage SVI and reformulate it as a stochastic MPEC:

$$\begin{aligned} \min & E[\|u_\xi - x\|^2] \\ \text{s.t.} & 0 \in G(x) + \mathcal{N}_D(x), \quad 0 \in_{a.e.} F(u_\xi, \xi) + \mathcal{N}_{C(\xi)}(u_\xi), \quad \xi \in \Xi, \end{aligned} \tag{3.3}$$

and then, by ERM approach, reformulate it as

$$\begin{aligned} \min & \frac{1}{\lambda\rho} \theta(x) + \frac{1}{\rho} E[r(u_\xi, \xi)] + E[\|u_\xi - x\|^2] \\ \text{s.t.} & x \in D, \mu_\xi \in_{a.s.} C(\xi), \xi \in \Xi, \end{aligned} \tag{3.4}$$

where $\theta(\cdot)$ and $r(\cdot, \xi)$ are residual functions of $0 \in G(x) + \mathcal{N}_D(x)$ and $0 \in_{a.e.} F(u_\xi, \xi) + \mathcal{N}_{C(\xi)}(u_\xi)$, respectively.

Noncooperative multi-agent games In [48], Pang et al. formally introduced and studied a noncooperative multi-agent game under uncertainty and focused mainly on a two-stage setting of the game where each agent is risk-averse as follows.

Consider a noncooperative game with N risk-averse players, each of whom, labeled $i = 1, \dots, n$, has a private strategy set $X_i \subset \mathbb{R}^{s_i}$ that is closed and convex, a (deterministic) first-stage objective function θ_i that depends on all players’ strategies $x := \{x_i\}_{i=1}^n$ and a second-stage risk-averse recourse function $\phi_i(x) = \mathcal{R}_i(\psi_i(x_i, x_{-i}, \omega))$ with a kind of risk measure \mathcal{R}_i and a realization of a random vector ξ defined as above. Then, the two-stage noncooperative games with risk-averse players is

$$\min_{x_i \in X_i} \theta_i(x_i, x_{-i}) + \mathcal{R}_i(\psi_i(x_i, x_{-i}, \xi)) \tag{3.5}$$

and

$$\psi_i(x, \omega) := \min_{y_i \in Y_i} f_i(y_i, x, \xi), \quad (3.6)$$

for $i = 1, \dots, n$.

When the risk measure $\mathcal{R}_i = E$, the players are risk-neutral. Moreover, the risk measure $\mathcal{R}_i, i = 1, \dots, n$ can be considered as several deviation measures include the standard deviation, lower and upper semideviations, mean absolute (semi)deviations, absolute (semi)deviations, the median and deviations derived from the CVaR. They then investigated several properties of mean-deviation composite games with quadratic recourse, such as the continuity, regularization and differentiability of the second-stage optimal value function, and the reformulation of mean-deviation composite game.

Pang et al. [48] bypassed the SVI framework and dealt with the risk-averse SNEP based largely on smoothing, regularization, sampled solution approach and the best-response scheme. But now, by [23,24] and the sample average approximation method [31], we can solve the two-stage SVI by the progressive hedging method.

5 Final Remarks

Although the two-stage stochastic optimization has been investigated deeply and two-stage SNEP has been applied widely, the research of two-stage SVI and multistage SVI has just begun. There are several important questions have to be investigated, e.g.,

- Can we solve non-monotone two-stage SVI using PHM?
- How could we achieve better convergence rate by using sampling technology?
- How to extend the discrete approximation methods to multistage SVI when The random vector follows a continuous distribution?
- How to extend the SIV to dynamic two-stage SVI?
- How could we solve the two-stage or even multistage SVI more effectively?
- In the case when we only have limited information about the distribution of random vectors, could we model the two-stage variational inequalities in the sense of distributional robustness or even robustness?

In summary, there are many challenges and opportunities for developing both theoretical analysis and numerical algorithms when we handle the two-stage (multistage) VI in an uncertain environment.

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

References

- [1] Cottle, R.W., Pang, J.S., Stone, R.E.: The Linear Complementarity Problem. Academic Press, New York (1992)

- [2] Facchinei, F., Pang, J.S.: *Finite-Dimensional Variational Inequalities and Complementarity Problems*. Springer, New York (2003)
- [3] Lin, G.H., Fukushima, M.: Stochastic equilibrium problems and stochastic mathematical programs with equilibrium constraints: a survey. *Pac. J. Optim.* **6**, 455–482 (2010)
- [4] Shanbhag, U.: Stochastic variational inequality problems: applications, analysis, and algorithms. In: *INFORMS Tutorials in Operations Research*. pp. 71–107 (2013)
- [5] Chen, X., Wets, R.J.-B.: Special Issue: Stochastic Equilibrium and Variational Inequalities. *Math. Program.* **165** (2017)
- [6] Chen, X., Fukushima, M.: Expected residual minimization method for stochastic linear complementarity problems. *Math. Oper. Res.* **30**, 1022–1038 (2005)
- [7] Chen, X., Zhang, C., Fukushima, M.: Robust solution of monotone stochastic linear complementarity problems. *Math. Program.* **117**, 51–80 (2009)
- [8] Ling, C., Qi, L., Zhou, G., Caccetta, L.: The SC1 property of an expected residual function arising from stochastic complementarity problems. *Oper. Res. Lett.* **36**, 456–460 (2008)
- [9] Luo, M.J., Lin, G.H.: Expected residual minimization method for stochastic variational inequality problems. *J. Optim. Theory Appl.* **140**, 103–116 (2009)
- [10] Zhang, C., Chen, X.: Smoothing projected gradient method and its application to stochastic linear complementarity problems. *SIAM J. Optim.* **20**, 627–649 (2009)
- [11] Chen, X., Wets, R.J.-B., Zhang, Y.: Stochastic variational inequalities: residual minimization smoothing sample average approximations. *SIAM J. Optim.* **22**, 649–673 (2012)
- [12] King, A.J., Rockafellar, R.T.: Asymptotic theory for solutions in statistical estimation and stochastic programming. *Math. Oper. Res.* **18**, 148–162 (1993)
- [13] Jiang, H., Xu, H.: Stochastic approximation approaches to the stochastic variational inequality problem. *IEEE Trans. Autom. Control* **53**, 1462–1475 (2008)
- [14] Ravat, U., Shanbhag, U.V.: On the characterization of solution sets of smooth and nonsmooth convex stochastic Nash games. *SIAM J. Optim.* **21**, 1168–1199 (2011)
- [15] Gürkan, G., Özge, A.Y., Robinson, S.M.: Sample-path solution of stochastic variational inequalities. *Math. Program.* **84**, 313–333 (1999)
- [16] Fukushima, M.: Equivalent differentiable optimization problems and descent methods for asymmetric variational inequality problems. *Math. Program.* **53**, 99–110 (1992)
- [17] Rockafellar, R.T., Uryasev, S.: Optimization of conditional value-at-risk. *J. Risk* **2**, 21–41 (2000)
- [18] Tibshirani, R., Saunders, M., Rosset, S., Zhu, Ji: Sparsity and smoothness via the fused lasso. *J. R. Stat. Soc. B* **67**, 91–108 (2005)
- [19] Ralph, D., Xu, H.: Convergence of stationary points of sample average two stage stochastic programs: a generalized equation approach. *Math. Oper. Res.* **36**, 568–592 (2011)
- [20] Chen, X., Pong, T.K., Wets, R.J.-B.: Two-stage stochastic variational inequalities: an ERM-solution procedure. *Math. Program.* **165**, 71–111 (2017)
- [21] Luo, Z., Pang, J.S., Ralph, D.: *Mathematical Programs with Equilibrium Constraints*. Cambridge University Press, Cambridge (1996)
- [22] Rockafellar, R.T., Wets, R.J.-B.: Stochastic variational inequalities: single-stage to multistage. *Math. Program.* **165**, 331–360 (2017)
- [23] Rockafellar, R.T., Sun, J.: Solving monotone stochastic variational inequalities and complementarity problems by progressive hedging. *Math. Program.* **174**, 453–471 (2018)
- [24] Rockafellar R. T., Sun, J.: Solving lagrangian variational inequalities with applications to stochastic programming, *Manuscripts* (2018)
- [25] Van Slyke, R., Wets, R.: L-shaped linear programs with application to optimal control and stochastic programming. *SIAM J. Appl. Math.* **17**, 638–663 (1969)
- [26] Rockafellar, R.T., Wets, R.: *Variational Analysis*. Springer, Berlin (1998)
- [27] Douglas, J., Rachford, H.: On the numerical solution of heat conduction problems in two or three space variables. *Trans. Am. Math. Soc.* **82**, 421–439 (1956)
- [28] Li, G., Pong, T.K.: Douglas-Rachford splitting for nonconvex optimization with application to non-convex feasibility problems. *Math. Program.* **159**, 371–401 (2016)
- [29] Rockafellar, R.T., Wets, R.J.-B.: Scenarios and policy aggregation in optimization under uncertainty. *Math. Oper. Res.* **16**, 235–290 (1991)
- [30] Chen, X., Sun, H., Xu, H.: Discrete approximation of two-stage stochastic and distributionally robust linear complementarity problems. *Math. Program.* **177**, 255–289 (2019)

- [31] Chen, X., Shapiro, A., Sun, H.: Convergence analysis of sample average approximation of two-stage stochastic generalized equations. *SIAM J. Optim.* **29**, 135–161 (2019)
- [32] Clarke, F.H.: *Optimization and Nonsmooth Analysis*. Wiley, New York (1983)
- [33] Shapiro, A., Dentcheva, D., Ruszczyński, A.: *Lectures on Stochastic Programming: Modeling and Theory*. MPS/SIAM Series on Optimization, vol. 9. Society for Industrial and Applied Mathematics, Philadelphia (2009)
- [34] Shapiro, A., Xu, H.: Stochastic mathematical programs with equilibrium constraints, modelling and sample average approximation. *Optimization* **57**, 395–418 (2008)
- [35] Xu, H.: Uniform exponential convergence of sample average random functions under general sampling with applications in stochastic programming. *J. Math. Anal. Appl.* **368**, 692–710 (2010)
- [36] Izmailov, A.F.: Strongly regular nonsmooth generalized equations. *Math. Program.* **147**, 581–590 (2014)
- [37] Genc, T.S., Reynolds, S.S., Sen, S.: Dynamic oligopolistic games under uncertainty: a stochastic programming approach. *J. Econ. Dyn. Control* **31**, 55–80 (2007)
- [38] Ehrenmann, A., Smeers, Y.: Generation capacity expansion in a risky environment: a stochastic equilibrium analysis. *Oper. Res.* **59**, 1332–1346 (2011)
- [39] Gürkan, G., Ozdemir, O., Smeers, Y.: Generation capacity investments in electricity markets: perfect competition. *CentER Discussion Paper, Tilburg: Econometrics*, vol. 045 (2013)
- [40] Gürkan, G., Pang, J.S.: Approximations of Nash equilibria. *Math. Program.* **117**, 223–253 (2009)
- [41] Kannan, A., Shanbhag, U.V., Kim, H.M.: Addressing supply-side risk in uncertain power markets: stochastic Nash models, scalable algorithms and error analysis. *Optim. Methods Softw.* **28**, 1095–1138 (2013)
- [42] Kannan, A., Shanbhag, U.V., Kim, H.M.: Strategic behavior in power markets under uncertainty. *Energy Syst.* **2**, 115–141 (2011)
- [43] Yao, J., Adler, I., Oren, S.S.: Modeling and computing two-settlement oligopolistic equilibrium in a congested electricity network. *Oper. Res.* **56**, 34–47 (2008)
- [44] Gabriel, S.A., Fuller, J.D.: A Benders decomposition method for solving stochastic complementarity problems with an application in energy. *Comput. Econ.* **35**, 301–329 (2010)
- [45] Schiro, D.A., Hobbs, B.F., Pang, J.S.: Perfectly competitive capacity expansion games with risk-averse participants. *Comput. Optim. Appl.* **65**, 511–539 (2015)
- [46] Jiang, J., Shi, Y., Wang, X., Chen, X.: Regularized two-stage stochastic variational inequalities for Cournot–Nash equilibrium under uncertainty. *J. Comp. Math.* [arXiv:1907.07317](https://arxiv.org/abs/1907.07317) (2019)
- [47] Ferris, M., Pang, J.S.: Engineering and economic applications of complementarity problems. *SIAM Rev.* **39**, 669–713 (1997)
- [48] Pang, J.S., Sen, S., Shanbhag, U.V.: Two-stage non-cooperative games with risk-averse players. *Math. Program.* **165**, 119–147 (2017)