SOLVING TWO-STAGE STOCHASTIC VARIATIONAL INEQUALITIES BY A HYBRID PROJECTION SEMISMOOTH NEWTON ALGORITHM*

XIAOZHOU WANG† AND XIAOJUN CHEN†

Abstract. A hybrid projection semismooth Newton algorithm (PSNA) is developed for solving two-stage stochastic variational inequalities, which is globally and superlinearly convergent under suitable assumptions. PSNA is a hybrid algorithm of the semismooth Newton algorithm and extragradient algorithm. At each step of PSNA, the second stage problem is split into a number of small variational inequality problems and solved in parallel for a fixed first stage decision iterate. The projection algorithm and semismooth Newton algorithm are used to find a new first stage decision iterate. Numerical results for large-scale nonmonotone two-stage stochastic variational inequalities and applications in traffic assignments show the efficiency of PSNA.

Key words. stochastic variational inequalities, semismooth Newton, extragradient algorithm, global convergence, superlinear convergence rate

AMS subject classifications. 90C15, 90C33

1. Introduction.

Let $(\Xi, \mathcal{A}, P)$ be a probability space induced by a random vector ξ with the support set $\Xi \subseteq \mathbb{R}^d$. Let $\mathcal{Y}$ be the space consisting of $\mathcal{A}$-measurable functions from Ξ to $\mathbb{R}^m$. We are interested in developing a globally and superlinearly convergent algorithm for computing a pair $(x, y(\cdot)) \in \mathbb{R}^n \times \mathcal{Y}$ that solves the following two-stage stochastic variational inequality (SVI) [2]

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

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

where

- $G : \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^d \to \mathbb{R}^n$ is a vector-valued function, Lipschitz continuous with respect to (x, y) for a.e. $\xi \in \Xi$ with Lipschitz constant $L_G(\xi)$, 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 function, continuously differentiable with respect to (x, y) for a.e. $\xi \in \Xi$, and $\mathcal{A}$-measureable with respect to ξ;
- $\mathbb{E}[\cdot]$ denotes the expected value over Ξ, $D \subseteq \mathbb{R}^n$ is a nonempty closed convex set, $C(\xi) \subseteq \mathbb{R}^m$ is a polyhedral set for a.e. $\xi \in \Xi$, $\mathcal{N}_D(x)$ 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.

In a solution pair $(x, y(\cdot)) \in \mathbb{R}^n \times \mathcal{Y}$ of (1.1)-(1.2), x is the first stage decision variable independent of ξ and $y(\cdot)$ is the second stage decision variable. The two-stage SVI characterizes the first-order optimality condition of the two-stage stochastic programming [2] and models some equilibrium problems under uncertain environments. The research for the two-stage SVI has received much attention; see [4, 5, 25, 28] for references.

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

*Submitted to the editors DATE.

Funding: This work was supported by Hong Kong Research Grants Council grant PolyU15300021.

†Department of Applied Mathematics, The Hong Kong Polytechnic University, Kowloon, Hong Kong (xzhou.wang@connect.polyu.hk, maxjchen@polyu.edu.hk).

stochastic linear complementarity problem (SLCP) as follows:

$$0 \le x \perp Ax + \mathbb{E}[B(\xi)y(\xi)] + q_1 \ge 0, \tag{1.3}$$

$$0 \le y(\xi) \perp N(\xi)x + M(\xi)y(\xi) + q_2(\xi) \ge 0, \quad \text{for a.e. } \xi \in \Xi, \tag{1.4}$$

where $A \in \mathbb{R}^{n\times n}$, $q_1 \in \mathbb{R}^n$, $B : \mathbb{R}^d \to \mathbb{R}^{n\times m}$, $N : \mathbb{R}^d \to \mathbb{R}^{m\times n}$, $M : \mathbb{R}^d \to \mathbb{R}^{m\times m}$, $q_2 : \mathbb{R}^d \to \mathbb{R}^m$. In [5], the existence and uniqueness of a solution of the two-stage SLCP were established under the strong monotonicity assumption. In addition, a new discretization scheme was proposed and a distributionally robust two-stage SLCP was studied.

Numerically, we solve the sample approximation discretization problem of (1.1)-(1.2). More specifically, given a sample set $\Xi_\nu = \{\xi_1, \dots, \xi_\nu\}$ of the random vector ξ, its discrete approximation problem has the following form

$$-\sum_{\ell=1}^{\nu} p_\ell G(x, y(\xi_\ell), \xi_\ell) \in \mathcal{N}_D(x), \tag{1.5}$$

$$-F(x, y(\xi_\ell), \xi_\ell) \in \mathcal{N}_{C(\xi_\ell)}(y(\xi_\ell)), \quad \ell = 1, \dots, \nu, \tag{1.6}$$

where $p_\ell > 0$ for $\ell = 1, \dots, \nu$ and $\sum_{\ell=1}^{\nu} p_\ell = 1$. If the sample set is independent identically distributed (i.i.d.), then (1.5)-(1.6) is called a sample average approximation (SAA) discretization problem of (1.1)-(1.2). See [2, 4, 5] for the convergence analysis of the solution of the SAA discretization problem to that of the two-stage SVI (1.1)-(1.2). The dimension of variables in problem (1.5)-(1.6) is $n + m\nu$. In practice, the sample size ν is very large and thus (1.5)-(1.6) is a large-scale problem. Most deterministic VI solvers [3, 9, 12, 15, 18, 19, 20, 26] encounter difficulties in handling such large-scale problems. Hence, it is necessary to develop efficient algorithms for solving (1.5)-(1.6).

The progressive hedging algorithm (PHA) was first proposed by Rockafellar and Wets [23] to solve multi-stage stochastic optimization problems. Recently, it was extended to solve the monotone multi-stage SVI by Rockafellar and Sun with finite samples [22]. PHA decomposes the original large-scale problem into a sequence of independent small sample-based subproblems and solves them in parallel. Theoretically, PHA is globally convergent for the monotone multi-stage SVI. However, only 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 [28] to solve the elicited monotone (not necessarily monotone) two-stage SVI. But it is difficult to verify the elicited monotonicity of the problem, and the convergence rate is still linear. To the best of our knowledge, globally and superlinearly convergent algorithms have not been studied for solving the two-stage SVI.

In this paper, we propose a globally and superlinearly convergent projection semismooth Newton algorithm (PSNA) for solving (1.5)-(1.6), which is a hybrid algorithm of the semismooth Newton algorithm and extragradient algorithm. We assume that (1.5)-(1.6) has relatively complete recourse [4]; that is, for any $x \in D$ and $\xi \in \Xi_\nu$, the second stage problem (1.6) has at least one solution. Let $\mathcal{S}(x, \xi)$ be the solution set of the second stage problem (1.6) for a given $(x, \xi) \in D \times \Xi_\nu$. Then problem (1.5)-(1.6) can be equivalently written as

$$-\sum_{\ell=1}^{\nu} p_\ell G(x, y(\xi_\ell), \xi_\ell) \in \mathcal{N}_D(x), \quad y(\xi_\ell) \in \mathcal{S}(x, \xi_\ell), \quad \ell = 1, \dots, \nu. \tag{1.7}$$

From an iterate x^k, PSNA finds $y^k(\xi_\ell) \in \mathcal{S}(x^k, \xi_\ell), \ell = 1, \ldots, \nu$ in parallel, and then finds x^{k+1} by using the linear Newton approximation scheme with the projection algorithm for the variational inequality (VI) in (1.7).

In convergence analysis, we define a solution function $\hat{y} : D \times \Xi \to \mathbb{R}^m$ by selecting a vector $\hat{y}(x, \xi_\ell) \in \mathcal{S}(x, \xi_\ell)$ for any $x \in D$ and $\xi_\ell \in \Xi_\nu$, and two functions $\hat{G} : D \times \Xi_\nu \to \mathbb{R}^n$ and $H : D \to \mathbb{R}^n$ with

$$\hat{G}(x, \xi_\ell) = G(x, \hat{y}(x, \xi_\ell), \xi_\ell) \quad \text{and} \quad H(x) = \sum_{\ell=1}^{\nu} p_\ell \hat{G}(x, \xi_\ell). \tag{1.8}$$

It is easy to see that if x^* is a solution of the VI

$$-H(x) \in \mathcal{N}_D(x), \tag{1.9}$$

then $(x^*, \hat{y}(x^*, \xi_1), \ldots, \hat{y}(x^*, \xi_\nu))$ is a solution of (1.5)-(1.6).

The main contribution of this paper is the development of a globally and superlinearly convergent algorithm called PSNA for solving large-scale two-stage SVI (1.5)-(1.6). Convergence analysis and numerical experiments with over 10^7 variables show the effectiveness and efficiency of the proposed PSNA. To guarantee the global convergence of PSNA, we provide sufficient conditions for the function H being Lipschitz continuous and monotone. Moreover, we show that H is semismooth under these conditions, which ensures the superlinear convergence of PSNA. It is worth noting that if the two-stage SVI (1.5)-(1.6) is monotone, then H is monotone, but conversely it is not true. Hence the conditions for global convergence of PSNA are weaker than the conditions for global convergence of PHA [22]. Comparing PSNA and PHA regarding convergence rate, PSNA has the superlinear convergence rate under proper assumptions (see Theorems 3.3 and 4.5 and Corollary 4.6), while PHA has linear convergence rate for solving the affine monotone SVI [22, Theorem 2]. Moreover, preliminary numerical results show that PSNA can find a solution of (1.5)-(1.6) using much less CPU time than PHA.

The paper is organized as follows. In section 2, we investigate the Lipschitz continuity, semismoothness, linear Newton approximation scheme and monotonicity of the functions in the two-stage SVI (1.5)-(1.6). In section 3, we propose PSNA and give the convergence analysis. In section 4, PSNA is applied to solve a special class of (1.5)-(1.6), where the VI in the second stage is a linear complementarity problem (LCP) and in the first stage $\sum_{\ell=1}^{\nu} p_\ell G(x, y(\xi_\ell), \xi_\ell) = A(x) + \sum_{\ell=1}^{\nu} p_\ell B(\xi_\ell) y(\xi_\ell)$. In section 5, we conduct numerical experiments for large-scale nonmonotone two-stage SVI and applications in traffic assignments to show the efficiency of PSNA. Section 6 is devoted to the conclusions.

We use the following notation and terminology through out the paper. $\|\cdot\|$ represents the Euclidean norm. $\mathbb{R}^n_+$ is the nonnegative orthant of $\mathbb{R}^n$. $\Pi_D(x) = \arg\min_{y \in D} \|x - y\|^2$ denotes the projection of x onto the closed convex set D. $\mathcal{B}(x)$ represents an open neighborhood of x. $[m]$ denotes the set $\{1, \ldots, m\}$ for any positive integer m. If $K : \mathbb{R}^k \to \mathbb{R}^s$ is differentiable, $\nabla K(x)$ denotes its Jacobian at x and $K'(x; h)$ is the directional derivative at x along the direction h. A set-valued mapping $\Psi : \mathbb{R}^k \rightrightarrows \mathbb{R}^s$ is said to be outer semicontinuous (osc) at $\bar{x}$ relative to a set $X \subseteq \mathbb{R}^k$ if $\limsup_{x \to_X \bar{x}} \Psi(x) \subseteq \Psi(\bar{x})$ where $\limsup_{x \to_X \bar{x}} \Psi(x) := \{v \in \mathbb{R}^s : \exists x^k \to \bar{x}, \exists v^k \to v \text{ with } x^k \in X,\ v^k \in \Psi(x^k)\ \}$, see [24, Defintion 5.4]. A matrix M is called a P-matrix if all its principal minors are positive. A matrix M is called a Z-matrix if all its off-diagonal entries are non-positive. $M \succeq 0$ means that matrix M is positive semidefinite. We use VI(D, K) and LCP(q, M) to denote the problems

$-K(x) \in \mathcal{N}_D(x)$ and $0 \le x \perp Mx + q \ge 0$, respectively. SOL$(q, M)$ is the solution set of LCP(q, M). e_n denotes the n-dimensional vector with all components being 1.

2. Properties of problem (1.5)-(1.6).

In this section, we study the Lipschitz continuity, semismoothness, linear Newton approximation scheme and monotonicity of the functions in (1.5)-(1.6) and the function in the single-stage SVI with a finite support set Ξ_ν for the convergence analysis of 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 [10] 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. Function K is said to be semismooth at x if K is locally Lipschitz continuous around x and the limit

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

exists for any $h \in \mathbb{R}^k$; see [12, 20, 27] for details.

Throughout the paper, $\mathcal{D} \subseteq \mathbb{R}^n$ denotes an open set containing the set D. It is said that (1.5)-(1.6) has relatively complete recourse on $\mathcal{D}$ if for any $x \in \mathcal{D}$ and $\xi \in \Xi_\nu$, the second stage problem (1.6) has at least one solution.

We make the following basic assumption for Lipschitz continuous selection of $\mathcal{S}(x, \xi)$. For continuous selection of $\mathcal{S}(x, \xi)$, see [24, Defintion 5.58 (Michael representations)].

ASSUMPTION 2.1. *The two-stage SVI (1.5)-(1.6) has relatively complete recourse on $\mathcal{D}$; i.e., $\mathcal{S}(x, \xi)$ is nonempty for any $x \in \mathcal{D}$ and $\xi \in \Xi_\nu$. Moreover, for any $\xi \in \Xi_\nu$, there exists a Lipschitz continuous selection $\hat{y}(x, \xi) \in \mathcal{S}(x, \xi)$, i.e.*

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

where $L_{\hat{y}}(\xi) > 0$ is the Lipschitz constant.

Some sufficient conditions for Assumption 2.1 can be found in [12]. For example, the condition that for any $x \in \mathcal{D}$ and $\xi \in \Xi_\nu$, $F(x, \cdot, \xi)$ is strongly monotone on $C(\xi)$ in the sense that there is $\rho_\xi > 0$, independent of x, such that for any $u, v \in C(\xi)$,

$$(u - v)^T (F(x, u, \xi) - F(x, v, \xi)) \ge \rho_\xi \|u - v\|^2$$

holds. Other conditions for ensuring Assumption 2.1 will be discussed in section 4.

The following proposition studies the Lipschitz continuity of H and the solvability of (1.5)-(1.6).

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

(i) The function H is Lipschitz continuous on $\mathcal{D}$ with a Lipschitz constant $L_H = \sum_{\ell=1}^{\nu} p_\ell (L_G(\xi_\ell) L_{\hat{y}}(\xi_\ell) + L_G(\xi_\ell))$.

(ii) If D is bounded, then (1.5)-(1.6) is solvable.

(iii) If D is a box and $\mathcal{S}(x, \xi)$ is a singleton for any $x \in D$ and $\xi \in \Xi_\nu$, and H is a uniformly P function, then (1.5)-(1.6) has a unique solution.

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

$$\|H(x) - H(x')\| = \left\| \sum_{\ell=1}^{\nu} p_\ell (\hat{G}(x,\xi_\ell) - \hat{G}(x',\xi_\ell)) \right\|$$

$$\leq \sum_{\ell=1}^{\nu} p_\ell \|G(x, \hat{y}(x,\xi),\xi_\ell) - G(x', \hat{y}(x',\xi_\ell),\xi_\ell)\|$$

$$\leq \sum_{\ell=1}^{\nu} p_\ell (L_G(\xi_\ell) L_{\hat{y}}(\xi_\ell) + L_G(\xi_\ell)) \|x - x'\| = L_H \|x - x'\|.$$

(ii) Since D is bounded and H is Lipschitz continuous, from [12, Corollary 2.2.5], we immediately know that (1.9) is solvable, which implies that (1.5)-(1.6) is solvable.

(iii) From [12, Proposition 3.5.10], problem (1.9) has a unique solution x^*. From the assumption that $\mathcal{S}(x,\xi)$ is singleton for any $x \in D$ and $\xi \in \Xi_\nu$, we find that $(x^*, \hat{y}(x^*,\xi_1), \ldots, \hat{y}(x^*,\xi_\nu))$ is the unique solution of (1.5)-(1.6). ∎

Next, we will discuss the semismoothness and the linear Newton approximation scheme of H.

By the Lipschitz continuity of $G(\cdot,\cdot,\xi)$ and $\hat{y}(\cdot,\xi)$, $\hat{G}(\cdot,\xi)$ is Lipschitz continuous. The set-valued mapping $\mathcal{H} : \mathcal{D} \rightrightarrows \mathbb{R}^{n\times n}$ defined by

$$\mathcal{H}(x) = \mathbb{E}[\partial \hat{G}(x,\xi)] = \left\{ \sum_{\ell=1}^{\nu} p_\ell V(x,\xi_\ell) \,:\, V(x,\xi_\ell) \in \partial \hat{G}(x,\xi_\ell) \right\}$$

is Aumann's (set-valued) expectation of $\partial \hat{G}(x,\xi)$ [1].

The following proposition provides some properties of $\mathcal{H}$.

PROPOSITION 2.2. *Under Assumption 2.1, $\mathcal{H}(x)$ is nonempty, convex and compact at any $x \in \mathcal{D}$. Moreover, $\mathcal{H}$ is osc and closed at any $x \in \mathcal{D}$ relative to $\mathcal{D}$; that is, if $x^k \to_{\mathcal{D}} x$, $W^k \in \mathcal{H}(x^k)$ and $W^k \to W$, then $W \in \mathcal{H}(x)$.*

Proof. From Assumption 2.1, for any $\xi \in \Xi_\nu$, the generalized Jacobian $\partial \hat{G}(\cdot,\xi)$ of $\hat{G}(\cdot,\xi)$ is nonempty, convex, compact and osc at any $x \in \mathcal{D}$ relative to $\mathcal{D}$. By the definition of $\mathcal{H}$, we have the properties in this proposition. ∎

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

DEFINITION 2.3 ([12], Definition 7.5.13). *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, is osc at $\bar{x}$, and for any $h \to 0$, $W \in \Psi(\bar{x}+h)$*

$$\|K(\bar{x}+h) - K(\bar{x}) - Wh\| = o(\|h\|). \tag{2.1}$$

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

By Definition 2.3, ∂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 and it holds that $\partial H(x) \subseteq \sum_{\ell=1}^{\nu} p_\ell \partial \hat{G}(x,\xi_\ell)$ in general by [10, Corollary 2]. As we will see in Section 4, elements of $\partial \hat{G}(x,\xi_\ell)$ can be easily calculated for the two-stage semi-linear SVI, which allows us to obtain elements of $\mathcal{H}(x)$. Hence from a practical point of view, it is more appropriate to use $\mathcal{H}$ in the study of the linear Newton approximation scheme of H.

To establish that $\mathcal{H}$ is a linear Newton approximation scheme of H, the semismoothness of $\hat{G}(\cdot,\xi)$ is needed. 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 $\hat{y}(\cdot,\xi)$. To this end, we introduce the *Strong Regularity Condition* (SRC) proposed by Robinson [21]. Facchinei and Pang also thoroughly discussed this property in the monograph [12].

Without loss of generality, for $\xi \in \Xi_\nu$ let

$$C(\xi) := \{y \in \mathbb{R}^m : T(\xi)y \le b(\xi)\},$$

with $T : \mathbb{R}^d \to \mathbb{R}^{s\times m}$ and $b : \mathbb{R}^d \to \mathbb{R}^s$. For any given $x \in \mathcal{D}$ and $\xi \in \Xi_\nu$, define the critical cone of the pair $(C(\xi), F(x,\cdot,\xi))$ at $\hat{y}(x,\xi) \in C(\xi)$ as follows

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

where $\bar{T}(\xi)$ is a sub-matrix of $T(\xi)$ consisting of rows of $T(\xi)$ satisfying $\bar{T}(\xi)\hat{y}(x,\xi) = \bar{b}(\xi)$ with $\bar{b}(\xi)$ being the corresponding sub-vector of $b(\xi)$.

We make the following SRC assumption for the second stage problem. In the case of the VI with a polyhedral set, by [12, Theorem 5.3.17(e)], the SRC condition is equivalently defined as follows.

ASSUMPTION 2.2. *For any $\xi \in \Xi_\nu$, the SRC holds at $\hat{y}(x,\xi)$ for the* VI$(C(\xi), F(x,\cdot,\xi))$ *for any $x \in \mathcal{D}$; that is, for any $x \in \mathcal{D}$, the following affine VI admits a unique solution for each $q \in \mathbb{R}^m$*

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

By the SRC assumption, it is clear that Assumption 2.2 holds if $F(x,\cdot,\xi)$ is strongly monotone on $C(\xi)$ for any $x \in \mathcal{D}$ and $\xi \in \Xi_\nu$. In the case that $C(\xi) = \mathbb{R}^m_+$ for any $\xi \in \Xi_\nu$, a sufficient condition for guaranteeing Assumption 2.2 is that $F(x,\cdot,\xi)$ is a uniformly P function for any $x \in \mathcal{D}$ and $\xi \in \Xi_\nu$.

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

PROPOSITION 2.4. *Let $D \times C(\xi)$ be contained in an open set $\mathcal{D} \times \bar{\mathcal{C}}(\xi)$ for any $\xi \in \Xi_\nu$. Suppose that Assumptions 2.1-2.2 holds, and that for any fixed $\xi \in \Xi_\nu$, $G(\cdot,\cdot,\xi)$ is semismooth at $(x,\hat{y}(x,\xi)) \in \mathcal{D} \times \bar{\mathcal{C}}(\xi)$. Then we have the following assertions.*

(i) H is semismooth at $x \in \mathcal{D}$.

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

Proof. (i) With Assumption 2.2, by [12, Theorem 5.4.6], we know that for any fixed $\xi \in \Xi_\nu$, $\hat{y}(\cdot,\xi)$ is a piecewise smooth function on $\mathcal{D}$, and hence it is semismooth on $\mathcal{D}$. By [12, Proposition 7.4.4], the composition of semismooth functions is also semismooth. Then, we deduce that $\hat{G}(\cdot,\xi)$ is semismooth at $x \in \mathcal{D}$ for any fixed $\xi \in \Xi_\nu$. Since the sum of finite semismooth functions is also semismooth [20], we know that H is semismooth at $x \in \mathcal{D}$.

(ii) By Proposition 2.2, $\mathcal{H}$ has nonempty compact images and is osc at any $x \in \mathcal{D}$ relative to $\mathcal{D}$. For any $h \to 0$, $W \in \mathcal{H}(x+h)$, let $V(\xi_\ell) \in \partial\hat{G}(x+h,\xi_\ell)$ such that

$W = \sum_{\ell=1}^{\nu} p_\ell V(\xi_\ell)$. It follows that

$$\begin{aligned}
&\lim_{\substack{h\to 0,\\ W\in\partial\mathcal{H}(x+h)}} \frac{\|H(x+h) - Wh - H(x)\|}{\|h\|}\\
&= \lim_{\substack{h\to 0,\\ V(\xi_\ell)\in\partial\hat{G}(x+h,\xi_\ell)}} \frac{\|\sum_{\ell=1}^{\nu} p_\ell(\hat{G}(x+h,\xi_\ell) - V(\xi_\ell)h - \hat{G}(x,\xi_\ell))\|}{\|h\|}\\
&\le \lim_{\substack{h\to 0,\\ V(\xi_\ell)\in\partial\hat{G}(x+h,\xi_\ell)}} \frac{\sum_{\ell=1}^{\nu} p_\ell\|\hat{G}(x+h,\xi) - V(\xi)h - \hat{G}(x,\xi)\|}{\|h\|} = 0,
\end{aligned}$$

where the last equality is due to the semismoothness of $\hat{G}(\cdot,\xi)$ at x for any $\xi\in\Xi_\nu$. Hence $\mathcal{H}$ is a linear Newton approximation scheme of H at $x\in\mathcal{D}$. □

Next, we study the monotonicity of H. The function H is said to be monotone on $\mathcal{D}$ if for any $u, v\in\mathcal{D}$, the following inequality holds

$$(H(u) - H(v))^T(u - v) \ge 0. \tag{2.2}$$

Using the definition of the monotonicity of the two-stage SVI in [22], we define the monotonicity of (1.5)-(1.6). Define a mapping $\mathcal{T}:\mathbb{R}^n\times\mathcal{Y}_\nu\to\mathbb{R}^n\times\mathcal{Y}_\nu$ with $\mathcal{Y}_\nu$ being the linear space consisting of all mappings from Ξ_ν to $\mathbb{R}^m$ 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 $\mathcal{D}\times\bar{\mathcal{C}}(\cdot)$ if for any $(x,y(\cdot))$, $(x',y'(\cdot))\in\mathcal{D}\times\bar{\mathcal{C}}(\cdot)$[1], it holds [22] that

$$\begin{aligned}
&\left\langle \mathcal{T}(x,y(\cdot)) - \mathcal{T}(x',y'(\cdot)), \begin{pmatrix} x - x' \\ y(\cdot) - y'(\cdot)\end{pmatrix}\right\rangle\\
&= \sum_{\ell=1}^{\nu} p_\ell[(x - x')^T(G(x,y(\xi_\ell),\xi_\ell) - G(x',y'(\xi_\ell),\xi_\ell))\\
&\quad + (y(\xi_\ell) - y'(\xi_\ell))^T(F(x,y(\xi_\ell),\xi_\ell) - F(x',y'(\xi_\ell),\xi_\ell))] \ge 0.
\end{aligned}$$

The SVI (1.5)-(1.6) is said to be monotone if $\mathcal{T}$ is monotone on $D\times C(\cdot)$.

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

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

ASSUMPTION 2.3. *The function $\hat{G}(\cdot,\xi)$ defined in (1.8) is monotone on $\mathcal{D}$ for each fixed $\xi\in\Xi_\nu$.*

The following proposition gives sufficient conditions for Assumption 2.3.

PROPOSITION 2.5. *Let $H(x) = \sum_{\ell=1}^{\nu} p_\ell\hat{G}(x,\xi_\ell)$, where $\hat{G}(x,\xi) = G(x,\hat{y}(x,\xi),\xi)$ with $\hat{y}(x,\xi)$ being a Lipschitz continuous selection from $\mathcal{S}(x,\xi)$. Then Assumption 2.3 holds and H is monotone on $\mathcal{D}$, under Assumption 2.2 and the following two conditions:*

(i) For any $\xi\in\Xi_\nu$, $\Theta(\cdot,\cdot,\xi)$ is monotone on $\mathcal{D}\times\bar{\mathcal{C}}(\xi)$;

(ii) For any $\bar{x}\in\mathcal{D}$ and $\xi\in\Xi_\nu$ with $\bar{y} := \hat{y}(\bar{x},\xi)$, $\nabla_y F(\bar{x},\bar{y},\xi)v$ is contained in the column space of $\nabla_x F(\bar{x},\bar{y},\xi)$ for any $v\in\mathcal{C}_{\bar{x}}(\bar{y};C(\xi),F)$.

[1] $(x,y(\cdot))\in\mathcal{D}\times\bar{\mathcal{C}}(\cdot)$ if $(x,y(\xi))\in\mathcal{D}\times\bar{\mathcal{C}}(\xi)$ for any $\xi\in\Xi_\nu$.

Proof. It suffices to show that every element of $\partial\hat{G}_x(x,\xi)$ is positive semidefinite for any $x \in \mathcal{D}$ and $\xi \in \Xi_\nu$ by [16, Proposition 2.3]. Under Condition (i), for any $(x, y(\xi)) \in \mathcal{D} \times \bar{\mathcal{C}}(\xi)$ and $\xi \in \Xi_\nu$, it holds

$$\begin{pmatrix} V_x(x,y(\xi),\xi) & V_y(x,y(\xi),\xi) \\ \nabla_x F(x,y(\xi),\xi) & \nabla_y F(x,y(\xi),\xi) \end{pmatrix} \succeq 0, \tag{2.3}$$

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

For any $\nabla_y F(x,y(\xi),\xi)$ with $\text{rank}(\nabla_y F(x,y(\xi),\xi)) = r \geq 1$, define the set

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

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

For any $u \in \mathbb{R}^n$, let $v = U_Z(x,y(\xi),\xi)u \in \mathbb{R}^m$. Then from (2.3), 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

$$V_x(x,y(\xi),\xi) + V_y(x,y(\xi),\xi)U_Z(x,y(\xi),\xi) \succeq 0. \tag{2.4}$$

Under Assumption 2.2, $\hat{y}(\cdot,\xi)$ is a semismooth function by Proposition 2.4. Let $\Omega_{\hat{y}(\cdot,\xi)}$ be the set of differentiable points of $\hat{y}(\cdot,\xi)$. Under Assumptions 2.2 and (ii), by [12, Corollary 5.4.14], we have that $\mathcal{C}_{\bar{x}}(\hat{y}; C(\xi), F)$ is a linear subspace for any $\bar{x} \in \Omega_{\hat{y}(\cdot,\xi)}$, i.e., $\mathcal{C}_{\bar{x}}(\hat{y}; C(\xi), F) = \mathcal{C}_{\bar{x}}(\hat{y}; C(\xi), F) \cap -\mathcal{C}_{\bar{x}}(\hat{y}; C(\xi), F)$. Therefore, by [17, Theorem 2.2], the Jacobian $\nabla_x \hat{y}(\bar{x},\xi)$ at any $\bar{x} \in \Omega_{\hat{y}(\cdot,\xi)}$ can be represented as

$$\nabla_x \hat{y}(\bar{x},\xi) = U_Z(\bar{x},\hat{y}(\bar{x},\xi),\xi), \ Z \in \hat{\mathcal{Z}}(\bar{x},\hat{y}(\bar{x},\xi),\xi), \tag{2.5}$$

where $\hat{\mathcal{Z}}(\bar{x},\hat{y}(\bar{x},\xi),\xi)$ is a set consisting of matrices in $\mathbb{R}^{m\times l}$ with l being the dimension of $\mathcal{C}_{\bar{x}}(\hat{y}; C(\xi), F)$, and each element $Z \in \hat{\mathcal{Z}}(\bar{x},\hat{y}(\bar{x},\xi),\xi)$ satisfies that Z^TZ and $Z^T\nabla_y F(\bar{x},\hat{y}(\bar{x},\xi),\xi)Z$ are nonsingular and $z \in \mathcal{C}_{\bar{x}}(\hat{y}; C(\xi), F)$ if and only if $z = Zv$ for some $v \in \mathbb{R}^l$. Under the SRC assumption, by [17, Lemma 2.1], we know that $\hat{\mathcal{Z}}(\bar{x},\hat{y}(\bar{x},\xi),\xi)$ is not empty, and it is clear that $\hat{\mathcal{Z}}(\bar{x},\hat{y}(\bar{x},\xi),\xi) \subseteq \mathcal{Z}(\bar{x},\hat{y}(\bar{x},\xi),\xi)$.

Let $\mathcal{B}(x) \subset \mathcal{D}$ be an open neighborhood of $x \in \mathcal{D}$. Since $\hat{G}(\cdot,\xi)$ and $\hat{y}(\cdot,\xi)$ are Lipschitz continuous, 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

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

where the third equality is due to (2.5) and the last inclusion is due to $\hat{\mathcal{Z}}(x, \hat{y}(x,\xi), \xi) \subseteq \mathcal{Z}(x, \hat{y}(x,\xi), \xi)$ and the outer semicontinuity of $\partial \hat{y}(\cdot, \xi)$. By (2.4), we know that for any $\xi \in \Xi_\nu$, all elements in $\partial_x \hat{G}(x,\xi)$ are positive semidefinite for any $x \in \mathcal{D}$, which implies the monotonicity of $\hat{G}(\cdot,\xi)$ on $\mathcal{D}$ for any $\xi \in \Xi_\nu$. Therefore, we conclude that H is monotone on $\mathcal{D}$. ◻

Remark 2.6. It is worth noting that the monotonicity of H does not imply the monotonicity of (1.5)-(1.6). For example, for any $x \in \mathcal{D}$, let

$$\|G(x, y(\xi), \xi) - G(x, y'(\xi), \xi)\| \le L(\xi)\|y(\xi) - y'(\xi)\|, \ \forall\ y(\xi), y'(\xi) \in C(\xi). \tag{2.6}$$

If for any $\xi \in \Xi_\nu$ and $y(\xi) \in C(\xi)$, $G(\cdot, y(\xi), \xi)$ is strongly monotone such that

$$(x - x')^T(G(x, y(\xi), \xi) - G(x', y(\xi), \xi)) \ge \sigma(\xi)\|x - x'\|^2, \quad \forall\ x, x' \in \mathcal{D}, \tag{2.7}$$

with $\sigma(\xi) := L(\xi)L_{\hat{y}}(\xi) > 0$, then by the Lipschitz continuity of $G(\cdot,\cdot,\xi)$ and $\hat{y}(\cdot,\xi)$ and (2.6) we have

$$\begin{aligned}
&(x - x')^T(H(x) - H(x')) \\
=&(x - x')^T(\sum_{\ell=1}^{\nu} p_\ell[G(x, \hat{y}(x,\xi_\ell), \xi_\ell) - G(x', \hat{y}(x,\xi_\ell), \xi_\ell) \\
&\qquad\qquad + G(x', \hat{y}(x,\xi_\ell), \xi_\ell) - G(x', \hat{y}(x',\xi_\ell), \xi_\ell)]) \\
\ge& \sum_{\ell=1}^{\nu} p_\ell \left(\sigma(\xi_\ell)\|x - x'\|^2 - \|x - x'\|\, \|G(x', \hat{y}(x,\xi_\ell), \xi_\ell) - G(x', \hat{y}(x',\xi_\ell), \xi_\ell)\|\right) \\
\ge& \sum_{\ell=1}^{\nu} p_\ell(\sigma(\xi_\ell) - L(\xi_\ell)L_{\hat{y}}(\xi_\ell))\|x - x'\|^2 \ge 0, \quad \forall\ x, x' \in \mathcal{D},
\end{aligned}$$

which implies the monotonicity of H on $\mathcal{D}$. However, the conditions (2.6)-(2.7) do not imply that (1.5)-(1.6) is monotone. Thus, the global convergence of PHA for solving (1.5)-(1.6) cannot be guaranteed under (2.6)-(2.7).

3. The hybrid projection semismooth Newton algorithm. In this section, we propose the hybrid projection semismooth Newton algorithm (PSNA), which combines the semismooth Newton algorithm with the extrgradient projection algorithm. The global convergence and superlinear convergence rate are established under suitable assumptions.

Define the residual function of (1.9) as

$$\hat{Q}(x) := x - \Pi_D(x - H(x)). \tag{3.1}$$

Proposition 1.5.8 in [12] claims that x^* solves (1.9) if and only if $\hat{Q}(x^*) = 0$. The function $\hat{Q}$ is Lipschitz continuous due to the Lipschitz continuity of H and the non-expansiveness of the projection operator. Let $L_{\hat{Q}}$ denote the Lipschitz constant of $\hat{Q}$.

We define a linear approximation of H and let the solution of the corresponding linear VI subproblem

$$-H(x^k) - (W^k + \epsilon_k I)(x - x^k) \in \mathcal{N}_D(x), \quad W^k \in \mathcal{H}(x^k), \tag{3.2}$$

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 (3.2) to be strongly monotone provided that W^k is positive semidefinite.

A main issue for Newton-type algorithms is that they are locally convergent in general. Since H is nonsmooth and an implicit function, the line search technique frequently used in Newton-type algorithms cannot be directly applied to our problem. Therefore, we turn to the extragradient projection algorithm to globalize the semismooth Newton iteration (3.2).

Define a projection operator

$$\tilde{\Pi}_{D,\alpha}(x) := \Pi_D[x - \alpha H(\pi(x))], \quad \text{with } \pi(x) := \Pi_D[x - \alpha H(x)], \tag{3.3}$$

where $\alpha > 0$ is the step size. Notice that (3.3) is called the extragradient algorithm for solving (1.9) in [12, Algorithm 12.1.9].

Under Assumptions 2.1 and 2.3, choosing $0 < \alpha < \frac{1}{L_H}$ with L_H being the Lipschitz constant of function H in Proposition 2.1, by [12, 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 [12, 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 (1.9). However, the convergence rate is linear. To achieve a superlinear convergence rate, a hybrid algorithm with the semismooth Newton algorithm (3.2) is proposed in Algorithm 2.1.

Algorithm 3.1. The Hybrid Projection Semismooth Newton Algorithm

Step 0: Choose an initial point $x^0 \in D$, $\eta \in (0,1)$, 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 $\hat{y}(x^k, \xi_\ell)$ that solve the second stage problem (1.6).

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

$$-H(x^k) - (W^k + \epsilon_k I)(x - x^k) \in \mathcal{N}_D(x). \tag{3.4}$$

If $\|\hat{Q}(\hat{x}^{k+1})\| \le \eta \|\hat{Q}(x^k)\|$, 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

$$x^{k,j+1} = \tilde{\Pi}_{D,\alpha}(x^{k,j}), \quad j = 0, 1, \ldots, \tag{3.5}$$

until $\|\hat{Q}(x^{k,j+1})\| \le \eta \|\hat{Q}(x^k)\|$ 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 2.3, any element of $\mathcal{H}(x)$ is positive semidefinite for any $x \in D$. Thus, subproblem (3.4) is strongly monotone for any $\epsilon_k > 0$, which has a unique solution and is easy to solve. In Step 3 of PSNA, the projection iteration (3.5) is well-defined and is equivalent to solving a strongly convex program.

Lemma 3.1. *Under Assumptions 2.1 and 2.3, for any x^k with $\|\hat{Q}(x^k)\| > 0$, Step 3 of PSNA is terminated in finite times, i.e., there is $j \ge 0$ such that $\|\hat{Q}(x^{k,j+1})\| \le \eta \|\hat{Q}(x^k)\|$.*

Proof. By [12, Theorem 12.1.11], we know that $\{x^{k,j}\}_{j=1}^{\infty}$ generated by (3.5) converges to a solution x^* of (1.9). By the Lipschitz continuity of $\hat{Q}$, we have

$$\|\hat{Q}(x^{k,j+1})\| = \|\hat{Q}(x^{k,j+1}) - \hat{Q}(x^*)\| \le L_{\hat{Q}} \|x^{k,j+1} - x^*\|.$$

Hence $\|\hat{Q}(x^{k,j+1})\| \to 0$ as $j \to \infty$, which implies that there exists j such that the assertion of the lemma holds. ∎

ASSUMPTION 3.1. *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 [12, Corollary 3.6.5(c)], Assumption 3.1 is satisfied if H is monotone and the solution set of (1.9) 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 3.1.

THEOREM 3.2. *Suppose that Assumptions 2.1, 2.3 and 3.1 hold. Let $\{x^k\}$ be an infinite sequence generated by PSNA. Then every accumulation point of $\{x^k\}$ is a solution of (1.9). In particular, if the Newton iteration is performed finite times, then $\{x^k\}$ converges to a solution of (1.9).*

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, this implies that there exists an integer $\bar{k} > 0$ such that for all $k \geq \bar{k}$ the projection iteration (3.5) is always executed. By [12, Theorem 12.1.11], it follows that $\{x^k\}$ converges to a solution of (1.9).

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 $\hat{Q}$, we deduce that every accumulation point of $\{x^k\}$ is a solution of (1.9). ∎

Next, we study the superlinear convergence rate of PSNA.

THEOREM 3.3. *Suppose that Assumptions 2.1-2.3 and 3.1 hold and x^* is an accumulation point of $\{x^k\}$ generated by PSNA. If $G(\cdot,\cdot,\xi)$ is semismooth at $(x^*, \hat{y}(x^*, \xi))$ for any $\xi \in \Xi_\nu$, D is a polyhedron, and all $W^* \in \mathcal{H}(x^*)$ are positive definite, then $\{x^k\}$ converges to x^* superlinearly.*

Proof. By Proposition 2.4, 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 3.2, x^* is a solution of (1.9), which implies $\hat{Q}(x^*) = 0$.

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 (3.4) has a unique solution, denoted by $\hat{x}^{k+1}$. Hence 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

$$
\begin{aligned}
& 0 \leq [H(x^k) + W^k_{\epsilon_k}(\hat{x}^{k+1} - x^k) - H(x^*)]^T(x^* - \hat{x}^{k+1}) \\
\Leftrightarrow\ & (\hat{x}^{k+1} - x^*)^T W^k_{\epsilon_k}(\hat{x}^{k+1} - x^*) \leq [H(x^k) - H(x^*) + W^k_{\epsilon_k}(x^* - x^k)]^T(x^* - \hat{x}^{k+1}) \\
\Rightarrow\ & \frac{1}{2}\lambda\|\hat{x}^{k+1} - x^*\|^2 \leq (\|H(x^k) - H(x^*) - W^k(x^k - x^*)\| + \epsilon_k\|x^k - x^*\|)\|\hat{x}^{k+1} - x^*\| \\
\Rightarrow\ & \|\hat{x}^{k+1} - x^*\| \leq o(\|x^k - x^*\|), \qquad (3.6)
\end{aligned}
$$

where the last inequality is due to the semismoothness of H at x^* and $\epsilon_k \to 0$.

Next, we will prove that for all k sufficiently large

$$\|\hat{Q}(\hat{x}^{k+1})\| = o(\|\hat{Q}(x^k)\|). \qquad (3.7)$$

By (3.6), 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 [12, Theorem 2.3.3], there exists a positive constant $c' > 0$ such that

$$\|x^k - x^*\| \leq c'\|\hat{Q}(x^k)\|. \qquad (3.8)$$

The last two inequalities imply that

$$\|\hat{x}^{k+1} - x^k\| \leq c'\|\hat{Q}(x^k)\|. \qquad (3.9)$$

(3.6) also implies that

$$\|\hat{x}^{k+1} - x^*\| \leq \varepsilon\|x^k - x^*\|, \qquad (3.10)$$

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 [12, Theorem 4.1.1]. Since $\hat{Q}$ is directionally differentiable at x^* and Lipschitz continuous, by [20], 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

$$
\begin{aligned}
\|\hat{Q}(\hat{x}^{k+1})\| &\leq \|\hat{Q}'(x^*; \hat{x}^{k+1} - x^*)\| + \varepsilon\|\hat{x}^{k+1} - x^*\| \\
&\leq (L_{\hat{Q}} + 2\varepsilon)\|\hat{x}^{k+1} - x^*\| \\
&\leq (L_{\hat{Q}} + 2\varepsilon)\varepsilon\|x^k - x^*\|. \qquad (3.11)
\end{aligned}
$$

From (3.9) and (3.10), 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

$$\|x^k - x^*\| \le \frac{c'}{1-\varepsilon}\|\hat{Q}(x^k)\|. \tag{3.12}$$

Combining (3.11) with (3.12), it holds that

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

Since ε can be arbitrarily small when k is sufficiently large, the last inequality implies (3.7). This means that $\hat{x}^{k+1}$ computed from the Newton iteration (3.4) is always accepted when x^k is sufficiently close to x^*. Then, $x^{k+1} = \hat{x}^{k+1}$. Therefore, (3.6) becomes

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

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

Remark 3.4. The assumption on semismoothness of the function $G(\cdot,\cdot,\xi)$ is standard for Newton-type algorithms. If $F(x,\cdot,\xi)$ is a uniformly P function for any $x \in \mathcal{D}$ and $\xi \in \Xi_\nu$, then $G(\cdot,\cdot,\xi)$ is semismooth. The assumption on the positive definiteness of the elements in $\mathcal{H}(x^*)$ holds if the $\Theta(\cdot,\cdot,\xi)$ is strongly monotone in an open neighborhood of $(x^*, \hat{y}(x^*,\xi))$ according to the proof in Proposition 2.5. Moreover, from Theorem 4.5, the assumption on $\mathcal{H}(x^*)$ can be weaken if D is a box. The assumption that D is a polyhedron in Theorem 3.3 can be extended to $D = \{x \in \mathbb{R}^n : g(x) \le 0\}$ where g is twice continuously differentiable and convex with the constant rank constraint qualification at x^*. From [12, Theorem 4.5.2], function $\hat{Q}$ is piecewise smooth around x^* in such case.

4. A two-stage semi-linear SVI.

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

$$-A(x) - \mathbb{E}[B(\xi)y(\xi)] \in \mathcal{N}_D(x), \tag{4.1}$$

$$0 \le y(\xi) \perp M(\xi)y(\xi) + N(\xi)x + q(\xi) \ge 0, \qquad \forall\, \xi \in \Xi_\nu, \tag{4.2}$$

where the function $A : \mathcal{D} \supset D \to \mathbb{R}^n$ is continuously differentiable 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 $q : \mathbb{R}^d \to \mathbb{R}^m$. If A is a linear function, then (4.1)-(4.2) is a two-stage linear SVI.

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

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

ASSUMPTION 4.1. *(i) M_ℓ is a P-matrix for all ℓ; or,*

(ii) M_ℓ is a Z-matrix for all ℓ, and (4.1)-(4.2) *has relatively complete recourse on $\mathcal{D}$.*

LEMMA 4.1. *For any fixed $x \in \mathcal{D}$ and $\xi_\ell \in \Xi_\nu$, the second stage problem* (4.2) *has a unique solution (or a unique least-element solution[2]) $\hat{y}(x,\xi_\ell)$ if Assumption 4.1 (i)*

[2] A solution y^* of the LCP(q, M) is called the least-element solution if $y^* \le y$ (componentwise) for any $y \in \mathrm{SOL}(q, M)$, and the least-element solution can be computed by solving a linear program [11].

(or Assumption 4.1 (ii)) holds, which reads

$$\hat{y}(x,\xi_\ell) = -U(x,\xi_\ell)(N_\ell x + q_\ell), \tag{4.3}$$

with $U(x,\xi_\ell) := (I - \Lambda(x,\xi_\ell) + \Lambda(x,\xi_\ell)M_\ell)^{-1}\Lambda(x,\xi_\ell)$, *where* $\Lambda(x,\xi_\ell)$ *is a diagonal matrix with*

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

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

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

and

$$-U(x,\xi_\ell)N_\ell \in \partial\hat{y}(x,\xi_\ell). \tag{4.4}$$

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 [11, Theorem 3.3.7]. When M_ℓ is a Z-matrix and LCP$(N_\ell x + q_\ell, M_\ell)$ is feasible for all $x \in \mathcal{D}$, the existence of the unique least-element solution follows from [11, Theorem 3.11.6]. The expression (4.3) follows from Lemma 2.1 and Theorem 2.2 in [8]. It is clear that $\hat{y}(\cdot,\xi_\ell)$ is piecewise affine from the expression (4.3). According to [12, 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 [8, Corollary 2.1] and [8, Theorem 2.3], respectively.

The generalized Jacobian (4.4) is due to [8, Theorem 3.1]. ∎

As in the last section, substituting the Lipschitz continuous selection $\hat{y}(x,\xi_\ell)$ into (4.1), we can define $\hat{G}(x,\xi_\ell) := A(x) + B_\ell\hat{y}(x,\xi_\ell)$. Thus the single-stage SVI formulation (1.9) is as follows

$$H(x) := \sum_{\ell=1}^{\nu} p_\ell \hat{G}(x,\xi_\ell) = A(x) + \mathbf{B}_\nu \hat{\mathbf{y}}_\nu(x), \tag{4.5}$$

where $\mathbf{B}_\nu = (p_1B_1,\ldots,p_\nu B_\nu) \in \mathbb{R}^{n\times\nu m}$, $\hat{\mathbf{y}}_\nu(x) = (\hat{y}^T(x,\xi_1),\ldots,\hat{y}^T(x,\xi_\nu))^T \in \mathbb{R}^{\nu m}$ with $\hat{y}(x,\xi_\ell) \in \text{SOL}(N_\ell x + q_\ell, M_\ell)$, $\ell = 1,\ldots,\nu$. Moreover, function H is Lipschitz continuous on $\mathcal{D}$ with Lipschitz constant

$$L_H = L_A + \bar{\sigma}, \quad \text{where } \bar{\sigma} = \sum_{\ell=1}^{\nu} p_\ell \|B_\ell\| L_\ell. \tag{4.6}$$

In addition, the corresponding residual function $\hat{Q}$ is Lipschitz continuous on $\mathcal{D}$. Under Assumption 4.1(i), as discussed in Proposition 2.1, (4.5) is an equivalent formulation to (4.1)-(4.2). Under Assumption 4.1(ii), if D is bounded, then (4.5) is solvable. Thus, if x^* solves (4.5), then $(x^*, \hat{y}(x^*,\xi_1),\ldots,\hat{y}(x^*,\xi_\nu))$ is a solution to (4.1)-(4.2).

Let

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

Similar to the Assumption 2.3 in the last section, the monotonicity of H is needed.

[3] A locally Lipschitz 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 [20].

ASSUMPTION 4.2. *Function $\hat{G}(\cdot,\xi)$ is monotone on $\mathcal{D}$ for any fixed $\xi \in \Xi_\nu$.*

PROPOSITION 4.2. *If Assumptions 4.1 holds and $\Theta(\cdot,\cdot,\xi)$ is monotone on $\mathcal{D}\times\mathbb{R}^m$ for any fixed $\xi \in \Xi_\nu$, then Assumption 4.2 holds and H is monotone on $\mathcal{D}$.*

Proof. Under Assumptions 4.1, using [8, Lemma 2.1], it is known that $\nabla\hat{y}(\bar{x},\xi_\ell) = -U(\bar{x},\xi_\ell)N_\ell$ at every differentiable point $\bar{x}$ of $\hat{y}(\cdot,\xi_\ell)$. Then, the assertion follows by a similar argument in Proposition 2.5. ∎

Remark 4.3. The monotonicity of H does not necessarily imply the monotonicity of the original problem (4.1)-(4.2). For instance, without the monotonicity assumption on $\Theta(\cdot,\cdot,\xi)$, H is monotone if A is strongly monotone on $\mathcal{D}$ such that

$$(4.7)\qquad (x-x')^T(A(x)-A(x')) \geq \tilde{\sigma}\|x-x'\|^2, \quad \forall\, x,x' \in \mathcal{D},$$

where $\tilde{\sigma} = \max\{\|B_\ell\|L_\ell : \ell = 1,\ldots,\nu\}$ with L_ℓ defined in Lemma 4.1. But condition (4.7) and Assumptions 4.1 do not imply the monotonicity of (4.1)-(4.2).

Note that the nonmonotone problems with monotone single-stage SVI reformulations are not limited to the case given in Remark 4.3. For instance, consider the example

$$-\begin{pmatrix} 1 & 0 \\ 0 & -\frac{1}{2} \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} - \mathbb{E}\left[\begin{pmatrix} 0 & 0 \\ -3\xi & \xi \end{pmatrix}\begin{pmatrix} y_1(\xi) \\ y_2(\xi) \end{pmatrix}\right] \in \mathcal{N}_{[0,1]^2}(x),$$
$$0 \leq \begin{pmatrix} y_1(\xi) \\ y_2(\xi) \end{pmatrix} \perp \begin{pmatrix} 0 & 3\xi \\ 0 & -\xi \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} + \begin{pmatrix} \xi^2 & -3\xi^2 \\ 0 & \xi^2 \end{pmatrix}\begin{pmatrix} y_1(\xi) \\ y_2(\xi) \end{pmatrix} \geq 0, \quad \forall\, \xi \in \Xi_\nu,$$

where each realization of ξ is uniformly distributed on $[1,2]$ with probability $1/\nu$ and $[0,1]^2 := [0,1]\times[0,1]$. This example is nonmonotone since the second stage problem is a P-matrix LCP with respect to y for each fixed ξ, and the first stage problem is not monotone on $[0,1]^2$ with respect to x. Substituting the unique solution function $\hat{y}(x,\xi) = (0,(1/\xi)x_2)^T$ of the second stage problem into the first stage problem, we get the single-stage SVI as

$$0 \in \begin{pmatrix} 1 & 0 \\ 0 & \frac{1}{2} \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} + \mathcal{N}_{[0,1]^2}(x),$$

which is a strongly monotone VI.

Let

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

By Proposition 2.4, the set-valued mapping $\mathcal{H}$ is a linear Newton approximation scheme of H. By Lemma 4.1, one particular element of $\mathcal{H}(x)$ can be calculated by

$$(4.8)\qquad \nabla A(x) - \mathbf{B}_\nu\mathbf{U}_\nu(x) \in \mathcal{H}(x),$$

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 Lemma 4.1.

By the same argument as in the proof of Theorems 3.2 and 3.3, we can prove the global convergence and superlinear convergence rate of PSNA for solving (4.1)-(4.2). 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 $\mathrm{mid}(l,u,x)_i$ is equal to l_i if $x_i < l_i$, u_i if $x_i > u_i$ and x_i if $l_i \leq x_i \leq u_i$.

LEMMA 4.4. *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 [20, Proposition 3.1], there exist a neighborhood $\mathcal{B}(x)$ of x and positive constant $\hat{\beta}$ such that for any $\bar{x} \in \mathcal{B}(x)$, all $\bar{W} \in \mathcal{H}(\bar{x})$ are nonsingular and $\|\bar{W}^{-1}\| \le \hat{\beta}$. On the other hand, [13, 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. ∎

The superlinear convergence of PSNA whenever the second stage problems are P-matrix or Z-matrix LCPs can be established under weaker assumptions on the elements of $\mathcal{H}(x^*)$.

THEOREM 4.5. *Suppose that Assumptions 4.1(i) and 4.2 hold, the level set $\mathcal{L}_0$ is bounded, and $D = [l,u]$; or Assumptions 4.1(ii) and 4.2 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 3.2, 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 4.4, 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, (3.4) 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 [6, Lemma 2.1], there exists a diagonal matrix Λ_k with diagonal entries on $[0,1]$ such that

$$\begin{aligned} 0 &= \tilde{Q}(\hat{x}^{k+1}) - \hat{Q}(x^*) \\ &= (I - \Lambda_k)(\hat{x}^{k+1} - x^*) + \Lambda_k[H(x^k) + W^k_{\epsilon_k}(\hat{x}^{k+1} - x^k) - H(x^*)] \\ &= (I - \Lambda_k)(\hat{x}^{k+1} - x^*) + \Lambda_k[H(x^k) + W^k_{\epsilon_k}(\hat{x}^{k+1} - x^* + x^* - x^k) - H(x^*)]. \end{aligned} \tag{4.9}$$

The matrix $I - \Lambda_k + \Lambda_k W^k_{\epsilon_k}$ is nonsingular since $W^k_{\epsilon_k}$ is a P-matrix. Using (4.9), we get

$$
\begin{aligned}
\|\hat{x}^{k+1} - x^*\| &= \|(I - \Lambda_k + \Lambda_k W^k_{\epsilon_k})^{-1}\Lambda_k[H(x^k) - H(x^*) - W^k_{\epsilon_k}(x^k - x^*)]\| \\
&\le \|(I - \Lambda_k + \Lambda_k W^k_{\epsilon_k})^{-1}\Lambda_k\|\|H(x^k) - H(x^*) - W^k_{\epsilon_k}(x^k - x^*)\| \\
&\le \|(I - \Lambda_k + \Lambda_k W^k_{\epsilon_k})^{-1}\Lambda_k\|(\|H(x^k) - H(x^*) - W^k(x^k - x^*)\| \\
&\quad + \epsilon_k\|(x^k - x^*)\|) \\
&= o(\|x^k - x^*\|), \qquad (4.10)
\end{aligned}
$$

where the last equality is due to (2.1), Lemma 4.4 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

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

[12, 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 3.3, we can prove (3.7). This implies that $\hat{x}^{k+1}$ computed from Newton iteration (3.4) is always accepted when x^k is sufficiently close to x^*; that is $x^{k+1} = \hat{x}^{k+1}$. Therefore, (4.10) means that x^k converges to x^* superlinearly. ∎

COROLLARY 4.6. *Let D be a polyhedron. The sequence $\{x^k\}$ generated by PSNA globally and superlinearly converges to the unique solution of* (1.9) *if one of the following conditions holds.*

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

(ii) M_ℓ is a P-matrix for any $\xi_\ell \in \Xi_\nu$ and (4.7) *holds with strict inequality;*

(iii) M_ℓ is a Z-matrix for any $\xi_\ell \in \Xi_\nu$, (4.1)-(4.2) *has the relatively complete recourse, D is bounded and* (4.7) *holds with strict inequality.*

5. Numerical experiments.

In this section, we conduct numerical experiments to test the efficiency of PSNA for the large-scale two-stage SVI (4.1)-(4.2), and compare PSNA with PHA.

5.1. Randomly generated problems.

PSNA is terminated if

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

The starting point $x^0 \in \mathbb{R}^n_+$ is randomly chosen, $\alpha = 0.015$ and $\eta = 0.9$. The regularized parameter is set to $\epsilon_k = \min\{1, \|\hat{Q}(x^k)\|\}$. All codes were implemented in MATLAB R2018a on a laptop with Intel Core i7-4790 (3.6 GHz) and 32 GB RAM.

EXAMPLE 5.1. *Monotone two-stage SLCP in* [22][4]

[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 $\mu_k = 10^{-9}$.

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* $i = 1, \ldots, s$. *For* $\ell = 1, \ldots, \nu$, *randomly create* ν *antisymmetric matrices* $O_\ell \in \mathbb{R}^{(n+m)\times(n+m)}$. *Set*

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

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

EXAMPLE 5.2. *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* $B_\ell \in \mathbb{R}^{n\times m}$, $N_\ell \in \mathbb{R}^{m\times n}$, $U_\ell \in \mathbb{R}^{m\times m}$, $q_\ell \in \mathbb{R}^m$ *for* $\ell = 1, \ldots, \nu$, *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 [14], we set*

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

The second stage problem is as follows

$$0 \le y_\ell \perp M_\ell y_\ell + N_\ell f(x) + q_\ell \ge 0,\ \ \ell = 1, \ldots, \nu,$$

with $M_\ell = \Lambda_\ell + U_\ell$, $f(x) = (\sin x_1, \ldots, \sin x_n)^T$.

EXAMPLE 5.3. *Nonmonotone two-stage SVI with Z-matrix LCP in the second stage*

All parameters are generated in a same way as Example 5.2 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* $\bar{M}_{mm} = \bar{M}_{11} = 1$ *and* $\bar{M}_{21} = \ldots = \bar{M}_{k,k-1} = -2$. $q_k = \tilde{q}$ *and* $q_{k+1} = -\tilde{q}$ *with* $\tilde{q}$ *uniformly drawn from* $[0, 5]$, *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 $y_{k+1} = \ldots = y_{2k} = [(\xi_\ell + 1)\sum_{i=1}^n x_i + (\xi_\ell + 2)\tilde{q}]/\xi_\ell$ is a feasible point of the LCP$(N_\ell x + q_\ell, M_\ell)$.

EXAMPLE 5.4. *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*

5.2 except for $A(x)$, which is of the following form

$$A_1(x) = x_1^2 + \sum_{i=2}^{n-1}(x_i x_{i+1}) - \sum_{i=2}^{n} x_i + |x_1 - 1|,$$
$$A_2(x) = x_1(1 - x_3) + x_2^2 + |x_2 - 2|,$$
$$A_i(x) = x_1(1 - x_{i-1} - x_{i+1}) + x_i^2 + |x_i - i|,\ i = 3, \dots, n-1,$$
$$A_n(x) = x_1(1 - x_{n-1}) + x_n^2 + |x_n - n|,$$
$$A(x) = (A_1(x), \dots, A_n(x))^T + \lambda x + c, \quad \lambda > 1.$$

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 $x \in D$ when $\lambda > 2n + 1$. We set $\lambda = 2n + 2$ and 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 components $x_i^* = i$. The remaining components are set to 0 or n on a fifty-fifty basis, respectively.

By Remark 4.3, if $\min_i \bar{\Lambda}_{ii} - \tilde{\sigma} \geq 0$ in Examples 5.2-5.3 and $\lambda \geq 2n + \tilde{\sigma} + 1$ in Example 5.4 with $\tilde{\sigma}$ defined in (4.7), then the corresponding single-stage SVI reformulations of Examples 5.2-5.4 are monotone. However, since M_ℓ is a P-matrix or Z-matrix in Examples 5.2-5.4, these examples are not necessarily elicited monotone by [28, Theorem 3.5] and thus PHA and elicited PHA cannot be applied to solve them.

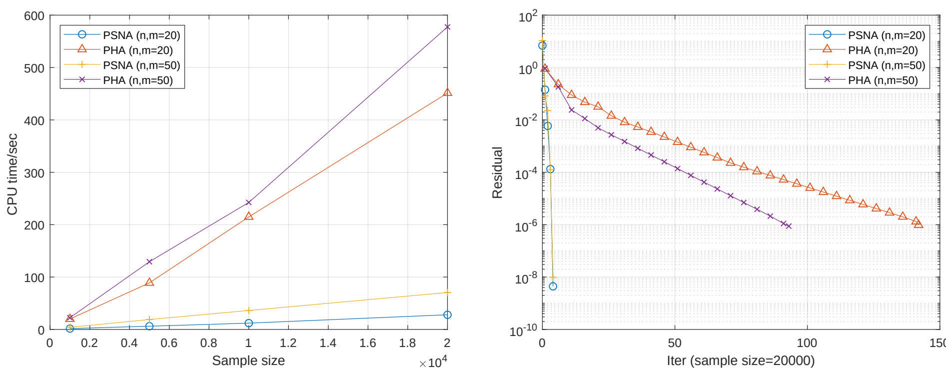
We compared our algorithm with PHA for solving Example 5.1, which is a monotone problem and also tested in [22]. Parameters of PHA in [22] are used in our numerical comparison. In Examples 5.1-5.4, each sample in the sample set $\{\xi_1, \dots, \xi_\nu\}$ has the equal probability $1/\nu$.

The numerical results for Example 5.1 are reported in Table 1 and Figure 1, in which the average performance profiles for 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 1, we set $n = m = 20$ and 50 and increase ν from 1,000 to 20,000. The dimensions of problems $(n+\nu m)$ are ranging from 20,020 to 1,000,050. For PSNA, the number of the Newton iteration (3.4) performed is denoted as "Iter/N". One can see that the Newton iteration is always used for all 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 1 shows that PSNA is much faster than PHA in terms of CPU time. The left figure of Figure 1 gives an intuitive comparison of the two algorithms for different n, m when ν increases from 1,000 to 20,000. 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, numerical results of PSNA for Example 5.2 are presented. We set $n = 30, m = 20$ and $n = 60, m = 50$, and increase ν from $10,000$ to $20,000$ to test 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 is observed. Similar results for Example 5.3 are presented in Table 3. Table 4 shows the results of Example 5.4, in which the influence of nonsmooth components (NSC) of the solution is explored, where NSC 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 more projection iterations when NSC is large. These results

Table 1: Comparison of PSNA and PHA for Example 5.1

		PSNA				PHA		
n, m	ν	Iter	Iter/N	CPU/sec	Res	Iter	CPU/sec	Res
20	1,000	4.5	4.5	1.6	6.0e-09	137.5	19.9	9.4e-07
	5,000	4.0	4.0	6.1	5.6e-08	134.0	88.9	9.6e-07
	10,000	4.0	4.0	12.0	2.0e-07	157.0	214.8	9.5e-07
	20,000	4.0	4.0	28.0	6.4e-09	161.5	451.4	9.5e-07
50	1,000	5.0	5.0	4.0	3.2e-15	72.5	22.4	9.6e-07
	5,000	4.5	4.5	18.9	9.4e-08	83.0	129.3	8.5e-07
	10,000	4.0	4.0	36.1	2.6e-07	78.5	242.5	9.1e-07
	20,000	4.0	4.0	70.5	2.0e-07	92.5	577.5	9.3e-07



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

Fig. 1: Comparison of PSNA and PHA.

suggest that PSNA is promising even for solving some nonmonotone problems. The good numerical performance for nonmonotone Examples 5.2-5.4 is partly supported by Corollary 4.6(ii)-(iii), which establishes the global and superlinear convergence of PSNA for some special nonmonotone problems, where the first stage problem is strongly monotone with respect to the first stage variable x and the second stage problem is a P-matrix LCP or Z-matrix LCP with respect to y.

5.2. Stochastic traffic assignments. In this subsection, 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 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; see [7, 12]. 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 traffic assignments.

- $\tilde{\mathcal{N}}, \mathcal{P}, \tilde{\mathcal{A}}, \mathcal{W}$: the node set, the path set, the link set and the origin destination (OD) pair set, respectively.
- $\mathcal{P}_w$: the set of paths that connect the OD pair $w \in \mathcal{W}$.

Table 2: Numerical results of PSNA for Example 5.2.

Case 1: $l = 0, u = \infty$									
		ν=10,000				ν=20,000			
n, m		Iter	Iter/N	CPU	Res	Iter	Iter/N	CPU	Res
	Max	3.0	3.0	4.0	2.6e-13	3.0	3.0	8.5	2.9e-13
$30, 20$	Ave	3.0	3.0	3.8	4.3e-10	3.0	3.0	8.1	3.0e-13
	Min	3.0	3.0	3.7	1.6e-13	3.0	3.0	7.9	2.5e-13
	Max	4.0	4.0	31.5	4.8e-13	3.0	3.0	55.2	4.8e-13
$60, 50$	Ave	3.3	3.3	26.8	3.6e-07	3.0	3.0	53.7	2.2e-07
	Min	3.0	3.0	24.6	3.3e-07	3.0	3.0	53.0	1.7e-07
Case 2: $l = -ne_n, u = ne_n$									
	Max	4.0	4.0	4.7	3.5e-13	4.0	4.0	10.3	5.8e-13
$30, 20$	Ave	3.2	3.2	4.0	3.5e-07	3.1	3.1	8.3	1.3e-07
	Min	3.0	3.0	3.7	7.2e-12	3.0	3.0	8.0	2.4e-07
	Max	5.0	5.0	40.2	7.5e-13	5.0	5.0	80.0	1.0e-12
$60, 50$	Ave	4.4	4.4	35.0	8.2e-08	4.1	4.1	67.9	1.3e-07
	Min	4.0	4.0	30.9	2.6e-08	4.0	4.0	65.8	5.7e-09
Case 3: $l_i = -n, u_i = n$ if i is even and $l_i = 0, u_i = \infty$ if i is odd									
	Max	3.0	3.0	3.8	3.5e-08	3.0	3.0	8.2	3.9e-13
$30, 20$	Ave	3.0	3.0	3.7	7.1e-08	3.0	3.0	8.0	8.8e-09
	Min	3.0	3.0	3.7	3.7e-07	3.0	3.0	7.9	4.0e-13
	Max	4.0	4.0	45.0	4.7e-13	4.0	4.0	95.9	1.1e-12
$60, 50$	Ave	4.0	4.0	34.8	7.1e-13	4.0	4.0	80.3	2.5e-10
	Min	4.0	4.0	31.2	8.1e-13	4.0	4.0	64.1	1.0e-12

Table 3: Numerical results of PSNA for Example 5.3.

n, m	ν	Iter	Iter/N	CPU/sec	Res
	2,000	1.7	1.7	3.6	1.7e-12
20	10,000	1.4	1.4	16.6	1.0e-11
	20,000	1.5	1.5	37.2	9.3e-12
	2,000	1.6	1.6	16.0	1.9e-11
50	10,000	1.3	1.3	73.8	6.8e-11
	20,000	1.4	1.4	162.8	7.3e-11

Table 4: Numerical results of PSNA for Example 5.4 with $\nu = 20,000$.

n, m	NSC	Iter	Iter/N	CPU/sec	Res
	0.2	5.7	4.8	13.7	1.2e-07
$30, 20$	0.4	6.2	4.5	15.8	1.9e-07
	0.6	6.0	4.6	15.9	1.6e-08
	0.8	6.5	4.3	16.7	7.0e-08
	0.2	6.2	4.9	102.2	2.8e-08
$60, 50$	0.4	6.0	5.0	101.8	4.9e-10
	0.6	6.9	5.0	110.7	2.0e-09
	0.8	11.3	5.0	153.2	3.4e-11

- $\Upsilon \in \mathbb{R}^{|\tilde{\mathcal{A}}|\times|\mathcal{P}|}$: 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 $v(\xi) = \Upsilon h(\xi)$.
- $c_a(\xi)$: the link capacity on link a, which is a positive scalar.
- $d_w(\xi)$: the nonnegative demand function for OD pair $w \in \mathcal{W}$.
- $R_p(h(\xi), \xi)$: the travel cost function through path p.
- $r_a(v(\xi), \xi)$: the travel cost function through link a.

Let

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

The matrix Γ has elements 0 or 1 only and each column of Γ has exactly one element being 1. By the boundedness of $d(\xi)$, it is known that D and $\hat{D}_\xi$ are bounded polyhedral sets. The function $r : \mathbb{R}^{|\tilde{\mathcal{A}}|} \times \mathbb{R}^d \to \mathbb{R}^{|\tilde{\mathcal{A}}|}$ is the generalized bureau of public road (GBPR) link travel time function [2] defined as

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

where t_a^0 and n_a are given positive numbers. Define the path travel cost functions $\bar{R} : \mathbb{R}^{|\mathcal{P}|} \to \mathbb{R}^{|\mathcal{P}|}$ and $R : \mathbb{R}^{|\mathcal{P}|} \times \mathbb{R}^d \to \mathbb{R}^{|\mathcal{P}|}$ as follows

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

The stochastic user equilibrium can be formulated as an SVI [7]: find $h(\xi) \in \hat{D}_\xi$ such that

$$(5.1) \qquad (h' - h(\xi))^T R(h(\xi), \xi) \geq 0, \quad \forall\ h' \in \hat{D}_\xi, \quad \text{for any } \xi \in \Xi_\nu.$$

To solve (5.1) with a fixed ξ, one can minimize the following optimization problem

$$(5.2) \qquad \min_{x \in \hat{D}_\xi}\ \max\{(x - h(\xi))^T R(x, \xi) \mid h(\xi) \in \hat{D}_\xi\},$$

which can be written as a two-stage optimization problem

$$(5.3) \qquad \begin{aligned} &\min\ x^T R(x, \xi) + Q(x, \xi) \\ &\text{s.t. } x \in \hat{D}_\xi, \\ &\quad Q(x, \xi) = \max\{-h(\xi)^T R(x, \xi) \mid h(\xi) \in \hat{D}_\xi\}. \end{aligned}$$

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

$$(5.4) \qquad \begin{aligned} &\min\ x^T \bar{R}(x) + \mathbb{E}[Q(x, \xi)] \\ &\text{s.t. } x \in D, \\ &\quad Q(x, \xi) = \min\{s(\xi)^T d(\xi) \mid \Gamma^T s(\xi) + R(x, \xi) \geq 0\}, \quad \text{for any } \xi \in \Xi_\nu. \end{aligned}$$

Following [2, Example 2.3], we can obtain the first-order optimality condition of (5.4) as follows

$$(5.5)\qquad -\left(\nabla\bar{R}(x)^T x+\bar{R}(x)-\mathbb{E}[\nabla R(x,\xi)^T\lambda(\xi)]\right)\in\mathcal{N}_D(x),$$

$$(5.6)\qquad -\left[\begin{pmatrix} 0 & -\Gamma \\ \Gamma^T & 0\end{pmatrix} y(\xi)+\begin{pmatrix} d(\xi) \\ R(x,\xi)\end{pmatrix}\right]\in\mathcal{N}_C(y(\xi)),\quad \text{any } \xi\in\Xi_\nu,$$

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 5.1. We can show that problem (5.5)-(5.6) 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 $\Gamma\bar{\lambda}(\xi)\geq d(\xi)$ and $\bar{z}(\xi)=0$. Thus, $(\bar{z}(\xi),\bar{\lambda}(\xi))$ is a feasible solution of the following LCP

$$(5.7)\qquad 0\leq\begin{pmatrix} z(\xi) \\ \lambda(\xi)\end{pmatrix}\perp\begin{pmatrix} 0 & \Gamma \\ -\Gamma^T & 0\end{pmatrix}\begin{pmatrix} z(\xi) \\ \lambda(\xi)\end{pmatrix}+\begin{pmatrix} -d(\xi) \\ R(x,\xi)\end{pmatrix}\geq 0.$$

Then, the LCP in (5.7) is solvable by [11, Theorem 3.1.2].

Let $(z^*(x,\xi),\lambda^*(x,\xi))^T$ be a solution of (5.7) for fixed $x\in D$ and $\xi\in\Xi_\nu$. Now we show that $(-z^*(x,\xi),\lambda^*(x,\xi))^T$ is a solution of (5.6). If there is $w'\in\mathcal{W}$ such that $(\Gamma\lambda^*(x,\xi)-d(\xi))_{w'}>0$, by the first complementarity condition in (5.7), we have $z^*_{w'}(x,\xi)=0$. Thus, $(R(x,\xi)-\Gamma^T z^*(x,\xi))_p=R_p(x,\xi)>0$ for any $p\in\mathcal{P}_{w'}$. Then, we have $\lambda^*_p(x,\xi)=0$ for any $p\in P_{w'}$ by the second complementarity condition in (5.7), which implies that $(\Gamma\lambda^*(x,\xi)-d(\xi))_{w'}=-d(\xi)_{w'}\leq 0$. This is a contradiction. Hence $(-z^*(x,\xi),\lambda^*(x,\xi))^T$ is a solution of (5.6).

By the positive semi-definiteness of the coefficient matrix of $y(\xi)$ in (5.6), it admits a unique least-norm solution[5] by [11, Theorem 3.1.7], denoted by $\hat{y}(x,\xi)$. By substituting $\hat{y}(x,\xi)$ into the first stage problem (5.5), we can get the single-stage SVI formulation of (5.5)-(5.6). We can calculate a solution of the original two-stage problem by solving the single-stage problem, since D is a bounded polyhedral set. To obtain the least-norm solution, we add a regularized term $\mu_k I$ with $\mu_k>0$ and $\mu_k\to 0$ as $k\to\infty$ to the coefficient matrix of $y(\xi)$, which forces the second stage problem to be strongly monotone and thus admit a unique solution $y_{\mu_k}(x,\xi)$ for any fixed x and ξ. In addition, the solution function $\hat{y}_{\mu_k}(x,\xi)$ of the regularized second stage problem with any $\mu_k>0$ is Lipschitz continuous with respect to x for any ξ and $\lim_{k\to\infty}\hat{y}_{\mu_k}(x,\xi)=\hat{y}(x,\xi)$ by [11, Theorem 5.6.2].

We test the efficiency of PSNA for solving (5.5)-(5.6) with Nguyen and Dupuis network, which has 13 nodes, 19 links, 25 paths and 4 OD pairs; see [7] for details. The data for demands $d(\xi)$, capacities $c(\xi)$ and the free travel time t^0 are set according to the data $\tilde{d}(\xi),\tilde{c}(\xi),\tilde{t}^0$ used in [7] after a scaling, i.e., $d(\xi)=0.1\times\tilde{d}(\xi)$, $c(\xi)=0.1\times\tilde{c}(\xi)$ and $t^0=0.1\times\tilde{t}^0$. Parameter n_a in $R(x,\xi)$ is set to $n_a=2,\ldots,5$, respectively. Note that PHA fails to solve problem (5.5)-(5.6), since the problem is nonmonotone for $n_a\geq 2$. The settings for PSNA are $\mu_k\equiv 10^{-12}$, $\epsilon_k\equiv 0$, $\eta=0.9$ and the step size for the projection iteration (3.5) is set to $\alpha=0.1,0.05,0.05,0.05$ for $n_a=2,\ldots,5$, respectively. The sample size is set to $\nu=100{,}000$ and $400{,}000$. Numerical results were reported in Table 5, which show that PSNA can solve these nonmonotone problems efficiently.

[5] A solution $\bar{y}$ of the LCP(q,M) is called the least-norm solution if $\|\bar{y}\|\leq\|y\|$ for any $y\in$ SOL(q,M).

Table 5: Results of PSNA for (5.5)-(5.6) with Nguyen and Dupuis network ($n = 25, m = 29$).

ν	n_α	Iter	Iter/N	CPU/sec	Res
100, 000	2	5.0	5.0	105.2	2.7e-07
	3	7.0	6.0	173.1	4.2e-07
	4	12.0	6.0	316.0	2.4e-07
	5	10.0	6.0	267.1	6.2e-08
400, 000	2	5.0	5.0	414.0	1.1e-08
	3	7.0	6.0	690.5	4.3e-07
	4	12.0	6.0	1247.5	2.3e-07
	5	10.0	6.0	1063.8	6.2e-08

6. Conclusions. Algorithm 3.1 describes a hybrid projection semismooth Newton algorithm (PSNA) for solving the two-stage SVI (1.5)-(1.6). We give sufficient conditions to guarantee that the sequence generated by Algorithm 3.1 globally and superlinearly converges to a solution of (1.5)-(1.6). Moreover, we show these conditions hold for Examples 5.1-5.4 and the example from stochastic traffic assignments with properly selected parameters. However Examples 5.2-5.4 are not (elicited) monotone two-stage SVI and cannot be solved by PHA and elicited PHA. Preliminary numerical experiments with over 10^7 variables show the effectiveness and efficiency of the proposed PSNA for solving large-scale two-stage SVI.

Acknowledgement. We would like to thank Professor Akil Narayan and two referees for their helpful comments which help us to improve the content and readability of this paper. We would also like to thank Professor Jie Sun and Dr. Min Zhang for providing us their PHA code for our numerical experiments.

REFERENCES

[1] R. J. Aumann, *Integrals of set-valued functions*, J. Math. Anal. Appl., 12 (1965), pp. 1–12.

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

[3] 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*, Math. Comput., 67 (1998), pp. 519–540.

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

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

[6] X. Chen and Z. Wang, *Computational error bounds for a differential linear variational inequality*, IMA J. Numer. Anal., 32 (2012), pp. 957–982.

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

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

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

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

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

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

[13] S. A. Gabriel and J. J. Moré, *Smoothing of mixed complementarity problems*, Complemen-

tarity and Variational Problems: State of the Art, (1997), pp. 105–116.

[14] 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.

[15] M. Hintermüller, K. Ito, and K. Kunisch, *The primal-dual active set strategy as a semismooth Newton method*, SIAM J. Optim., 13 (2002), pp. 865–888.

[16] H. Jiang and L. Qi, *Local uniqueness and convergence of iterative methods for nonsmooth variational inequalities*, J. Math. Anal. Appl., 196 (1995), pp. 314–331.

[17] J. Kyparisis, *Solution differentiability for variational inequalities*, Math. Program., 48 (1990), pp. 285–301.

[18] J.-S. Pang, *Newton's method for B-differentiable equations*, Math. Oper. Res., 15 (1990), pp. 311–341.

[19] L. Qi and X. Chen, *A globally convergent successive approximation method for severely nonsmooth equations*, SIAM J. Control Optim., 33 (1995), pp. 402–418.

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

[21] S. M. Robinson, *Strongly regular generalized equations*, Math. Oper. Res., 5 (1980), pp. 43–62.

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

[23] R. T. Rockafellar and R. J.-B. Wets, *Scenarios and policy aggregation in optimization under uncertainty*, Math. Oper. Res., 16 (1991), pp. 119–147.

[24] R. T. Rockafellar and R. J.-B. Wets, *Variational Analysis*, vol. 317, Springer Science & Business Media, 2009.

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

[26] J. Shen and T. M. Lebair, *Shape restricted smoothing splines via constrained optimal control and nonsmooth Newton's methods*, Automatica, 53 (2015), pp. 216–224.

[27] M. Ulbrich, *Semismooth Newton Methods for Variational Inequalities and Constrained Optimization Problems in Function Spaces*, SIAM, 2011.

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