A PROJECTION-BASED ALGORITHM FOR SOLVING STOCHASTIC INVERSE VARIATIONAL INEQUALITY PROBLEMS

Zeinab Alizadeh
Felipe Parra Polanco
Afrooz Jalilzadeh

Systems and Industrial Engineering
The University of Arizona
Tucson, AZ 85721, USA

ABSTRACT

We consider a stochastic Inverse Variational Inequality (IVI) problem defined by a continuous and co-coercive map over a closed and convex set. Motivated by the absence of performance guarantees for stochastic IVI, we present a variance-reduced projection-based gradient method. Our proposed method ensures an almost sure convergence of the generated iterates to the solution, and we establish a convergence rate guarantee. To verify our results, we apply the proposed algorithm to a network equilibrium control problem.

1 INTRODUCTION

In recent years, the inverse variational inequality (IVI) problem (He 1999) has become an active research area within the field of mathematical optimization. The IVI problem is pertinent to numerous fields, including transportation system operation, control policies, and electrical power network management (Yang and Bell 1997; He and Liu 2011; Scrimali 2012; He and Dong 2018). Depending on the specific application, factors such as the demand market, supply market, or transaction cost may exhibit stochastic behavior. In this paper, we focus on Stochastic Inverse Variational Inequality (SIVI). In particular, let $X \subseteq \mathbb{R}^n$ be a nonempty closed convex set, and $F : X \to \mathbb{R}^n$ be a continuous nonlinear map. Consider the following SIVI problem: find $x^* \in \mathbb{R}^n$ such that $F(x^*) \in X$ and

$$\langle y - F(x^*), x^* \rangle \geq 0, \quad \forall y \in X, \tag{1}$$

where $F(x) \triangleq \mathbb{E}[G(x, \xi)]$, $\xi : \Omega \to \mathbb{R}^d$, $G : X \times \mathbb{R}^d \to \mathbb{R}^n$, and the associated probability space is denoted by $(\Omega, \mathscr{F}, \mathbb{P})$. The SIVI problem involves finding a solution that satisfies a set of inequalities for all variable y lying within the set X. Indeed, SIVI in (1) can be viewed as a classical Stochastic Variational Inequality (SVI) problem, i.e., $\langle y - x^*, H(x^*) \rangle \geq 0$ for any $y \in X$, where $H \triangleq F^{-1}$. However, F^{-1} may not be available in some practical applications which necessitates developing a new set of schemes.

While VIs (Nesterov and Scrimali 2011; Facchinei and Pang 2007; Malitsky 2015) and SVIs (Jiang and Xu 2008; Koshal et al. 2012; Yousefian et al. 2017; Jalilzadeh and Shanbhag 2019) have been studied extensively over the last several decades, less is known about SIVIs. Projection-based algorithms to solve deterministic IVI problems have been developed by He and Liu (2011), He and Dong (2018) and Luo and Yang (2014). Moreover, Zou et al. (2016) introduced a neural network-based method to approximate the solutions to IVIs, however to the best of our knowledge there is no available algorithm with a convergence rate guarantee for solving the SIVI. In this paper, we introduce a variance-reduced projection-based algorithm for solving problem (1) when operator F is co-coercive, which is a weaker assumption than strong monotonicity plus Lipschitz continuity. We show that the proposed scheme produces

a sequence that converges to the solution of (1) almost surely, a convergence statement that was unavailable thus far for this class of problems. Then, we derive rate and complexity guarantees for the proposed algorithm.

The remainder of the paper is organized as follows. In Section 2, we define our notations and state assumptions and some technical lemmas that we use throughout the paper. In Section 3, we propose the Variance-reduced Inverse Projected Gradient (VR-IPG) method to solve the SIVI problem (1), and its convergence analysis is stated in Section 3.1. Finally, numerical experiments are presented in Section 4 to solve a network equilibrium control problem.

2 PRELIMINARIES

In this section, first, we define important notations and then the main assumptions that we need for the convergence analysis are stated.

Notations. Throughout the paper, $\|x\|$ denotes the Euclidean vector norm, i.e., $\|x\| = \sqrt{x^T x}$. $\mathbf{P}_X[u]$ is the projection of u onto the set X, i.e. $\mathbf{P}_X[u] = \operatorname{argmin}_{z \in X} \|z - u\|$. $\mathbb{E}[x]$ is used to denote the expectation of a random variable x. Moreover, $e_n \in \mathbb{R}^n$ denotes the vector of ones and for a given $a \in \mathbb{R}^n$, $\operatorname{diag}(a) \in \mathbb{R}^{n \times n}$ denotes a diagonal matrix whose main diagonal elements are a.

Assumption 1. *Operator $F : X \to \mathbb{R}^n$ is a co-coercive map on X, i.e., there exists $\mu > 0$ such that*

$$\langle F(x) - F(y), x - y \rangle \geq \mu \|F(x) - F(y)\|^2, \quad \forall x, y \in X.$$

Remark 1. *It is worth noting that for a Lipschitz continuous map F, co-coercivity is a weaker assumption than strong monotonicity. In other words, a co-coercive map may not be strongly monotone, such as a constant mapping. However, a map that is both strongly monotone and Lipschitz continuous is co-coercive. Moreover, co-coercive maps have some desirable properties. For example, co-coercivity is preserved under affine transformations. For instance, if F is a co-coercive map, then $A^T F(Ax) + d$ is also co-coercive, where A is a matrix and d is a vector. This property highlights the versatility of co-coercivity in various settings. See Zhu and Marcotte (1996) for more discussion about the properties of co-coercive maps.*

If $\mathscr{F}_k$ denotes the information history at epoch k, then we have the following requirements on the associated filtrations where $\bar{w}_{k,N_k} \triangleq \frac{1}{N_k} \sum_{j=1}^{N_k} (G(x_k, \xi_{j,k}) - F(x_k))$.

Assumption 2. *There exists $\nu > 0$ such that $\mathbb{E}[\bar{w}_{k,N_k} \mid \mathscr{F}_k] = 0$ and $\mathbb{E}[\|\bar{w}_{k,N_k}\|^2 \mid \mathscr{F}_k] \leq \frac{\nu^2}{N_k}$ holds almost surely for all k, where $\mathscr{F}_k \triangleq \sigma\{x_0, x_1, \ldots, x_{k-1}\}$.*

In our analysis, the following technical lemma for projection mappings is used.

Lemma 1. *(Bertsekas et al. 2003) Let $X \subseteq \mathbb{R}^n$ be a nonempty closed and convex set. Then the followings hold: (a) $\|\mathbf{P}_X[u] - \mathbf{P}_X[v]\| \leq \|u - v\|$ for all $u, v \in \mathbb{R}^n$; (b) $(u - \mathbf{P}_X[u])^T (x - \mathbf{P}_X[u]) \leq 0$ for all $u \in \mathbb{R}^n$ and $x \in X$.*

Moreover, to prove almost sure convergence of the iterates we use the following Robbins-Siegmund lemma (see Lemma 11 in Polyak (2003)).

Lemma 2. *[Robbins-Siegmund] Let $v_k, u_k, \alpha_k, \beta_k$ be nonnegative random variables and let $\mathbb{E}[v_{k+1} \mid \mathscr{F}_k] \leq (1 + \alpha_k)v_k - u_k + \beta_k$, $\sum_{k=0}^{\infty} \alpha_k < \infty$ and $\sum_{k=0}^{\infty} u_k < \infty$ almost surely, where $\mathbb{E}[v_{k+1} \mid \mathscr{F}_k]$ denotes the conditional mathematical expectation for the given $v_0, \ldots, v_k, u_0, \ldots, u_k, \alpha_0, \ldots, \alpha_k, \beta_0, \ldots, \beta_k$. Then, $v_k \to v$ almost surely and $\sum_{k=0}^{\infty} u_k < \infty$ almost surely, where $v \geq 0$ is some random variable.*

3 PROPOSED METHOD

In this section, we propose our algorithm for solving problem (1). First, we show that solving SIVI problem (1) is equivalent to finding $x^* \in \mathbb{R}^n$ such that

$$F(x^*) = \mathbf{P}_X(F(x^*) - \eta x^*). \tag{2}$$

Proposition 1. *x^* is a solution of problem (1) if and only if x^* is a solution of equation (2).*

Proof. We start by rewriting the projection equation (2) which is equivalent to $F(x^*) \in \operatorname{argmin}_{y \in X} \|y - (F(x^*) - \eta x^*)\|^2$. We observe that $F(x^*) \in X$ and since the objective function is convex, the first-order optimality condition is equivalent to finding a global solution. Hence, $\langle y - F(x^*), F(x^*) - (F(x^*) - \eta x^*) \rangle \geq 0$ for all $y \in X$. Therefore, $\langle y - F(x^*), x^* \rangle \geq 0$ for all $y \in X$ and $F(x^*) \in X$ which means that x^* is the solution of problem (1). □

To address the stochastic nature of the problem, we utilize a stochastic approximation (SA) scheme, which has shown effectiveness in various stochastic computational problems. Conventional SA schemes rely on a single or fixed batch of samples to estimate the expectation. Our proposed approach is a projected gradient-based scheme where, in each iteration, $F(x_k)$ is substituted by a gradually increasing sample average using a batch size of N_k, i.e., $\frac{1}{N_k} \sum_{j=1}^{N_k} G(x_k, \xi_{j,k})$. A projection step generates an intermediate iteration z_k and the new iterate x_{k+1} is then generated by taking a direction along $-(\frac{1}{N_k} \sum_{j=1}^{N_k} G(x_k, \xi_{j,k}) - z_k)$ with step-size $1/\eta_k$ from the current iterate point x_k. The outline of the proposed variance-reduced inverse projected gradient (VR-IPG) method is displayed in Algorithm 1. Moreover, to measure how far the iterates are from the optimal solution, we can examine the degree to which the optimality condition in (2) is violated. Therefore, we define the gap function $H : X \times \mathbb{R}_+ \to \mathbb{R}_+$ such that $H(x, \eta) \triangleq \frac{1}{\eta}(F(x) - \mathbf{P}_X(F(x) - \eta x))$ which will be used to analyze the convergence rate of the proposed method.

Algorithm 1 Variance-reduced Inverse Projected Gradient (VR-IPG) method

---

Input: $x_0 \in X$, $\{\eta_k, N_k\}_{k \geq 0} \subseteq \mathbb{R}_+$;

for $k = 0, \ldots T - 1$ **do**

$$z_k = \mathbf{P}_X \left[\frac{\sum_{j=1}^{N_k} G(x_k, \xi_{j,k})}{N_k} - \eta_k x_k \right];$$

$$x_{k+1} = x_k - \frac{1}{\eta_k} \left(\frac{\sum_{j=1}^{N_k} G(x_k, \xi_{j,k})}{N_k} - z_k \right);$$

end for

3.1 Convergence Analysis

To obtain the convergence results of Algorithm 1, we state a one-step analysis in Lemma 4. To this end, we first establish the following technical result that will be utilized in the proof of Lemma 4. Related proofs are provided in the appendix.

Lemma 3. *Define $H(x_k, \eta_k) \triangleq \frac{1}{\eta_k}(F(x_k) - \mathbf{P}_X(F(x_k) - \eta_k x_k))$ and let $x^* \in X$ be a solution point for problem (1). For any $x \in \mathbb{R}^n$, the following holds*

$$(x - x^*)^T H(x, \eta) \geq \left(1 - \frac{1}{4\mu\eta}\right) \|H(x, \eta)\|^2.$$

Lemma 4. *Consider iterates $\{x_k\}_{k \geq 0}$ generated by Algorithm 1 and suppose Assumptions 1 and 2 hold, then the following for any $k \geq 0$:*

$$\mathbb{E}[\|x_{k+1} - x^*\|^2 \mid \mathscr{F}_k] \leq \|x_k - x^*\|^2 - \left(1 - \frac{1}{2\eta_k\mu}\right)\|H(x_k, \eta_k)\|^2 + \frac{6\nu^2}{\eta_k^2 N_k} + \frac{2\nu}{\eta_k\sqrt{N_k}}\|x^*\|. \tag{3}$$

Now we are ready to present the main results of this paper. In Theorem 1, we show that the iterates generated by Algorithm 1 converge to a solution of problem (1) almost surely.

Theorem 1 (Almost sure convergence). *Consider iterates $\{x_k\}_{k \geq 0}$ generated by Algorithm 1 and suppose Assumptions 1 and 2 hold. Let $\eta_k = \eta$ for some $\eta > 0$ and suppose $\{N_k\}_k$ is an increasing sequence, such that $\sum_{k=0}^{\infty} \frac{1}{\sqrt{N_K}} < \infty$. Then $x_k \to x^*$ almost surely.*

Proof. Consider (3), if $\eta_k = \eta$ for some $\eta > 0$ and $\sum_{k=0}^{\infty} \frac{1}{\sqrt{N_K}} < \infty$, the requirements of Lemma 2 are satisfied and we can conclude that:

(i) $\|x_k - x^*\|^2$ is a convergent sequence. Hence, $\{x_k\}_k$ is a bounded sequence which implies that it has a convergent subsequence $x_{k_q} \to x_\#$ as $q \to \infty$ for some $x_\# \in X$;

(ii) $\sum_{k=0}^{\infty} \left(1 - \frac{1}{2\eta_k\mu}\right) \|H(x_k, \eta)\|^2 < \infty$ almost surely. Since $\left(1 - \frac{1}{2\eta_k\mu}\right) > 0$, hence we have that $\|H(x_k, \eta)\|^2 \to 0$ almost surely. From the definition of $H(x_k, \eta)$, we conclude that $\lim_{k\to\infty} F(x_k) - \mathbf{P}_X(F(x_k) - \eta x_k) \to 0$ which implies that $\lim_{q\to\infty} F(x_{k_q}) - \mathbf{P}_X(F(x_{k_q}) - \eta x_{k_q}) \to 0$.

From (i), we know that $x_{k_q} \to x_\#$ almost surely, so from (ii), we have that $F(x_\#) - \mathbf{P}_X(F(x_\#) - \eta x_\#) = 0$ which means that $x_\#$ is a solution of (1). Now we show that $z_\#$ is the unique limit point. Note that (3) is true for any optimal solution, so one can write (3) for $x^* = x_\#$. Invoking Lemma (2) again we conclude that $\|x_k - x_\#\|^2$ is a convergent sequence, i.e., there exists $a \geq 0$ such that $a = \lim_{k\to\infty} \|x_k - x_\#\|^2$. Since x_{k_q} is a subsequence of x_k, then we know that $a = \lim_{k\to\infty} \|x_k - x_\#\|^2 = \lim_{q\to\infty} \|x_{k_q} - x_\#\|^2 = 0$. Therefore, we conclude that $\lim_{k\to\infty} \|x_k - x_\#\|^2 = 0$ almost surely which means that $x_k \to x_\#$ almost surely. $\qquad\square$

Now, in Theorem 2 we demonstrate the convergence rate of the proposed method.

Theorem 2 (Convergence rate). *Let $\{x_k\}_{k\geq 0}$ be the iterates generated by Algorithm 1 and suppose Assumptions 1 and 2 hold. Choose $\eta_k = \eta \geq \frac{1}{2\mu}$. Let $N_k = \lceil (k+1)^{2+2\delta} \rceil$ where $\delta > 0$ for all $k \geq 0$, Then the following holds:*

$$\min_{k\in\{0,\dots,T-1\}} \mathbb{E}[\|H(x_k,\eta)\|^2] \leq \frac{1}{T\left(1-\frac{1}{2\eta\mu}\right)}[\|x_0 - x^*\|^2 + \frac{\pi^2 v^2}{\eta^2} + \frac{2v\|x^*\|}{\eta}(1+\frac{1}{\delta})] = \mathcal{O}(1/T).$$

Proof. Taking expectations from both sides of (3) and use the fact that $\mathbb{E}[\mathbb{E}[\|x_{k+1} - x^*\|^2 \mid \mathscr{F}_k]] = \mathbb{E}[\|x_{k+1} - x^*\|^2]$, we have:

$$\mathbb{E}[\|x_{k+1} - x^*\|^2] \leq \mathbb{E}[\|x_k - x^*\|^2] - \left(1 - \frac{1}{2\eta_k\mu}\right)\mathbb{E}[\|H(x_k,\eta_k)\|^2] + \frac{6v^2}{\eta_k^2 N_k} + \frac{2v}{\eta_k\sqrt{N_K}}\|x^*\|,$$

Now, by summing over $k = 0, \dots, T-1$ and putting $\eta_k = \eta$ the following holds:

$$\sum_{k=0}^{T-1}\left(1 - \frac{1}{2\eta\mu}\right)\mathbb{E}[\|H(x_k,\eta)\|^2] \leq \|x_0 - x^*\|^2 + \sum_{k=0}^{T-1}\frac{6v^2}{\eta^2 N_k} + \sum_{k=0}^{T-1}\frac{2v}{\eta\sqrt{N_K}}\|x^*\|,$$

Now, we can provide a lower-bound for the left-hand-side of the above inequality by $\min_k \mathbb{E}[\|H(x_k,\eta)\|^2]$, then we have:

$$T\left(1 - \frac{1}{2\eta\mu}\right)\min_k \mathbb{E}[\|H(x_k,\eta)\|^2] \leq \|x_0 - x^*\|^2 + \sum_{k=0}^{T-1}\frac{6v^2}{\eta^2 N_k} + \sum_{k=0}^{T-1}\frac{2v}{\eta\sqrt{N_K}}\|x^*\|. \qquad (4)$$

Let $\eta > \frac{1}{2\mu}$, and select $N_k = \lceil (k+1)^{2+2\delta} \rceil$, for some $\delta > 0$. Therefore, one can deduce the following inequalities using simple algebra.

$$\sum_{k=0}^{T-1}\frac{1}{N_k} \leq \sum_{k=0}^{T-1}\frac{1}{(k+1)^{2+2\delta}} \leq \frac{\pi^2}{6},$$

$$\sum_{k=0}^{T-1}\frac{1}{\sqrt{N_k}} \leq \sum_{k=0}^{T-1}\frac{1}{(k+1)^{1+\delta}} \leq 1 + \sum_{k=1}^{T-1}\frac{1}{(k+1)^{1+\delta}} \leq 1 + \int_0^{T-1}\frac{1}{(x+1)^{1+\delta}}\,dx \leq 1 + \frac{1}{\delta} - \frac{1}{T\delta} \leq 1 + \frac{1}{\delta}.$$

Next, using the above inequalities in (4), we obtain

$$T\left(1 - \frac{1}{2\eta\mu}\right)\min_k \mathbb{E}[\|H(x_k,\eta)\|^2] \leq \|x_0 - x^*\|^2 + \frac{\pi^2 v^2}{\eta^2} + \frac{2v\|x^*\|}{\eta}(1+\frac{1}{\delta}).$$

Finally, dividing both sides of previous inequality by $T\left(1-\frac{1}{2\eta\mu}\right)$ leads to

$$\min_{k\in\{0,...,T-1\}}\mathbb{E}[\|H(x_k,\eta)\|^2]\le\frac{1}{T\left(1-\frac{1}{2\eta\mu}\right)}[\|x_0-x^*\|^2+\frac{\pi^2v^2}{\eta^2}+\frac{2v\|x^*\|}{\eta}(1+\frac{1}{\delta})]=\mathscr{O}(1/T).$$

$\square$

Remark 2. *In Theorem 2, we established the convergence rate of the proposed algorithm. In particular, we showed that the proposed method finds an ε-gap in $\mathscr{O}(1/\varepsilon)$ iterations. Since at each iteration of the method, N_k samples are required, the total number of samples (oracle complexity) to achieve ε-gap is $\sum_{k=0}^{T-1}N_k\ge\int_0^{T-1}(x+1)^{2+2\delta}dx=\mathscr{O}(1/\varepsilon^{3+2\delta})$.*

4 NUMERICAL EXPERIMENTS

In this section, to illustrate the effectiveness of the proposed scheme, we first solve a test problem (Example 1) and then we solve a network equilibrium control problem (Example 2). The simulations are conducted in MATLAB.

Example 1. Consider problem (1) for $G(x,\xi)=Ax+b(\xi)$, where

$$A=\begin{pmatrix}5.0&2.0&1.0\\2.0&5.0&0.0\\1.0&0.0&6.0\end{pmatrix},\quad b(\xi)=\begin{pmatrix}0.0\\-3.0\\-5.5\end{pmatrix}+\xi,$$

$\xi\in\mathbb{R}^3$ is generated randomly from a standard Gaussian distribution, and $X=\{x\in\mathbb{R}^3\mid x_i\in[-1,10],\ i=1,2,3\}$. It is easy to check that $F(x)=\mathbb{E}[G(x,\xi)]$ is a co-coercive map and the unique optimal solution of the problem is $x^*=[0,0.4,0.75]^T$. Our proposed method is implemented for 20 replications with the same starting point and the mean of replications are plotted. Figure 1 (left) shows the convergence of VR-IPG for different choices of stepsizes when sample size $N_k=(k+1)^{2+2\delta}$ with $\delta=1/2$.
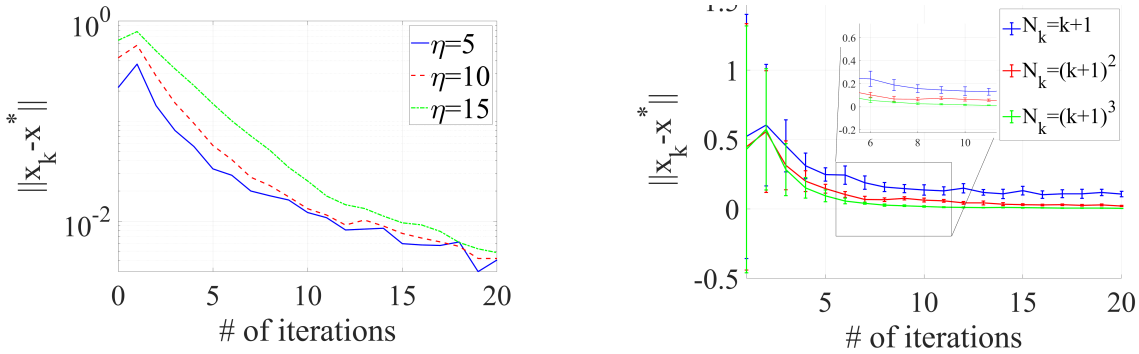

Figure 1: Example 1: Convergence of VR-IPG for different choices of stepsize (left); Almost sure convergence of VR-IPG for different sample sizes.

We also implement VR-IPG for different choices of N_k and the mean of 20 replications and their 95% confidence intervals are plotted in Figure 1 (right). From the plots, we observe that the confidence bands are becoming tighter as $k\to\infty$.

Example 2. In this example, we consider a network equilibrium control problem described in He et al. (2010). The problem is motivated by calculating the equilibrium in a supply and demand network. The goal is to control resource exploitations on the supply market and commodity consumptions on the demand markets by regulating taxes as a control variable denoted by $x_1\in\mathbb{R}^m$ and $x_2\in\mathbb{R}^n$, respectively. Let $x=[x_1^\top\ x_2^\top]^\top\in\mathbb{R}^{m+n}$ and the commodity shipment between supply and demand markets are denoted

by $\mathbf{a} = [[a_{ij}]_{i=1}^n]_{j=1}^m \in \mathbb{R}^{nm}$. The problem can be formulated as an SIVI in (1) where the constraint set is described by $X = \{x \in \mathbb{R}^{m+n} \mid Lx \leq b,\ F^{\min} \leq x \leq F^{\max}\}$ which restricts the total resource consumption for some $L \in \mathbb{R}^{q \times (m+n)}$, $b \in \mathbb{R}^q$, and $q > 0$, moreover, the map F is described by a linear model based on an optimal unit of shipments $\mathbf{a}^*$. Different from He et al. (2010), here we consider a stochastic linear model to include uncertainty. In particular, let $G(x,\xi) = [(A\mathbf{a}^*(x) + \xi_1)^\top,\ (B\mathbf{a}^*(x) + \xi_2)^\top]^\top$ where $\xi_1 \in \mathbb{R}^m$, $\xi_2 \in \mathbb{R}^n$ are random variables with standard Gaussian distribution, and

$$A \triangleq \begin{bmatrix} e_n^T & 0 & \cdots & 0 \\ 0 & e_n^T & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & e_n^T \end{bmatrix} \in \mathbb{R}^{m*mn}, \qquad B \triangleq [I_n, \ldots, I_n] \in \mathbb{R}^{n*mn}.$$

For a given tax policy x, $\mathbf{a}^*(x)$ can be obtained by solving the following equilibrium problem

$$\langle \mathbf{a}' - \mathbf{a}, \ell(\mathbf{a}) + A^T(g(\mathbf{a}) + (\alpha + x_1)) - B^T(h(\mathbf{a}) - (\beta + x_2)) \rangle \geq 0, \quad \forall \mathbf{a}' \geq 0.$$

where $\alpha \in \mathbb{R}^m$ is the resource tax imposed on the supply market, $\beta \in \mathbb{R}^n$ is the sale tax imposed in the demand market, $\ell(x) \triangleq \mathrm{diag}(c)x + \tau$, $g(\mathbf{a}) \triangleq \mathrm{diag}(a)A\mathbf{a} + a^0$, and $h(\mathbf{a}) \triangleq \mathrm{diag}(\rho)B\mathbf{a} + \rho^0$ for some $c \in \mathbb{R}^{nm}$, $a, a^0 \in \mathbb{R}^m$ and $\rho, \rho^0 \in \mathbb{R}^n$. In this experiment, similar to He et al. (2010) we set $m = 10$, $n = 30$, $q = 2$, $F^{\min} = [\mathbf{0}_m^\top\ 20e_n^\top]^\top$, $F^{\max} = [160e_m^\top\ 60e_n^\top]^\top$, and other problem's parameter are generated uniformly at random according to Table 1.

Table 1: Distribution of randomly generated problem's parameters.

Parameter	c	τ	a	a^0	ρ	ρ^0
Distribution	$U[0.1, 0.2]$	$U[1,2]$	$U[1,2]$	$U[270, 370]$	$U[1,2]$	$U[620, 720]$

We execute the code 20 times with the same initial point $x_0 = \mathbf{0}$, and visualize the mean of replications. Figure 2 (left) depicts the convergence of VR-IPG for varying step sizes when the sample size $N_k = (k+1)^{2+2\delta}$ and $\delta = 1/2$. In addition, Figure 2 (right) depicts the results of implementing VR-IPG for a variety of N_k
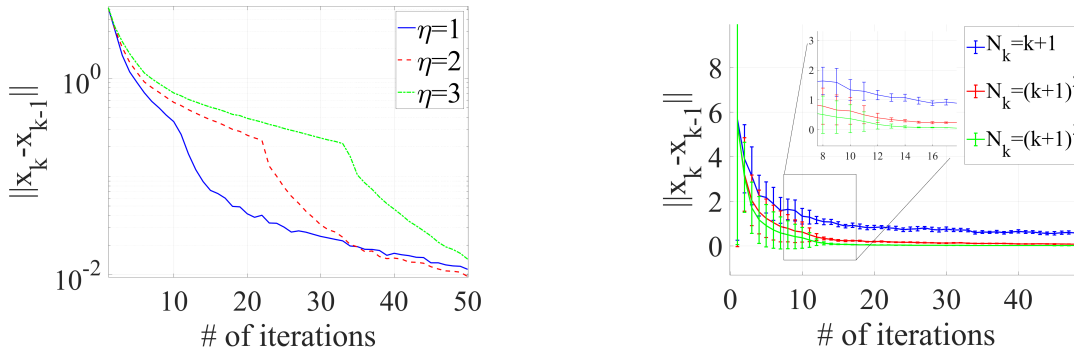

Figure 2: Convergence of VR-IPG implemented on Example 2 for different choices of stepsize (left); Almost sure convergence of VR-IPG for different sample sizes (right).

values, along with the mean of 20 replications, with stepsize equal to 1 and their respective 95% confidence intervals. We observe that as k goes to infinity, the confidence intervals become tighter.

A   APPENDICES

In this section, proof of Lemma 3 and Lemma 4 are provided.

Proof of Lemma 3. Since x^* is a solution and $\mathbf{P}_X(F(x) - \eta x) \in X$ by using definition of IVI we have:

$$\langle \eta x^*, \mathbf{P}_X(F(x) - \eta x) - F(x^*)\rangle \geq 0. \tag{5}$$

Choosing $u = F(x) - \eta x$ in Lemma 1. Since $F(x^*) \in X$ then following holds:

$$\langle (F(x) - \eta x) - \mathbf{P}_X(F(x) - \eta x), \mathbf{P}_X(F(x) - \eta x) - F(x^*)\rangle \geq 0. \tag{6}$$

Combining (5) and (6), we get :

$$\langle \eta H(x, \eta) - \eta(x - x^*), (F(x) - F(x^*)) - \eta H(x, \eta)\rangle \geq 0.$$

and consequently

$$(x - x^*)^T H(x, \eta) \geq \|H(x, \eta)\|^2 + \tfrac{1}{\eta}(x - x^*)^T(F(x) - F(x^*)) - \tfrac{1}{\eta}(F(x) - F(x^*))^T H(x, \eta).$$

Applying the co-coercivity of mapping F, we obtain:

$$(x - x^*)^T H(x, \eta) \geq \|H(x, \eta)\|^2 + \tfrac{\mu}{\eta}\|F(x) - F(x^*)\|^2 - \tfrac{1}{\eta}(F(x) - F(x^*))^T H(x, \eta)$$
$$= \|H(x, \eta)\|^2 - \tfrac{1}{4\eta\mu}\|H(x, \eta)\|^2 + \|\sqrt{\tfrac{\mu}{\eta}}(F(x) - F(x^*)) - \tfrac{1}{2\sqrt{\mu\eta}}H(x, \eta)\|^2$$
$$\geq (1 - \tfrac{1}{4\eta\mu})\|H(x, \eta)\|^2. \qquad \square$$

Proof of Lemma 4. Recall that $\bar{w}_{k,N_k} = \frac{1}{N_k}\sum_{j=1}^{N_k}(G(x_k, \xi_{j,k}) - F(x_k))$. Using the update rule of x_{k+1} in Algorithm 1, one can obtain the following.

$$\|x_{k+1} - x^*\|^2$$
$$= \|x_k - x^* - \tfrac{1}{\eta_k}(F(x_k) + \bar{w}_{k,N_k} - \mathbf{P}_X(F(x_k) + \bar{w}_{k,N_k} - \eta_k x_k))\|^2$$
$$= \|x_k - x^* - \tfrac{1}{\eta_k}(F(x_k) - \mathbf{P}_X(F(x_k) - \eta_k x_k)) - \tfrac{1}{\eta_k}(\bar{w}_{k,N_k} - \mathbf{P}_X(F(x_k) + \bar{w}_{k,N_k} - \eta_k x_k) + \mathbf{P}_X(F(x_k) - \eta_k x_k))\|^2.$$

For notational simplicity, define $e_k \triangleq \bar{w}_{k,N_k} - \mathbf{P}_X(F(x_k) + \bar{w}_{k,N_k} - \eta_k x_k) + \mathbf{P}_X(F(x_k) - \eta_k x_k)$ and $H(x_k, \eta_k) \triangleq \tfrac{1}{\eta_k}(F(x_k) - \mathbf{P}_X(F(x_k) - \eta_k x_k))$. Using the fact that $(a - b - c)^2 = a^2 + b^2 - 2ab + c^2 - 2ac + 2bc$, we obtain

$$\|x_{k+1} - x^*\|^2$$
$$\leq \|x_k - x^*\|^2 + \|H(x_k, \eta_k)\|^2 - 2\langle x_k - x^*, H(x_k, \eta_k)\rangle + \frac{1}{\eta_k^2}\|e_k\|^2 - \tfrac{2}{\eta_k}\langle x_k - x^*, e_k\rangle + \tfrac{2}{\eta_k}\langle H(x_k, \eta_k), e_k\rangle. \tag{7}$$

Using the definition of e_k and Lemma 1, we have that

$$\tfrac{1}{\eta_k^2}\|e_k\|^2 \leq \tfrac{2}{\eta_k^2}\|\bar{w}_{k,N_k}\|^2 + \tfrac{2}{\eta_k^2}\|\mathbf{P}_X(F(x_k) + \bar{w}_{k,N_k} - \eta_k x_k) - \mathbf{P}_X(F(x_k) - \eta_k x_k)\|^2 \leq \tfrac{4}{\eta_k^2}\|\bar{w}_{k,N_k}\|^2. \tag{8}$$

Moreover, from definition of e_k and $H(x_k, \eta_k)$, one can show that

$$\tfrac{2}{\eta_k}\langle H(x_k, \eta_k) - (x_k - x^*), e_k\rangle$$
$$= \tfrac{2}{\eta_k}\langle H(x_k, \eta_k) - (x_k - x^*), \bar{w}_{k,N_k}\rangle$$
$$\quad - \langle \tfrac{2}{\eta_k}(H(x_k, \eta_k) - (x_k - x^*)), \mathbf{P}_X(F(x_k) + \bar{w}_{k,N_k} - \eta_k x_k) - \mathbf{P}_X(F(x_k) - \eta_k x_k)\rangle$$
$$= \tfrac{2}{\eta_k}\langle H(x_k, \eta_k) - (x_k - x^*), \bar{w}_{k,N_k}\rangle$$

$$-\left\langle \tfrac{2}{\eta_k^2}\left((F(x_k)-\eta_k x_k)+\eta_k x^*-\mathbf{P}_X(F(x_k)-\eta_k x_k)\right),\mathbf{P}_X\left(F(x_k)+\bar{w}_{k,N_k}-\eta_k x_k\right)-\mathbf{P}_X\left(F(x_k)-\eta_k x_k\right)\right\rangle,$$

By adding and subtracting $\bar{w}_{k,N_k}$ and $\mathbf{P}_X\left(F(x_k)+\bar{w}_{k,N_k}-\eta_k x_k\right)$, we obtain:

$$\tfrac{2}{\eta_k}\langle H(x_k,\eta_k)-(x_k-x^*),e_k\rangle$$
$$=\tfrac{2}{\eta_k}\langle H(x_k,\eta_k)-(x_k-x^*),\bar{w}_{k,N_k}\rangle$$
$$\underbrace{-\tfrac{2}{\eta_k^2}\langle(F(x_k)+\bar{w}_{k,N_k}-\eta_k x_k)-\mathbf{P}_X\left(F(x_k)+\bar{w}_{k,N_k}-\eta_k x_k\right),\mathbf{P}_X\left(F(x_k)+\bar{w}_{k,N_k}-\eta_k x_k\right)-\mathbf{P}_X\left(F(x_k)-\eta_k x_k\right)\rangle}_{\text{term (a)}}$$
$$-\tfrac{2}{\eta_k^2}\langle\mathbf{P}_X\left(F(x_k)+\bar{w}_{k,N_k}-\eta_k x_k\right)-\mathbf{P}_X\left(F(x_k)-\eta_k x_k\right),\mathbf{P}_X\left(F(x_k)+\bar{w}_{k,N_k}-\eta_k x_k\right)-\mathbf{P}_X\left(F(x_k)-\eta_k x_k\right)\rangle$$
$$+\tfrac{2}{\eta_k^2}\langle\bar{w}_{k,N_k},\mathbf{P}_X\left(F(x_k)+\bar{w}_{k,N_k}-\eta_k x_k\right)-\mathbf{P}_X\left(F(x_k)-\eta_k x_k\right)\rangle$$
$$-\tfrac{2}{\eta_k}\langle x^*,\mathbf{P}_X\left(F(x_k)+\bar{w}_{k,N_k}-\eta_k x_k\right)-\mathbf{P}_X\left(F(x_k)-\eta_k x_k\right)\rangle,$$

Based on Lemma 1 part (b), one can easily verify that term (a) ≤ 0, then the following can be obtained:

$$\tfrac{2}{\eta_k}\langle H(x_k,\eta_k)-(x_k-x^*),e_k\rangle$$
$$\leq\tfrac{2}{\eta_k}\langle H(x_k,\eta_k)-(x_k-x^*),\bar{w}_{k,N_k}\rangle-\tfrac{2}{\eta_k}\langle x^*,\mathbf{P}_X\left(F(x_k)+\bar{w}_{k,N_k}-\eta_k x_k\right)-\mathbf{P}_X\left(F(x_k)-\eta_k x_k\right)\rangle$$
$$-\tfrac{2}{\eta_k^2}\langle\mathbf{P}_X\left(F(x_k)+\bar{w}_{k,N_k}-\eta_k x_k\right)-\mathbf{P}_X\left(F(x_k)-\eta_k x_k\right),\mathbf{P}_X\left(F(x_k)+\bar{w}_{k,N_k}-\eta_k x_k\right)-\mathbf{P}_X\left(F(x_k)-\eta_k x_k\right)\rangle$$
$$+\tfrac{2}{\eta_k^2}\langle\bar{w}_{k,N_k},\mathbf{P}_X\left(F(x_k)+\bar{w}_{k,N_k}-\eta_k x_k\right)-\mathbf{P}_X\left(F(x_k)-\eta_k x_k\right)\rangle$$
$$=\tfrac{2}{\eta_k}\langle H(x_k,\eta_k)-(x_k-x^*),\bar{w}_{k,N_k}\rangle-\tfrac{2}{\eta_k}\langle x^*,\mathbf{P}_X\left(F(x_k)+\bar{w}_{k,N_k}-\eta_k x_k\right)-\mathbf{P}_X\left(F(x_k)-\eta_k x_k\right)\rangle$$
$$\underbrace{-\tfrac{2}{\eta_k^2}\|\mathbf{P}_X\left(F(x_k)+\bar{w}_{k,N_k}-\eta_k x_k\right)-\mathbf{P}_X(F(x_k)-\eta_k x_k)\|^2}_{\text{term (b)}}$$
$$+\tfrac{2}{\eta_k^2}\langle\bar{w}_{k,N_k},\mathbf{P}_X\left(F(x_k)+\bar{w}_{k,N_k}-\eta_k x_k\right)-\mathbf{P}_X\left(F(x_k)-\eta_k x_k\right)\rangle$$
$$\leq\tfrac{2}{\eta_k}\langle H(x_k,\eta_k)-(x_k-x^*),\bar{w}_{k,N_k}\rangle+\tfrac{2}{\eta_k}\|x^*\|\|\bar{w}_{k,N_k}\|+\tfrac{2}{\eta_k^2}\|\bar{w}_{k,N_k}\|^2, \tag{9}$$

where in the last two inequalities we used the fact that term (b) ≤ 0, Lemma 1 and Cauchy-Schwartz inequality. Using Lemma 3, inequalities (8) and (9) in (7), we obtain

$$\|x_{k+1}-x^*\|^2\leq\|x_k-x^*\|^2-\left(1-\tfrac{1}{2\eta_k\mu}\right)\|H(x_k,\eta_k)\|^2+\tfrac{2}{\eta_k}\langle H(x_k,\eta_k)-(x_k-x^*),\bar{w}_{k,N_k}\rangle$$
$$+\tfrac{6}{\eta_k^2}\|\bar{w}_{k,N_k}\|^2+\tfrac{2}{\eta_k}\|x^*\|\|\bar{w}_{k,N_k}\|.$$

Taking conditional expectation and using Assumption 2, we have the following.

$$\mathbb{E}[\|x_{k+1}-x^*\|^2\mid\mathscr{F}_k]\leq\|x_k-x^*\|^2-\left(1-\tfrac{1}{2\eta_k\mu}\right)\|H(x_k,\eta_k)\|^2+\tfrac{6\nu^2}{\eta_k^2 N_k}+\tfrac{2\nu}{\eta_k\sqrt{N_K}}\|x^*\|. \quad\square$$

REFERENCES

Bertsekas, D., A. Nedić, and A. Ozdaglar. 2003. *Convex Analysis and Optimization, Ser*, Volume 1. Belmont, Massachusetts: Athena Scientific.

Facchinei, F., and J.-S. Pang. 2007. *Finite-dimensional Variational Inequalities and Complementarity Problems*. Berlin/Heidelberg, Germany: Springer Science & Business Media.

He, B., X.-Z. He, and H. X. Liu. 2010. "Solving a Class of Constrained 'Black-Box' Inverse Variational Inequalities". *European Journal of Operational Research* 204(3):391–401.

He, B.-s. 1999. "A Goldstein's Type Projection Method for a Class of Variant Variational Inequalities". *Journal of Computational Mathematics* 17(4):425–434.

He, S., and Q.-L. Dong. 2018. "An Existence-Uniqueness Theorem and Alternating Contraction Projection Methods for Inverse Variational Inequalities". *Journal of Inequalities and Applications* 2018(1):1–19.

He, X., and H. X. Liu. 2011. "Inverse Variational Inequalities with Projection-Based Solution Methods". *European Journal of Operational Research* 208(1):12–18.

Jalilzadeh, A., and U. V. Shanbhag. 2019. "A Proximal-point Algorithm with Variable Sample-sizes (PPAWSS) for Monotone Stochastic Variational Inequality Problems". In *Proceedings of the 2019 Winter Simulation Conference (WSC)*, edited by N. Mustafee, K.-H. G. Bae, S. Lazarova-Molnar, M. Rabe, C. Szabo, P. Haas, and Y.-J. Son, 3551–3562. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

Jiang, H., and H. Xu. 2008. "Stochastic Approximation Approaches to the Stochastic Variational Inequality Problem". *Institute of Electrical and Electronics Engineers Transactions on Automatic Control* 53(6):1462–1475.

Koshal, J., A. Nedić, and U. V. Shanbhag. 2012. "Regularized Iterative Stochastic Approximation Methods for Stochastic Variational Inequality Problems". *Institute of Electrical and Electronics Engineers Transactions on Automatic Control* 58(3):594–609.

Luo, X.-P., and J. Yang. 2014. "Regularization and Iterative Methods for Monotone Inverse Variational Inequalities". *Optimization Letters* 8(4):1261–1272.

Malitsky, Y. 2015. "Projected Reflected Gradient Methods for Monotone Variational Inequalities". *Society for Industrial and Applied Mathematics Journal on Optimization* 25(1):502–520.

Nesterov, Y., and L. Scrimali. 2011. "Solving strongly monotone variational and quasi-variational inequalities". *Discrete and Continuous Dynamical Systems* 31(4):1383–1396.

Polyak, B. T. 2003. *Introduction to Optimization*, Volume 1. New York: Optimization Software, Inc.

Scrimali, L. 2012. "An Inverse Variational Inequality Approach to the Evolutionary Spatial Price Equilibrium Problem". *Optimization and Engineering* 13:375–387.

Yang, H., and M. G. Bell. 1997. "Traffic Restraint, Road Pricing and Network Equilibrium". *Transportation Research Part B: Methodological* 31(4):303–314.

Yousefian, F., A. Nedić, and U. V. Shanbhag. 2017. "On Smoothing, Regularization, and Averaging in Stochastic Approximation Methods for Stochastic Variational Inequality Problems". *Mathematical Programming* 165:391–431.

Zhu, D. L., and P. Marcotte. 1996. "Co-coercivity and its Role in the Convergence of Iterative Schemes for Solving Variational Inequalities". *Society for Industrial and Applied Mathematics Journal on Optimization* 6(3):714–726.

Zou, X., D. Gong, L. Wang, and Z. Chen. 2016. "A Novel Method to Solve Inverse Variational Inequality Problems Based on Neural Networks". *Neurocomputing* 173:1163–1168.

AUTHOR BIOGRAPHIES

ZEINAB ALIZADEH is a Ph.D. student in Systems and Industrial Engineering at the University of Arizona. She received her bachelor's degree in Mathematics from the University of Tehran. Her research interests include stochastic optimization, variational inequality problems, and data analytics. Her email address is zalizadeh@arizona.edu.

FELIPE PARRA POLANCO is an undergraduate student in Systems and Industrial Engineering at the University of Arizona. His research interests include smart transportation, data analytics, stochastic optimization, and machine learning. His email address is fparrapolanco@arizona.edu.

AFROOZ JALILZADEH is an assistant professor in Systems and Industrial Engineering at the University of Arizona. She holds a Ph.D. in Industrial Engineering and Operations Research from Pennsylvania State University. Her research is focused on the design, analysis, and implementation of stochastic approximation methods for solving convex optimization and stochastic variational inequality problems with applications in machine learning, game theory, healthcare, and power systems. Her website is https://profiles.arizona.edu/person/afrooz and her email address is afrooz@arizona.edu.