# Regularized Iterative Stochastic Approximation Methods for Variational Inequality Problems

Jayash Koshal, Angelia Nedić and Uday V. Shanbhag

**Abstract**

We consider a Cartesian stochastic variational inequality problem with a monotone map. For this problem, we develop and analyze distributed iterative stochastic approximation algorithms. Such a problem arises, for example, as an equilibrium problem in monotone stochastic Nash games over continuous strategy sets. We introduce two classes of *stochastic approximation methods*, each of which requires exactly one projection step at every iteration, and we provide convergence analysis for them. Of these, the first is the *stochastic iterative Tikhonov regularization method* which necessitates the update of regularization parameter after *every* iteration. The second method is a *stochastic iterative proximal-point method*, where the centering term is updated after *every* iteration. Notably, we present a generalization of this method where the weighting in the proximal-point method can also be updated after every iteration. Conditions are provided for recovering global convergence in limited coordination extensions of such schemes where players are allowed to choose their steplength sequences independently, while meeting a suitable coordination requirement. We apply the proposed class of techniques and their limited coordination versions to stochastic networked rate allocation problem.

## I. Introduction

In this paper, we consider a Cartesian monotone stochastic variational inequality problem, which is a problem that requires finding a vector $x = (x_1, \ldots, x_N)$ satisfying

$$\begin{aligned} (x_1 - y_1)^T \mathbb{E}[f_1(x,\xi)] &\geq 0, \quad \forall y_1 \in K_1, \\ &\vdots \\ (x_N - y_N)^T \mathbb{E}[f_N(x,\xi)] &\geq 0, \quad \forall y_N \in K_N, \end{aligned} \tag{1}$$

where $\mathbb{E}$ denotes the expectation with respect to some uncertainty $\xi$. Our interest is in developing *distributed* stochastic approximation schemes for such a problem when the mapping

$$F(x) = (\mathbb{E}[f_1(x,\xi)], \ldots, \mathbb{E}[f_N(x,\xi)])$$

Department of Industrial and Enterprise Systems Engineering, University of Illinois, Urbana IL 61801, Email: {koshal1,angelia,udaybag}@illinois.edu. This work has been supported by NSF CMMI 09-48905 EAGER ARRA award. A subset of this work has appeared in [1]

is monotone[1] over the Cartesian product of the sets $K_i$, i.e., $(F(x)-F(y))^T(x-y) \geq 0$ for all $x,y \in K_1 \times \cdots \times K_N$.

Cartesian stochastic variational inequalities arise from both stochastic optimization and game-theoretic problems, as discussed next.

*a) Multiuser stochastic optimization problems:* Consider the following problem:

$$\begin{aligned} &\text{minimize} && f(x) \triangleq \sum_{i=1}^{N} \mathbb{E}[f_i(x_i, x_{-i}, \xi_i)] \\ &\text{subject to} && x_i \in K_i, \quad i = 1, \ldots, N, \end{aligned} \tag{2}$$

where $x_{-i}$ denotes the sub-vector of $x$ that does not contain the component $x_i$. The problem is associated with a system of users, where $f_i$ and $K_i$ are the objective function and constraint set of user $i$. When $f(x)$ is convex over $K_1 \times \cdots \times K_N$, the global optimality conditions of (2) are given by (1) where $F_i(x) = \nabla_{x_i} f(x)$ for $i = 1, \ldots, N$, and the resulting mapping $F(x)$ is monotone over $K_1 \times \cdots \times K_N$ **CITE**. Often, the informational restrictions dictate that the $i$th user only has access to his objective $f_i$ and constraint set $K_i$. Consequently, distributed schemes that abide by these requirements are of relevance. Multiuser optimization problems have been studied extensively in context of network resource management, such as rate allocation in communication networks [2], [3], [4]. Notably, such problems do not allow for interactions in the objective but introduce a coupling across the user decisions. By using a Lagrangian relaxation, in [1], primal-dual schemes for such problems have been considered in a deterministic regime.

*b) Continuous-strategy Nash games:* Consider a Nash game in which the $i$th player solves the following problem:

$$\begin{aligned} &\text{minimize} && \mathbb{E}[f_i(x_i, x_{-i}, \xi_i)] \\ &\text{subject to} && x_i \in K_i. \end{aligned} \tag{3}$$

When $f_i(x)$ is convex over $K_i$ for all $i = 1, \ldots, N$, then the equilibrium (sufficient) conditions of this problem are given by (1) where $F_i(x) = \nabla_{x_i} \mathbb{E}[f_i(x, \xi)]$ for $i = 1, \ldots, N$, and the resulting mapping $F$ is monotone over $K_1 \times \cdots \times K_N$. While Nash games may be natural models for capturing the strategic behavior of the multiuser system, competitive counterparts of multiuser optimization problems are effectively Nash games. Game-theoretic models find application in a range of settings ranging from wireline and wireless communication networks [5], bandwidth allocation [6], [7], [8], spectrum allocation in radio networks [9] and optical networks [10], [11]. In fact, in many of the above mentioned applications, monotonicity of the mappings is seen to be a natural outcome of the model assumptions including communication networks [8], [12], cognitive radio networks [13], **Nash-Cournot games [?]** REF MISSING and optical networks [10], [11].

[1] A mapping $F: K \to \mathbb{R}^n$ is said to be monotone over a set $K \subset \mathbb{R}^n$ if $(F(x)-F(y))^T(x-y) \geq 0$ for all $x,y \in K$. It is said to be strictly monotone over $K$ if $(F(x)-F(y))^T(x-y) > 0$ for all $x,y \in K$ with $x \neq y$. In addition, it is said to be strongly monotone if there exists a positive scalar $\eta$ such that $(F(x)-F(y))^T(x-y) \geq \eta \|x-y\|^2$ for all $x,y \in K$.

For the Cartesian stochastic variational inequalities, it seems natural to exploit the presence of decoupled constraint sets and distribute the computations of a solution **to an extent possible** CHECK WORDING. When considering multiuser optimization problems, distributed optimization approaches have been natural candidates (cf. [3]). Yet, there appears to have been markedly little on stochastic variational inequalities that naturally extend multiuser optimization. Our work in this paper intends to fill the lacuna through this framework.

In a deterministic setting, distributed schemes for computing equilibria have received significant attention recently [8], [6], [14], [15], [11]. Of particular relevance to this paper is the work in [11], [15], the latter employing an extragradient scheme [16] capable of accommodating deterministic monotone Nash games. Finally, Scutari et al. [13] examine an array of monotone Nash games and consider proximal-point based distributed schemes in a basic prox-setting where a sub-problem is solved at each iteration.

In game-theoretic regimes, best-response schemes are often the de-facto choice under an assumption of full rationality [13]. In such schemes, every player takes the best response, given the decisions of his adversaries. In many settings, the best-response step requires solving a large-scale optimization problem and this necessitates that players possess the computational ability and time to obtain such solutions. We consider a *bounded rationality* where players rule out strategies of high complexity [17]. This notion is rooted in the influential work by Simon [18] where it was suggested that if reasoning and computation were considered to be costly, then agents may not invest in these resources for marginal benefits. Similarly, by noting that solving a convex program may be computationally burdensome for the agents, we instead allow for agents to compute low-overhead gradient steps. We provide mathematical substantiation for claiming that if players independently operated in this regime, convergence of the overall scheme ensures. While gradient-response techniques are not necessarily superior from the standpoint of convergence rate, they remain advantageous from the standpoint of implementability.

Within the framework of stochastic variational inequalities, stochastic approximation schemes have been recently employed in [19] under somewhat stringent assumptions on the mapping (strongly monotone and co-coercive maps, see [16] for a definition). To the best of our knowledge, [19] appears to be the only existing work considering stochastic approximation methods for variational inequalities. However, the work in [19] is more along the lines of standard stochastic-gradient approximation for strongly convex problems. In contrast, our work builds on different deterministic algorithms, including Tikhonov regularization and proximal-point methods, and combines them with the stochastic approximation approach. At the same time, our convergence results require less stringent monotonicity assumption on the map.

The typical stochastic approximation procedure, first introduced by Robbins and Monro [20], works toward finding an extremum of a function $h(x)$ using the following iterative scheme:

$$x^{k+1} = x^k + a_k(\nabla h(x^k) + M^{k+1}),$$

where $M^{k+1}$ is a martingale difference sequence. Under reasonable assumptions on the stochastic errors $M^k$, stochastic approximation schemes ensure that $\{x^k\}$ converges almost surely to an optimal solution of the problem. Jiang and Xu [19] consider the use of stochastic approximation for strongly monotone and Lipschitz continuous maps in the realm of stochastic variational inequalities, rather than optimization problems. The use of stochastic approximation methods has a long tradition in stochastic optimization for both differentiable and nondifferentiable problems, starting with the work of Robbins and Monro [20] for differentiable problems and Ermoliev [21], [22], [23] for nondifferentiable problems while a subset of more recent efforts include [24], [25], [26].

An alternative approach towards the solution of stochastic optimization problems is through sample-average approximation (SAA) methods [27], [28], [29], while a short discussion of SAA techniques to the stochastic VI is provided in [27]. More recently, Xu [30] applied an SAA approach to stochastic variational inequalities problems where the existence of a solution of the sample average approximated stochastic VI is established almost surely. Exponential convergence of a solution to its true counterpart is established under moderate conditions. Asymptotics of the associated estimators has been studied extensively over the years, with early work attributed to King and Rockafellar [31], while Monte-Carlo bounding approaches have been studied by Mak et al. [32]. A related sample-path approach for solving stochastic variational problems can be found in [33] while matrix-splitting schemes for addressing stochastic complementarity problems is examined in [34].

This paper is motivated by the challenges associated with solving stochastic variational problems when the mappings lose strong monotonicity. In solving deterministic variational inequalities, such a departure is ably handled through techniques, such as Tikhonov regularization [35], [36] and proximal-point [37]. It should be remarked that Tikhonov-based regularization methods have a long history in the solution of ill-posed optimization and variational problems [38], [16]. Such schemes, in general, require a solution of a regularized (well-posed) problem and an iterative process is often needed to obtain the solution. While Tikhonov regularization approach provides a framework for solving monotone games, the approach suffers from a potential drawback due to the computational aspect. The complexity of solving ill-posed problem increases as the regularization parameter decays to zero. Proximal point method proposed by Martinet [37] alleviates this problem (cf. [16]). Rockafellar investigated this scheme in the context of maximal monotone operators [39] and established convergence of some variants of this scheme. Proximal point methods have gained attention from various application areas communication network being one of them. More recently, the work of Nesterov [40] and Nemirovski [41] provides error bounds for Lipschitzian monotone operators and compact sets. They show that a variation of the standard proximal-point scheme leads to convergent algorithms that require the number of iterations in the order of $O(1/\varepsilon)$ to reach an error level of $\varepsilon$. Importantly, while the schemes require two nested iterative methods, the inner level can be shown to be solvable in a finite number of iterations.

In both Tikhonov regularization and proximal point methods, two nested iterative procedures are involved, where the outer procedure updates a parameter after an increasingly accurate solution of an inner subproblem is available. In networked stochastic regimes, this is challenging for two reasons: (1) First, obtaining increasingly accurate solutions of stochastic variational problems via simulation techniques requires significant effort; and (2) Assessing solution quality formally requires validation analysis that needs to be conducted over the network, a somewhat challenging task. We obviate this challenge by considering algorithms where the parameter is updated after every iteration and, thus, the update of the steplength and the regularization parameter is synchronized. Such an approach is popularly referred to as *iterative regularization.* While there have been efforts to use such techniques for optimization problems (cf. [38]), there has been noticeably less in the realm of variational inequalities, barring [42] and more recently in [12], [43]. However, all this work has been restricted to deterministic regimes. Accordingly, the present work emphasizes iterative regularization for stochastic variational inequalities with monotone maps. We present and analyze two stochastic iterative regularization schemes:

1) *Stochastic iterative Tikhonov regularization method:* We consider a stochastic iterative Tikhonov regularization method for monotone stochastic variational inequalities where the steplength and regularization parameter are updated at every iteration. Partially coordinated method is presented where users independently select stepsize and regularization sequences. Under some restrictions on the deviations across the users' choices, we establish convergence properties of the method in almost sure sense.
2) *Stochastic iterative proximal-point method:* An alternative to the stochastic iterative Tikhonov method lies in a stochastic iterative proximal-point method where the prox-parameter is updated at *every iteration*. As in the case of iterative Tikhonov method, we present convergence results for partially coordinated implementation. Our convergence results are established for strictly monotone mappings.

The rest of the paper is organized as follows. In Section II, we discuss an $N$-person stochastic Nash game and its associated equilibrium conditions. In Section III, we propose and analyze a stochastic iterative Tikhonov regularization method. Analogous results for a stochastic iterative proximal point scheme are provided in Section IV. In Section V, the performance of our schemes and their relative sensitivity to parameters is examined in the context of networked rate allocation game. We conclude the paper with some remarks in Section VI.

Throughout this paper, we view vectors as columns. We write $x^T$ to denote the transpose of a vector $x$, and $x^T y$ to denote the inner product of vectors $x$ and $y$. We use $\|x\|$ to denote the Euclidean norm of a vector $x$, i.e., $\|x\| = \sqrt{x^T x}$. We use $\Pi_X$ to denote the Euclidean projection operator onto a set $X$, i.e., $\Pi_X(x) \triangleq \text{argmin}_{z\in X} \|x-z\|$. We write $\prod_{i=1}^N K_i$ to denote the Cartesian product of sets $K_1,\ldots,K_N$. We use $\text{VI}(K,F)$ to denote a variational inequality problem specified by a set $K$ and a mapping $F$. Also, we use $\text{SOL}(K,F)$ to denote the solution set for a given $\text{VI}(K,F)$. The expectation of a random variable $V$ is denoted by $\mathbb{E}[V]$ and *a.s.* is used for *almost surely*.

## II. Problem Description

We consider an $N$-person stochastic Nash game in which the $i$th agent solves the parameterized problem

$$\begin{aligned} &\text{minimize} && \mathbb{E}[f_i(x_i,x_{-i},\xi_i)] \\ &\text{subject to} && x_i \in K_i, \end{aligned} \tag{4}$$

where $x_{-i}$ denotes the collection $\{x_j, j\neq i\}$ of decisions of all players $j$ other than player $i$. For each $i$, the variable $\xi_i$ is random with $\xi_i:\Omega_i \to \mathbb{R}^{n_i}$, and the function $\mathbb{E}[f_i(x_i,x_{-i},\xi_i)]$ is convex in $x_i$ for all

$x_{-i} \in \prod_{j \neq i} K_j$. For every $i$, the set $K_i \subseteq \mathbb{R}^{n_i}$ is a closed convex set.

The equilibrium conditions of the game in (4) can be characterized by a variational inequality problem, denoted by $\mathrm{VI}(K,F)$, where the mapping $F: K \to \mathbb{R}^n$ and the set $K$ are given by

$$F(x) \triangleq \begin{pmatrix} \nabla_{x_1} \mathbb{E}[f_1(x,\xi_1)] \\ \vdots \\ \nabla_{x_N} \mathbb{E}[f_N(x,\xi_N)] \end{pmatrix}, \qquad K = \prod_{i=1}^{N} K_i, \tag{5}$$

with $x \triangleq (x_1,\ldots,x_N)^T$ and $x_i \in K_i$ for $i = 1,\ldots,N$. We let $n = \sum_{i=1}^{N} n_i$, and note that the set $K$ is closed and convex set in $\mathbb{R}^n$, whenever the sets $K_i$ are closed and convex. Recall that $\mathrm{VI}(K,F)$ requires determining a vector $x^* \in K$ such that

$$(x - x^*)^T F(x^*) \geq 0 \quad \text{for all } x \in K. \tag{6}$$

Standard deterministic algorithms for obtaining solutions to a variational inequality $\mathrm{VI}(K,F)$ require an analytical form for the gradient of the expected-value function. Yet, when the expectation is over a general measure space, analytical forms of the expectation are often hard to obtain. In such settings, stochastic approximation schemes assume relevance. In the remainder of this section, we describe the basic framework of stochastic approximation and the supporting convergence results.

Consider the Robbins-Monro stochastic approximation scheme for solving the stochastic variational inequality $\mathrm{VI}(K,F)$ in (5)–(6), given by

$$x^{k+1} = \Pi_K[x^k - \alpha_k(F(x^k) + w^k)] \quad \text{for } k \geq 0, \tag{7}$$

where $x^0 \in K$ is a random initial vector that is independent of the random variables $\xi_i$ for all $i$ and such that $\mathbb{E}[\|x^0\|^2]$ is finite. The vector $F(x^k)$ is the true value of $F(x)$ at $x = x^k$, $\alpha_k > 0$ is the stepsize, while the vector $w^k$ is the stochastic error given by

$$w^k = -F(x^k) + \tilde{F}(x^k,\xi^k),$$

with

$$\tilde{F}(x^k,\xi^k) \triangleq \begin{pmatrix} \nabla_{x_1} f_1(x^k,\xi_1^k) \\ \vdots \\ \nabla_{x_N} f_N(x^k,\xi_N^k) \end{pmatrix} \qquad \text{and} \qquad \xi^k \triangleq \begin{pmatrix} \xi_1^k \\ \vdots \\ \xi_N^k \end{pmatrix}.$$

The projection scheme (7) is shown to be convergent when the mapping $F$ is strongly monotone and Lipschitz continuous in [19]. In the sequel, we examine how the use of regularization methods can alleviate the strong monotonicity requirement.

In our analysis we use some well-known results on supermartingale convergence, which we provide for convenience. The following result is from [38], Lemma 10, page 49.

**Lemma 1.** *Let $V_k$ be a sequence of non-negative random variables adapted to $\sigma$-algebra $\mathcal{F}_k$ and such that almost surely*

$$\mathbb{E}[V_{k+1} \mid \mathcal{F}_k] \leq (1 - u_k)V_k + \beta_k \qquad \textit{for all } k \geq 0,$$

*where* $0 \le u_k \le 1$, $\beta_k \ge 0$, *and*

$$\sum_{k=0}^{\infty} u_k = \infty, \quad \sum_{k=0}^{\infty} \beta_k < \infty, \quad \lim_{k\to\infty} \frac{\beta_k}{u_k} \to 0.$$

*Then,* $V_k \to 0$ *a.s.*

The result of the following lemma can be found in [38], Lemma 11, page 50.

**Lemma 2.** *Let* $V_k, u_k, \beta_k$ *and* $\gamma_k$ *be non-negative random variables adapted to* $\sigma$*-algebra* $\mathcal{F}_k$. *If almost surely* $\sum_{k=0}^{\infty} u_k < \infty$, $\sum_{k=0}^{\infty} \beta_k < \infty$, *and*

$$\mathbb{E}[V_{k+1} \mid \mathcal{F}_k] \le (1+u_k)V_k - \gamma_k + \beta_k \quad \textit{for all } k \ge 0,$$

*then almost surely* $\{V_k\}$ *is convergent and* $\sum_{k=0}^{\infty} \gamma_k < \infty$.

## III. Stochastic Iterative Tikhonov Schemes

In this section, we propose and analyze a stochastic iterative Tikhonov algorithm for solving the variational inequality $\mathrm{VI}(K,F)$ in (5)–(6). We consider the case when the mapping $F$ is monotone over the set $K$, i.e., $F$ is such that

$$(F(x) - F(y))^T(x-y) \ge 0 \qquad \text{for all } x, y \in K.$$

A possible approach for addressing monotone variational problems is through a Tikhonov regularization scheme [16, Ch. 12] (cf. [36], [35]). In the context of variational inequalities, this avenue typically requires solving a sequence of variational inequality problems, denoted by $\mathrm{VI}(K, F+\varepsilon_k \mathbf{I})$. Each $\mathrm{VI}(K, F+\varepsilon_k \mathbf{I})$ is a perturbation of the original variational inequality $\mathrm{VI}(K,F)$ obtained by using a perturbed mapping $F+\varepsilon_k I$ instead of $F$ for some positive scalar $\varepsilon_k$. In this way, each of the variational inequality problems $\mathrm{VI}(K, F+\varepsilon_k \mathbf{I})$ is strongly monotone and, hence, it has a unique solution denoted by $y_k$, (see Theorem 2.3.3 in [16]). Under suitable conditions, it can be seen that the Tikhonov sequence $\{y_k\}$ satisfies

$$\lim_{k\to\infty} y^k = x^*,$$

where $x^*$ is the least norm solution of $\mathrm{VI}(K,F)$ (see Theorem 12.2.3 in [16]). Thus, to reach a solution of $\mathrm{VI}(K,F)$, one has to solve a sequence of variational inequality problems, namely $\mathrm{VI}(K, F+\varepsilon_k y)$ along some sequence $\{\varepsilon_k\}$. However, determining a solution $y^k$ for a regularized variational inequality $\mathrm{VI}(K, F+\varepsilon_k y)$ in the current setting, requires either the exact or approximate solution of a strongly monotone stochastic variational inequality (see Section 12.2 in [16]).

In deterministic regimes, a solution to the regularized Tikhonov subproblem may be obtained in a distributed fashion via a projection scheme. However, in stochastic regimes, this is a more challenging proposition. While an almost sure convergence theory for a stochastic approximation method for strongly monotone variational problems is provided in [19], termination criteria are generally much harder to provide. As a consequence, one often provides confidence intervals in practice by generating a fixed number of sample paths. Furthermore, the convergence theory of Tikhonov-based schemes necessitates that the solutions to the subproblem are computed with increasing accuracy. Implementing such algorithms in stochastic regime is significantly harder since simulation-based schemes are being employed for obtaining confidence intervals for each regularized problem and we require that these intervals get increasingly tighter. In the numerical results, we revisit this challenge by considering the behavior of the standard regularization schemes (operating in two nested iterative updates).

Accordingly, we consider an alternative iterative method that avoids solving a sequence of variational inequality problems; instead, each user takes a *single* projection step associated with his regularized problem. By imposing appropriate assumptions on the steplength and regularization sequences, we may recover convergence to the least-norm Nash equilibrium. To summarize, our intent lies in developing algorithms that are characterized by (a) a single iterative process; (b) a distributed architecture that can accommodate computation of equilibria; and (c) finally, the ability to accommodate uncertainty via an expected-value objectives. An important characteristic of distributed schemes is the autonomous choice of parameters that users are expected to be provided with. Thus, we consider a situation where users choose their individual stepsize, which leads to the following coupled user-specific Tikhonov updates:

$$x_i^{k+1} = \Pi_{K_i}[x_i^k - \alpha_{k,i}(F_i(x^k) + w_i^k + \varepsilon_{k,i} x_i^k)], \tag{PITR}$$

where $x_i^0 \in K_i$ is a random initial point with finite expectation $\mathbb{E}[\|x_i^0\|^2]$ and $F_i(x^k)$ denotes the $i$th component of mapping $F$ evaluated at $x^k$. The vector $w_i^k$ is a stochastic error for user $i$ in evaluating $F_i(x^k)$, while $\alpha_{k,i}$ is the stepsize and $\varepsilon_{k,i}$ is the regularization parameter chosen by user $i$ at the $k$th iteration. The iterate updates can be compactly written as

$$x^{k+1} = \Pi_K[x^k - D(\alpha_k)(F(x^k) + D(\varepsilon_k)x^k + w^k)], \tag{8}$$

where $F = (F_1, \ldots, F_N)$, $K = \prod_{i=1}^N K_i$ and $w^k = (w_1^k, \ldots, w_N^k)$, while $D(\alpha_k)$ and $D(\varepsilon_k)$ denote the diagonal matrices with diagonal entries $\alpha_{k,i}$ and $\varepsilon_{k,i}$, respectively. We use abbreviation PITR to refer to the iterative projection algorithm in (PITR) and its equivalent compact version in (8). The abbreviation stands for Partially-coordinated Iterative Tikhonov Projection method, which is motivated by the partial coordination of the user parameters $\alpha_{k,i}$ and $\varepsilon_{k,i}$ that is required for convergence of the method.

Typically, an iterative Tikhonov method is studied by at first analyzing the behavior of the Tikhonov sequence $\{y^k\}$, where each $y^k$ is a (unique) solution to $\mathrm{VI}(K, F + \bar{\varepsilon}_k I)$ and the sequence $\{y_k\}$ is obtained as the parameter $\bar{\varepsilon}_k \geq 0$ is let to go to zero. Under certain conditions the Tikhonov sequence $\{y^k\}$ converges to the smallest norm solution of $\mathrm{VI}(K, F)$. Then, the sequence of iterates $\{x^k\}$ is related to the Tikhonov sequence to assert the convergence of the iterates $x^k$.

We adopt the same approach. However, we cannot directly use the existing results for Tikhonov sequence $\{y^k\}$ such as, for example, those given in Chapter 12.2 of [16]. In particular, arising from user-specific Tikhonov regularization parameters $\varepsilon_{k,i}$ (cf. (8), our variational inequalities have the form $\mathrm{VI}(K, F + D(\varepsilon_k))$ instead of $\mathrm{VI}(K, F + \bar{\varepsilon}_k I)$ (which would be obtained if all the users choose the same regularization parameter $\varepsilon k, i = \bar{\varepsilon}_k$). In the next two subsections, we develop a necessary result for Tikhonov sequence and investigate the convergence of the method.

*A. Tikhonov Sequence*

Here, we analyze the behavior of our Tikhonov sequence $\{y^k\}$ as each user lets its own regularization parameter $e_{k,i}$ go to zero with $y^k$ being a solution to the coupled variational inequality $\mathrm{VI}(K, F + D(\varepsilon_k))$. Recall that $D(\varepsilon_k)$ is the diagonal matrix with diagonal entries $\varepsilon_{k,i} > 0$ and note that each $\mathrm{VI}(K, F + D(\varepsilon_k))$ is strongly monotone. Thus, the sequence $\{y^k\}$ is uniquely determined by the choice of user sequences $\{\varepsilon_{k,i}\}$, $i = 1, \ldots, N$. For this sequence, we have the following result.

**Lemma 3.** *Let the set $K \subseteq \mathbb{R}^n$ be closed and convex, and let the map $F : K \to \mathbb{R}^n$ be continuous and monotone over K. Assume that SOL$(K, F)$ is nonempty. Let the sequences $\{\varepsilon_{k,i}\}$, $i = 1, \ldots, N$, be monotonically decreasing to 0 and such that*

$$\limsup_{k\to\infty} \frac{\varepsilon_{k,\max}}{\varepsilon_{k,\min}} < \infty,$$

*where $\varepsilon_{k.max} = \max_i \varepsilon_{k,i}$ and $\varepsilon_{k,\min} = \min_i \varepsilon_{k,i}$. Then, for the Tikhonov sequence $\{y^k\}$ we have*

*(a) $\{y^k\}$ is bounded and every accumulation point of $\{y^k\}$ is a solution of VI$(K, F)$.*

*(b) The following inequality holds*

$$\|y^k - y^{k-1}\| \leq M_y \frac{\varepsilon_{k-1,\max} - \varepsilon_{k,\min}}{\varepsilon_{k,\min}} \|y^{k-1}\| \qquad \text{for all } k \geq 1,$$

*where $M_y$ is a norm bound on the Tikhonov sequence, i.e., $\|y^k\| \leq M_y$ for all $k \geq 0$.*

*(c) If $\limsup_{k\to\infty} \frac{\varepsilon_{k,\max}}{\varepsilon_{k,\min}} \leq 1$, then $\{y^k\}$ converges to the smallest norm solution of VI$(K, F)$.*

*Proof:* (a) Since $\mathrm{SOL}(K, F) \neq \emptyset$, by letting $x^*$ be any solution of $\mathrm{VI}(K, F)$ we have

$$(x - x^*)^T F(x^*) \geq 0 \qquad \text{for all } x \in K. \tag{9}$$

Since $y^k \in K$ solves $\mathrm{VI}(K, F + D(\varepsilon_k))$ for each $k \geq 0$ we have

$$(y - y^k)^T (F(y^k) + D(\varepsilon_k) y^k) \geq 0 \qquad \text{for all } y \in K \text{ and } k \geq 0. \tag{10}$$

By letting $x = y^k$ in Eq. (9) and $y = x^*$ in Eq. (10), we obtain for all $k \geq 0$,

$$(y^k - x^*)^T F(x^*) \geq 0$$
$$(x^* - y^k)^T (F(y^k) + D(\varepsilon_k) y^k) \geq 0.$$

By the monotonicity of $F$ we have $(y^k - x^*)^T (F(x^*) - F(y^k)) \leq 0$, implying that

$$(x^* - y^k)^T D(\varepsilon_k) y^k \geq 0.$$

By rearranging the terms in above expression we have

$$(x^*)^T D(\varepsilon_k) y^k \geq (y^k)^T D(\varepsilon_k) y^k \geq \varepsilon_{k,\min} \|y^k\|^2,$$

where $\varepsilon_{k,\min} = \min_{1\leq i\leq N} \varepsilon_{k,i}$. By using the Cauchy-Schwartz inequality, we see that

$$\varepsilon_{k,\max} \|x^*\| \, \|y^k\| \geq (x^*)^T D(\varepsilon_k) y^k,$$

where $\varepsilon_{k,\max} = \max_{1\le i\le N} \varepsilon_{k,i}$. Combining the preceding two inequalities, we obtain

$$\|y^k\| \le \frac{\varepsilon_{k,\max}}{\varepsilon_{k,\min}} \|x^*\|. \tag{11}$$

Let $\limsup_{k\to\infty} \frac{\varepsilon_{k,\max}}{\varepsilon_{k,\min}} = c$. Since $c$ is finite (by our assumption), it follows that the sequence $\{y^k\}$ is bounded. By choosing any accumulation point $\bar{y}$ of $\{y^k\}$ and letting $k\to\infty$ in Eq. (10) over a corresponding convergent subsequence of $\{y^k\}$, in view of continuity of $F$ and $\varepsilon_{k,i}\to 0$ as $k\to\infty$, we conclude that

$$(y-\tilde{y})^T F(\tilde{y}) \ge 0 \qquad \text{for all } y\in K.$$

Thus, every accumulation point $\tilde{y}$ of $\{y^k\}$ is a solution to VI$(K,F)$.

(b) Now, we establish the inequality satisfied by the Tikhonov sequence $\{y^k\}$. Since $y^k$ solves VI$(K,F+D(\varepsilon_k))$ for each $k\ge 0$, we have for $k\ge 1$,

$$(y^{k-1}-y^k)^T(F(y^k)+D(\varepsilon_k)y^k)\ge 0 \ \text{ and } \ (y^k-y^{k-1})^T(F(y^{k-1})+D(\varepsilon_{k-1})y^{k-1})\ge 0.$$

By adding the preceding relations, we obtain

$$(y^{k-1}-y^k)^T(F(y^k)-F(y^{k-1}))+(y^{k-1}-y^k)^T(D(\varepsilon_k)y^k-D(\varepsilon_{k-1})y^{k-1})\ge 0.$$

By the monotonicity of the mapping $F$, it follows

$$(y^{k-1}-y^k)^T(D(\varepsilon_k)y^k-D(\varepsilon_{k-1})y^{k-1})\ge 0,$$

and thus

$$(y^{k-1}-y^k)^T(D(\varepsilon_k)y^k-D(\varepsilon_k)y^{k-1}+D(\varepsilon_k)y^{k-1}-D(\varepsilon_{k-1})y^{k-1})\ge 0.$$

By rearranging the terms in the above expression, we obtain

$$\begin{aligned}(y^{k-1}-y^k)^T(D(\varepsilon_k)-D(\varepsilon_{k-1}))y^{k-1} &\ge D(\varepsilon_k)(y^{k-1}-y^k)^T(y^{k-1}-y^k)\\ &\ge \varepsilon_{k,\min}\|y^k-y^{k-1}\|^2.\end{aligned}$$

In the view of the Cauchy-Schwartz inequality, the left hand side is bounded from above as

$$\begin{aligned}(y^{k-1}-y^k)^T(D(\varepsilon_k)-D(\varepsilon_{k-1}))y^{k-1} &\le \|y^{k-1}-y^k\|\left\|(D(\varepsilon_k)-D(\varepsilon_{k-1}))\,y^{k-1}\right\|\\ &\le (\varepsilon_{k-1,\max}-\varepsilon_{k,\min})\,\|y^{k-1}-y^k\|\|y^{k-1}\|,\end{aligned}$$

where we use the monotonically decreasing property of the regularization sequences $\{\varepsilon_{k,i}\}$ $i=1,\dots,N$, to bound the norm $\|D(\varepsilon_k)-D(\varepsilon_{k-1})\|$. Combining the preceding relations we obtain

$$\|y^k-y^{k-1}\|\le \|y^{k-1}\|\frac{\varepsilon_{k-1,\max}-\varepsilon_{k,\min}}{\varepsilon_{k,\min}}. \tag{12}$$

From the part (a) we have that the Tikhonov sequence is bounded. Let $M_y>0$ be such that $\|y^k\|\le M_y$ for all $k$. Then, from relation (12) we obtain

$$\|y^k-y^{k-1}\|\le M_y\frac{\varepsilon_{k-1,\max}-\varepsilon_{k,\min}}{\varepsilon_{k,\min}} \qquad \text{for all } k\ge 1.$$

(c) Suppose now $\limsup_{k\to\infty} \frac{\varepsilon_{k,\max}}{\varepsilon_{k,\min}} \le 1$. Then, by part (a), the sequence $\{y^k\}$ is bounded. Furthermore, in view of relation (11) (where the solution $x^*$ is arbitrary), it follows that every accumulation point $\tilde{y}$ of $\{y^k\}$ satisfies

$$\|\tilde{y}\| \le \limsup_{k\to\infty} \frac{\varepsilon_{k,\max}}{\varepsilon_{k,\min}} \|x^*\| \le \|x^*\| \qquad \text{for all } x^* \in \mathrm{SOL}(K,F).$$

Therefore, every accumulation point $\tilde{y}$ of $\{y^k\}$ is bounded above the least-norm solution of $\mathrm{VI}(K,F)$. At the same time, by part (a), we have $\tilde{y} \in \mathrm{SOL}(K,F)$ for every accumulation point $\tilde{y}$, thus implying that the sequence $\{y^k\}$ must converge to the smallest norm solution of $\mathrm{VI}(K,F)$. ■

Lemma 3 plays a key role in the convergence analysis of the stochastic iterative Tikhonov method (PITR). Aside from this, Lemma 3 may be of its own interest as it extends the existing results for Tikhonov regularization to the case when the regularization mapping is a time varying diagonal matrix.

### *B. Almost Sure Convergence of Stochastic Iterative Tikhonov Method*

We now focus on the method in (PITR). We introduce some notation and state assumptions on the stochastic errors $w^k$ that are standard in stochastic approximation schemes. Specifically, throughout this section and the remainder of the paper, we use $\mathcal{F}_k$ to denote the sigma-field generated by the initial point $x^0$ and errors $w^\ell$ for $\ell = 0,1,\ldots,k$, i.e., $\mathcal{F}_0 = \{x^0\}$ and

$$\mathcal{F}_k = \{x^0,\ (w^\ell,\ \ell = 0,1,\ldots,k-1)\} \qquad \text{for } k \ge 1.$$

Now, we specify our assumptions for $\mathrm{VI}(K,F)$ in (5)–(6) and the stochastic errors $w^k$.

**Assumption 1.** *Let the following hold:*

*(a) The sets $K_i \subseteq \mathbb{R}^{n_i}$ are closed and convex;*

*(b) The mapping $F : K \to \mathbb{R}^n$ is monotone and Lipschitz continuous with a constant $L$ over the set $K$;*

*(c) The stochastic error is such that $\mathbb{E}[w^k \mid \mathcal{F}_k] = 0$ almost surely for all $k \ge 0$.*

Expectedly, convergence of the method (PITR) does rely on some coordination requirements across steplengths and the regularization parameters. Specifically, we impose the following conditions.

**Assumption 2.** *Let $\alpha_{k,\max} = \max_{1\le i\le N}\{\alpha_{k,i}\}$, $\alpha_{k,\min} = \min_{1\le i\le N}\{\alpha_{k,i}\}$, $\varepsilon_{k,\max} = \max_{1\le i\le N}\{\varepsilon_{k,i}\}$, $\varepsilon_{k,\min} = \min_{1\le i\le N}\{\varepsilon_{k,i}\}$. Let $\{\varepsilon_{k,i}\}$ be a monotonically decreasing sequence for each $i$. Furthermore, with $L$ being the Lipschitz constant of mapping $F$, let the following hold:*

*(a)* $\lim_{k\to\infty} \frac{\alpha_{k,\max}}{\alpha_{k,\min}} \frac{\alpha_{k,\max}}{\varepsilon_{k,\min}} (\varepsilon_{k,\max} + L)^2 = 0$ *and* $\lim_{k\to\infty} \frac{\alpha_{k,\max} - \alpha_{k,\min}}{\alpha_{k,\min}\varepsilon_{k,\min}} = 0$*;*

*(b)* $\lim_{k\to\infty} \alpha_{k,\min}\varepsilon_{k,\min} = 0$ *and* $\lim_{k\to\infty} \varepsilon_{k,i} = 0$ *for all $i$;*

*(c)* $\sum_{k=0}^{\infty} \alpha_{k,\min}\varepsilon_{k,\min} = \infty$*;*

*(d)* $\sum_{k=1}^{\infty} \frac{(\varepsilon_{k-1,\max} - \varepsilon_{k,\min})^2}{\varepsilon_{k,\min}^2} \left(1 + \frac{1}{\alpha_{k,\min}\varepsilon_{k,\min}}\right) < \infty$*;*

*(e)* $\lim_{k\to\infty} \frac{(\varepsilon_{k-1,\max} - \varepsilon_{k,\min})^2}{\varepsilon_{k,\min}^3 \alpha_{k,\min}} \left(1 + \frac{1}{\alpha_{k,\min}\varepsilon_{k,\min}}\right) \to 0$*;*

*(f)* $\lim_{k\to\infty} \frac{\alpha_{k,\max}}{\varepsilon_{k,\min}} \mathbb{E}[\|w^k\|^2 \mid \mathcal{F}_k] = 0$ *and* $\sum_{k=0}^{\infty} \alpha_{k,\max}^2 \mathbb{E}[\|w^k\|^2 \mid \mathcal{F}_k] < \infty$ *almost surely.*

When all the stepsizes $\alpha_{k,i}$ and the regularization parameters $\varepsilon_{k,i}$ across the users are the same, the conditions in Assumption 2 are a combination of the conditions typically assumed for deterministic Tikhonov algorithms and the stepsize conditions imposed in stochastic approximation methods. Later in forthcoming Lemma 4, we demonstrate that Assumption 2 can be satisfied by a simple choice of steplength and regularization sequences of the form $(k+\eta_i)^{-a}$ and $(k+\zeta_i)^{-b}$.

In the following proposition, using Assumption 2, we prove that the random sequence $\{x^k\}$ of the method (PITR) and the Tikhonov sequence $\{y^k\}$ associated with the problems VI$(K, F+D(\varepsilon_k))$, $k \geq 0$, have the same accumulation points almost surely. Assumption 2 basically provides the conditions on the sequences $\{\varepsilon_{k,i}\}$ and $\{\alpha_{k,i}\}$ ensuring that the sequence $\{\|x^k - y^{k-1}\|^2\}$ is a convergent supermartingale.

**Proposition 1.** *Let Assumptions 1 and 2 hold. Also, assume that $SOL(K,F)$ is nonempty. Let the sequence $\{x^k\}$ be generated by stochastic iterative Tikhonov algorithm* (PITR). *Then, we have*

$$\lim_{k\to\infty} \|x^k - y^{k-1}\| = 0 \qquad \textit{almost surely}.$$

*Proof:* By using the relation $y_i^k = \Pi_{K_i}[y_i^k - \alpha_{k,i}(F_i(y^k) + \varepsilon_{k,i} y_i^k)]$ and non-expansive property of the projection operator, we have

$$\begin{aligned}\|x_i^{k+1} - y_i^k\|^2 &= \|\Pi_{K_i}[x_i^k - \alpha_{k,i}(F_i(x^k) + \varepsilon_{k,i} x_i^k + w_i^k)] - \Pi_{K_i}[y_i^k - \alpha_{k,i}(F_i(y^k) + \varepsilon_{k,i} y_i^k)]\|^2 \\ &\leq \|x_i^k - \alpha_{k,i}(F_i(x^k) + \varepsilon_{k,i} x_i^k + w_i^k) - y_i^k + \alpha_{k,i}(F_i(y^k) + \varepsilon_{k,i} y_i^k)\|^2.\end{aligned}$$

Further, on expanding the expression on left of the preceding relation it can be verified that

$$\begin{aligned}\|x_i^{k+1} - y_i^k\|^2 \leq \|x_i^k - y_i^k\|^2 &- 2\alpha_{k,i}(x_i^k - y_i^k)^T(F_i(x^k) - F_i(y^k)) - 2\alpha_{k,i}\varepsilon_{k,i}\|x_i^k - y_i^k\|^2 \\ &- 2\alpha_{k,i}(x_i^k - y_i^k)^T w_i^k + \alpha_{k,i}^2\|F_i(x^k) - F_i(y^k) + w_i^k + \varepsilon_{k,i}(x_i^k - y_i^k)\|^2. \qquad (13)\end{aligned}$$

The last term in the inequality can be expanded as

$$\begin{aligned}\|F_i(x^k) - F_i(y^k) + w_i^k + \varepsilon_{k,i}(x_i^k - y_i^k)\|^2 &= \|F_i(x^k) - F_i(y^k)\|^2 + \|w_i^k\|^2 + \varepsilon_{k,i}^2\|x_i^k - y_i^k\|^2 + 2(F_i(x^k) - F_i(y^k))^T w_i^k \\ &+ 2\varepsilon_{k,i}(x_i^k - y_i^k)^T w_i^k + 2\varepsilon_{k,i}(F_i(x^k) - F_i(y^k))^T(x_i^k - y_i^k). \qquad (14)\end{aligned}$$

Now, we take the expectation of (13) and (14) conditional on the past $\mathcal{F}_k$, and use $\mathbb{E}[w_i^k \mid \mathcal{F}_k] = 0$ (cf. Assumption 1(c)). By combining the resulting two relations we get

$$\begin{aligned}\mathbb{E}[\|x_i^{k+1} - y_i^k\|^2 \mid \mathcal{F}_k] \leq (1 - 2\alpha_{k,i}\varepsilon_{k,i} + \alpha_{k,i}^2\varepsilon_{k,i}^2)\|x_i^k - y_i^k\|^2 + \alpha_{k,i}^2\left(\|F_i(x^k) - F_i(y^k)\|^2 + \mathbb{E}[\|w_i^k\|^2 \mid \mathcal{F}_k]\right) \\ + 2\alpha_{k,i}^2\varepsilon_{k,i}(x_i^k - y_i^k)^T(F_i(x^k) - F_i(y^k)) - 2\alpha_{k,i}(x_i^k - y_i^k)^T(F_i(x^k) - F_i(y^k)).\end{aligned}$$

Summing over all $i$ and using $\alpha_{k,\min} \leq \alpha_{k,i} \leq \alpha_{k,\max}$, $\varepsilon_{k,\min} \leq \varepsilon_{k,i} \leq \varepsilon_{k,\max}$ together with the Lipschitz continuity of $F$ yields

$$\begin{aligned}\mathbb{E}[\|x^{k+1} - y^k\|^2 \mid \mathcal{F}_k] \leq (1 - 2\alpha_{k,\min}\varepsilon_{k,\min} + \alpha_{k,\max}^2\varepsilon_{k,\max}^2 + \alpha_{k,\max}^2 L^2)\|x^k - y^k\|^2 + \alpha_{k,\max}^2\mathbb{E}[\|w^k\|^2 \mid \mathcal{F}_k] \\ + 2\sum_{i=1}^N \alpha_{k,i}^2\varepsilon_{k,i}(x_i^k - y_i^k)^T(F_i(x^k) - F_i(y^k)) - 2\sum_{i=1}^N \alpha_{k,i}(x_i^k - y_i^k)^T(F_i(x^k) - F_i(y^k)). \quad (15)\end{aligned}$$

Next we estimate the last two sums in (15) with the inner product terms $(x_i^k - y_i^k)^T(F_i(x^k) - F_i(y^k))$. The first sum involving $\alpha_{k,i}^2 \varepsilon_{k,i}$ can be estimated as follows:

$$\sum_{i=1}^N 2\alpha_{k,i}^2 \varepsilon_{k,i}(x_i^k - y_i^k)^T(F_i(x^k) - F_i(y^k)) \le \alpha_{k,\max}^2 \varepsilon_{k,\max} \sum_{i=1}^N \|x_i^k - y_i^k\| \|F_i(x^k) - F_i(y^k)\|.$$

By Hölder's inequality, we have $\sum_{i=1}^N \|x_i^k - y_i^k\|\|F_i(x^k) - F_i(y^k)\| \le \|x^k - y^k\|\|F(x^k) - F(y^k)\|$, which through the use of Lipschitz continuity of $F$ yields

$$\sum_{i=1}^N 2\alpha_{k,i}^2 \varepsilon_{k,i}(x_i^k - y_i^k)^T(F_i(x^k) - F_i(y^k)) \le \alpha_{k,\max}^2 \varepsilon_k L \|x^k - y^k\|^2. \tag{16}$$

Adding and subtracting the terms $\alpha_{k,\min}(x_i^k - y_i^k)^T(F_i(x^k) - F_i(y^k))$ in the last term of (15) we have

$$\begin{aligned} -\sum_{i=1}^N 2\alpha_{k,i}(x_i^k - y_i^k)^T(F_i(x^k) - F_i(y^k)) &\le -2\alpha_{k,\min}(x^k - y^k)^T(F(x^k) - F(y^k)) \\ &\quad + 2(\alpha_{k,\max} - \alpha_{k,\min}) \sum_{i=1}^N \|x_i^k - y_i^k\|\|F_i(x^k) - F_i(y^k)\|. \end{aligned}$$

Using monotonicity of $F$ we have $(x^k - y^k)^T(F(x^k) - F(y^k)) \ge 0$. Further by the use of Hölder's inequality and Lipschitz continuity of $F$, we get

$$-\sum_{i=1}^N 2\alpha_{k,i}(x_i^k - y_i^k)^T(F_i(x^k) - F_i(y^k)) \le 2\delta_k L \|x^k - y^k\|^2, \tag{17}$$

where $\delta_k \triangleq \alpha_{k,\max} - \alpha_{k,\min}$. Combining the results of relation (15), (16) and (17) we obtain

$$\begin{aligned} \mathbb{E}[\|x^{k+1} - y^k\|^2 \mid \mathcal{F}_k] &\le (1 - 2\alpha_{k,\min}\varepsilon_{k,\min} + \alpha_{k,\max}^2 \varepsilon_{k,\max}^2 + \alpha_{k,\max}^2 L^2)\|x^k - y^k\|^2 \\ &\quad + \alpha_{k,\max}^2 \mathbb{E}[\|w^k\|^2 \mid \mathcal{F}_k] + (2\alpha_{k,\max}^2 \varepsilon_{k,\max} L + 2\delta_k L)\|x^k - y^k\|^2. \end{aligned}$$

Letting $q_k \triangleq 1 - 2\alpha_{k,\min}\varepsilon_{k,\min} + \alpha_{k,\max}^2(\varepsilon_{k,\max} + L)^2 + 2\delta_k L$, we can write

$$\mathbb{E}[\|x^{k+1} - y^k\|^2 \mid \mathcal{F}_k] \le q_k \|x^k - y^k\|^2 + \alpha_{k,\max}^2 \mathbb{E}[\|w^k\|^2 \mid \mathcal{F}_k]. \tag{18}$$

Now, we relate $\|x^k - y^k\|$ to $\|x^k - y^{k-1}\|$. By the triangle inequality $\|x^k - y^k\| \le \|x^k - y^{k-1}\| + \|y^{k-1} - y^k\|$ while from Lemma 3 we have

$$\|y^k - y^{k-1}\| \le M_y \frac{\varepsilon_{k-1,\max} - \varepsilon_{k,\min}}{\varepsilon_{k,\min}}.$$

Therefore, it follows that

$$\begin{aligned} \|x^k - y^k\|^2 &\le \|x^k - y^{k-1}\|^2 + \|y^k - y^{k-1}\|^2 + 2\|x^k - y^{k-1}\|\|y^k - y^{k-1}\| \\ &\le \|x^k - y^{k-1}\|^2 + \left(M_y \frac{\varepsilon_{k-1,\max} - \varepsilon_{k,\min}}{\varepsilon_{k,\min}}\right)^2 + 2M_y \frac{\varepsilon_{k-1,\max} - \varepsilon_{k,\min}}{\varepsilon_{k,\min}} \|x^k - y^{k-1}\|. \end{aligned}$$

Further we use Cauchy-Schwartz inequality to estimate the last term as follows:

$$\begin{aligned} M_y \frac{\varepsilon_{k-1,\max} - \varepsilon_{k,\min}}{\varepsilon_{k,\min}} \|x^k - y^{k-1}\| &= 2\sqrt{\alpha_{k,\min}\varepsilon_{k,\min}} \|x^k - y^{k-1}\| \frac{\varepsilon_{k-1,\max} - \varepsilon_{k,\min}}{\sqrt{\alpha_{k,\min}\varepsilon_{k,\min}}\varepsilon_{k,\min}} M_y \\ &\le \alpha_{k,\min}\varepsilon_{k,\min}\|x^k - y^{k-1}\|^2 + \frac{(\varepsilon_{k-1,\max} - \varepsilon_{k,\min})^2}{\alpha_{k,\min}\varepsilon_{k,\min}^3} M_y^2 \end{aligned}$$

Using this in the preceding relation we obtain

$$\|x^k - y^k\|^2 \le (1+\alpha_{k,\min}\varepsilon_{k,\min})\|x^k - y^{k-1}\|^2 + \left(M_y \frac{\varepsilon_{k-1,\max}-\varepsilon_{k,\min}}{\varepsilon_{k,\min}}\right)^2 \left(1+\frac{1}{\alpha_{k,\min}\varepsilon_{k,\min}}\right). \tag{19}$$

Combining the relations of (18) and (19) we obtain the following estimate:

$$\begin{aligned}\mathbb{E}[\|x^{k+1}-y^k\|^2 \mid \mathcal{F}_k] &\le q_k(1+\alpha_{k,\min}\varepsilon_{k,\min})\|x^k-y^{k-1}\|^2 \\ &+ q_k\left(M_y \frac{\varepsilon_{k-1,\max}-\varepsilon_{k,\min}}{\varepsilon_{k,\min}}\right)^2\left(1+\frac{1}{\alpha_{k,\min}\varepsilon_{k,\min}}\right) + \alpha_{k,\max}^2\mathbb{E}[\|w^k\|^2 \mid \mathcal{F}_k].\end{aligned} \tag{20}$$

Next, we estimate the coefficient of $\|x^k - y^{k-1}\|^2$ in (20). Recall that according to the modified definition $q_k = 1 - 2\alpha_{k,\min}\varepsilon_k + \alpha_{k,\max}^2(\varepsilon_k + L)^2 + 2\delta_k L$. First we show that $q_k \in (0,1)$ for all $k$ large enough. Note that we can write

$$q_k = 1 - \alpha_{k,\min}\varepsilon_{k,\min}\left(2 - \frac{\alpha_{k,\max}^2}{\alpha_{k,\min}\varepsilon_{k,\min}}(\varepsilon_{k,\max}+L)^2 - \frac{2\delta_k L}{\alpha_{k,\min}\varepsilon_{k,\min}}\right).$$

By Assumption 2(a) we have

$$\frac{\alpha_{k,\max}^2}{\alpha_{k,\min}\varepsilon_{k,\min}}(\varepsilon_{k,\max}+L)^2 + \frac{2\delta_k L}{\alpha_{k,\min}\varepsilon_{k,\min}} \to 0,$$

implying that there exists a large enough integer $K \ge 0$ such that

$$\frac{\alpha_{k,\max}^2}{\alpha_{k,\min}\varepsilon_{k,\min}}(\varepsilon_{k,\max}+L)^2 + \frac{2\delta_k L}{\alpha_{k,\min}\varepsilon_{k,\min}} \le c \qquad \text{for all } k \ge K \text{ and some } c \in (0,1). \tag{21}$$

Thus,

$$1 \le 2 - \frac{\alpha_{k,\max}^2}{\alpha_{k,\min}\varepsilon_{k,\min}}(\varepsilon_{k,\max}+L)^2 - \frac{2\delta_k L}{\alpha_{k,\min}\varepsilon_{k,\min}} \le 2 \qquad \text{for all } k \ge K,$$

implying that for $q_k$ we have

$$1 - 2\alpha_{k,\min}\varepsilon_{k,\min} \le q_k \le 1 - \alpha_{k,\min}\varepsilon_{k,\min} \qquad \text{for all } k \ge K.$$

Furthermore, since $\alpha_{k,\min}\varepsilon_{k,\min} \to 0$ by Assumption 2(b), we can choose $K$ large enough so that $q_k \in (0,1)$ for $k \ge K$. Hence, for $k \ge K$ we obtain $0 \le q_k(1+\alpha_{k,\min}\varepsilon_{k,\min}) \le q_k + \alpha_{k,\min}\varepsilon_{k,\min}$ and using the definition of $q_k$ we further have for $k \ge K$,

$$0 \le q_k(1+\alpha_{k,\min}\varepsilon_{k,\min}) \le 1 - \alpha_{k,\min}\varepsilon_{k,\min}\left(1 - \frac{\alpha_{k,\max}^2}{\alpha_{k,\min}\varepsilon_k}(\varepsilon_{k,\max}+L)^2 - \frac{2\delta_k L}{\alpha_{k,\min}\varepsilon_{k,\min}}\right) \le 1 - \alpha_{k,\min}\varepsilon_{k,\min}(1-c), \tag{22}$$

where the last inequality follows from (21). Using relations (22) and (20), we obtain

$$\mathbb{E}[\|x^{k+1}-y^k\|^2 \mid \mathcal{F}_k] \le (1-u_k)\|x^k - y^{k-1}\|^2 + v_k \qquad \text{for all } k \ge K,$$

where $u_k \triangleq (1-c)\alpha_{k,\min}\varepsilon_{k,\min}$ and

$$v_k = q_k\left(M_y \frac{\varepsilon_{k-1,\max}-\varepsilon_{k,\min}}{\varepsilon_{k,\min}}\right)^2\left(1+\frac{1}{\alpha_{k,\min}\varepsilon_{k,\min}}\right) + \alpha_{k,\max}^2\mathbb{E}[\|w^k\|^2 \mid \mathcal{F}_k].$$

We now verify that the conditions of Lemma 1 are satisfied for $k \geq K$. Since $c < 1$, in view of relation (22) we have $0 \leq u_k \leq 1$ for all $k \geq K$, while from Assumption 2(c) we have $\sum_{k=K}^{\infty} u_k = \infty$. Under stepsize conditions Assumption 2(d)–(f), it can be verified that $\lim_{k\to\infty} \frac{v_k}{u_k} = 0$ and $\sum_{k=0}^{\infty} v_k < \infty$. Thus, the conditions of Lemma 1 are satisfied for $k \geq K$. Noting that Lemma 1 applies to a process delayed by a deterministic time-offset, we can conclude that $\|x^k - y^{k-1}\| \to 0$ almost surely. ■

As an immediate consequence of Proposition1 and the properties of Tikhonov sequence established in Lemma 3, we have the following result.

**Proposition 2.** *Let Assumptions 1 and 2 hold. Also, assume that SOL$(K,F)$ is nonempty. Then, for the sequence $\{x^k\}$ generated by stochastic iterative Tikhonov algorithm (PITR), we have*

*(a) If $\limsup_{k\to\infty} \frac{\varepsilon_{k,\max}}{\varepsilon_{k,\min}} < \infty$, then every accumulation point of $\{x^k\}$ is a solution of VI$(K,F)$.*

*(b) If $\limsup_{k\to\infty} \frac{\varepsilon_{k,\max}}{\varepsilon_{k,\min}} \leq 1$, then $\{x^k\}$ converges to the smallest-norm solution of VI$(K,F)$.*

A further extension of Proposition 1 is obtained when the mapping $F$ is strictly monotone over the set $K$. In this case, the uniqueness of solution of VI$(K,F)$ is guaranteed provided a solution exists. Hence, from Lemma 3(a) we have $\{y^k\}$ converging to the unique solution of VI$(K,F)$, which in view of Proposition 1 implies that $\{x^k\}$ is converging to the solution almost surely. This result is precisely presented in the following corollary.

**Corollary 1.** *Let Assumption 1 hold with $F$ being strictly monotone over the set $K$. Also let Assumption 2 hold, and assume that SOL$(K,F)$ is nonempty. Then, the sequence $\{x^k\}$ generated by iterative Tikhonov method (PITR) converges to the unique solution of VI$(K,F)$ almost surely.*

We conclude this section by providing an example of steplength and regularization sequences that satisfy the conditions of Assumption 2(a)–(e).

**Lemma 4.** *Consider the choice $\alpha_{k,i} = (k+\eta_i)^{-a}$ and $\varepsilon_k = (k+\zeta_i)^{-b}$ for all $k$, where each $\eta_i$ and $\zeta_i$ are selected from a uniform distribution on the interval $[-\eta,\eta]$ for some $\eta > 0$, while $a,b \in (0,1)$, $a+b<1$, and $a > b$. Then, the sequences $\{\alpha_{k,i}\}$ and $\{\varepsilon_{k,i}\}$ satisfy Assumption 2(a)–(e).*

*Proof:* The first limit condition in Assumption 2(a) holds trivially as we see that for $a > b$,

$$\lim_{k\to\infty} \frac{\alpha_{k,\max}}{\alpha_{k,\min}} \frac{\alpha_{k,\max}}{\varepsilon_{k,\min}} (\varepsilon_{k,\max}+L)^2 = \lim_{k\to\infty} \frac{(k+\eta_{\min})^{-a}}{(k+\eta_{\max})^{-a}} \frac{(k+\eta_{\min})^{-a}}{(k+\zeta_{\max})^{-b}} ((k+\zeta_{\min})^{-b}+L)^2,$$

where $\eta_{\max} = \max_{1\leq i\leq N}\{\eta_i\}$ and $\eta_{\min} = \min_{1\leq i\leq N}\{\eta_i\}$. Moreover,

$$\begin{aligned}\lim_{k\to\infty} \frac{(k+\eta_{\min})^{-a}}{(k+\eta_{\max})^{-a}} &= \lim_{k\to\infty} \frac{1}{\left(\frac{(k+\eta_{\max})}{(k+\eta_{\min})}\right)^{-a}} \\ &= \lim_{k\to\infty} \frac{1}{\left(1+\frac{(\eta_{\max}-\eta_{\min})}{(k+\eta_{\min})}\right)^{-a}} = 1.\end{aligned}$$

Therefore,

$$\lim_{k\to\infty}\frac{(k+\eta_{\min})^{-a}}{(k+\eta_{\max})^{-a}}\frac{(k+\eta_{\min})^{-a}}{(k+\zeta_{\max})^{-b}}((k+\zeta_{\min})^{-b}+L)^2 = \lim_{k\to\infty}\frac{(k+\eta_{\min})^{-a}}{(k+\eta_{\max})^{-a}}\lim_{k\to\infty}\frac{(k+\eta_{\min})^{-a}}{(k+\zeta_{\max})^{-b}}((k+\zeta_{\min})^{-b}+L)^2$$
$$= \lim_{k\to\infty}\frac{(k+\eta_{\min})^{-a}}{(k+\zeta_{\max})^{-b}}((k+\zeta_{\min})^{-b}+L)^2 = 0,$$

where the last equality follows by $a > b$. The second condition of Assumption 2(a) can be seen to follow by noticing that the argument of the limit can be written as

$$\frac{\alpha_{k,\max}-\alpha_{k,\min}}{\alpha_{k,\min}\varepsilon_{k,\max}} = \frac{(k+\eta_{\min})^{-a}-(k+\eta_{\max})^{-a}}{(k+\eta_{\max})^{-a}(k+\zeta_{\max})^{-b}} = \frac{\frac{(k+\eta_{\min})^{-a}}{(k+\eta_{\max})^{-a}}-1}{k^{-b}(1+\frac{\zeta_{\max}}{k})^{-b}}$$
$$= \frac{\left(1-\frac{\eta_{\max}-\eta_{\min}}{k+\eta_{\max}}\right)^{-a}-1}{k^{-b}(1+\frac{\zeta_{\max}}{k})^{-b}} \approx \frac{1+a\frac{\eta_{\max}-\eta_{\min}}{k+\eta_{\max}}+O(1/k^2)-1}{k^{-b}(1+\frac{\zeta_{\max}}{k})^{-b}} = O(1/k^{1-b}).$$

As $k\to\infty$, the required result follows. Also, Assumption 2(b) and (c) hold since $\alpha_{k,\min}\varepsilon_{k,\min} = k^{-a-b}(1+\eta_{\max}/k)^{-a}(1+\zeta_{\max}/k)^{-b} > k^{-1}$. Under the given form of $\varepsilon_{k,i}$ and $\alpha_{k,i}$ the expression in the summation of Assumption 2(d) becomes

$$\frac{((k-1+\zeta_{\min})^{-b}-(k+\zeta_{\max})^{-b})^2}{(k+\zeta_{\max})^{-2b}}\left(1+\frac{1}{(k+\eta_{\max})^{-a}(k+\zeta_{\max})^{-b}}\right) \le 2\frac{((1+(\zeta_{\min}-1)/k)^{-b}(1+\zeta_{\max}/k)^{b}-1)^2}{(k+\eta_{\max})^{-a}(k+\zeta_{\max})^{-b}},$$

where the inequality follows from the fact that

$$\frac{1}{(k+\eta_{\max})^{-a}(k+\zeta_{\max})^{-b}} \ge 1 \qquad \text{for } k\ge 1.$$

Using the expansion of $(1-x)^{-b}$ for $x$ small and ignoring higher order terms we have

$$\left((1+(\zeta_{\min}-1)/k)^{-b}(1+\zeta_{\max}/k)^{b}-1\right)^2 \approx \left(\left(1-b\frac{\zeta_{\min}-1}{k}\right)\left(1+b\frac{b\zeta_{\max}}{k}\right)-1\right)^2$$
$$\approx \frac{b^2(\zeta_{\max}-\zeta_{\min}+1)^2}{k^2}$$

Also for $k$ large we have $(k+\eta_{\max})^{-a}(k+\zeta_{\max})^{-b}\approx k^{-a-b}$. Thus we have

$$2\frac{((1+(\zeta_{\min}-1)/k)^{-b}(1+\zeta_{\max}/k)^{b}-1)^2}{(k+\eta_{\max})^{-a}(k+\zeta_{\max})^{-b}} \approx 2\frac{b^2(\zeta_{\max}-\zeta_{\min}+1)^2}{k^{2-a-b}} = O(k^{-(1+\delta)}),$$

where in the equality we use $a+b<1$ and $\delta = 1-(a+b) > 0$. Following a similar argument, it can be verified that the term in Assumption 2(e) reduces to

$$\frac{((k-1+\zeta_{\min})^{-b}-(k+\zeta_{\max})^{-b})^2}{(k+\zeta_{\max})^{-2b}}\left(1+\frac{1}{(k+\eta_{\max})^{-a}(k+\zeta_{\max})^{-b}}\right)\frac{1}{(k+\eta_{\max})^{-a}(k+\zeta_{\max})^{-b}}$$
$$\approx 2\frac{b^2(\zeta_{\max}-\zeta_{\min}+1)^2}{k^{2-a-b}}\frac{1}{k^{-a-b}},$$

and the limit in Assumption 2(e) follows from $a+b<1$. ■

## IV. Stochastic iterative proximal-point schemes

An alternative to using iterative Tikhonov regularization techniques is available through proximal-point methods, a class of techniques that appear to have been first studied by Martinet [37], and subsequently by Rockafellar [44]. A more recent description in the context of maximal-monotone operators can be found in [16]. We begin with a description of such methods in the context of a monotone variational inequality $\mathrm{VI}(K,F)$ where $F$ is a continuous and monotone mapping. In the standard proximal-point methods, the convergence to a single solution of $\mathrm{VI}(K,F)$ is obtained through the addition of a proximal term $\theta(x_k - x_{k-1})$, where $\theta$ is a fixed positive parameter. In effect, $x_k = \mathrm{SOL}(K, F + \theta(\mathbf{I} - x_{k-1})$ and convergence may be guaranteed under suitable assumptions.

When employing a proximal-point method for solving the $\mathrm{VI}(K,F)$ associated with problem (4), a crucial shortcoming of standard proximal-point schemes lies in the need to solve a sequence of variational problems. Analogous to our efforts in constructing an iterative Tikhonov regularization technique, we consider an iterative proximal-point method. In such a method, the centering term $x_{k-1}$ is updated after *every* projection step rather than when it obtains an accurate solution of $\mathrm{VI}(K, F + \theta(\mathbf{I} - x_{k-1}))$.

Before providing a detailed analysis of the convergence properties of this scheme, we examine the relationship between the proposed iterative proximal point method and the standard gradient projection method, in the context of variational inequalities. An iterative proximal-point scheme for $\mathrm{VI}(K,F)$ necessitates an update given by

$$\begin{aligned} x_{k+1} &= \Pi_K[x_k - \gamma_k(F(x_k) + \theta(x_k - x_{k-1}))] \\ &= \Pi_K[((1-\gamma_k\theta)x_k + \gamma_k\theta x_{k-1}) - \gamma_k F(x_k)] \\ &= \Pi_K[x_k(\theta) - \gamma_k F(x_k)], \end{aligned}$$

where $x_k(\theta) \triangleq (1-\gamma_k\theta)x_k + \gamma_k\theta x_{k-1}$. Therefore, when $\theta \equiv 0$, the method reduces to the standard gradient projection scheme. More generally, one can view the proximal-point method as employing a convex combination of the old iterate $x_{k-1}$ and $x_k$ instead of $x_k$ in the standard gradient scheme. Furthermore, as $\gamma_k \to 0$, the update rule starts resembling the standard gradient scheme more closely. In our discussion, we allow $\theta$ to vary at every iteration; in effect, we employ a sequence $\theta_k$ which can grow to $+\infty$ but at a sufficiently slow rate.

Analogous to the partially coordinated iterative Tikhonov (PITR) method, we consider a limited coordination generalization of the iterative proximal-point scheme (PIPP) where users independently choose their individual stepsizes. More precisely we have the following algorithm:

$$x_i^{k+1} = \Pi_{K_i}[x_i^k - \alpha_{k,i}(F_i(x^k) + \theta_{k,i}(x_i^k - x_i^{k-1}) + w_i^k)] \quad \text{for } i = 1, \ldots, N, \tag{PITR}$$

where $\alpha_{k,i}$ is the stepsize and $\theta_{k,i}$ is the centering term parameter chosen by the $i$th user at the $k$th iteration. We make the following assumption on user steplengths and parameters $\theta_{k,i}$.

**Assumption 3.** *Let* $\alpha_{k,\max} = \max_{1 \le i \le N}\{\alpha_{k,i}\}$, $\alpha_{k,\min} = \min_{1 \le i \le N}\{\alpha_{k,i}\}$, $\theta_{k,\max} = \max_{1 \le i \le N}\{\theta_{k,i}\}$, $\theta_{k,\min} = \min_{1 \le i \le N}\{\theta_{k,i}\}$, *and let the following hold:*

*(a)* $\alpha_{k,\max}\theta_{k,\max} \le \left(1+2\alpha_{k,\max}^2 L^2\right)\alpha_{k-1,\min}\theta_{k-1,\min}$ *for all* $k\ge 1$, *and*

$$\lim_{k\to\infty}\frac{\alpha_{k,\max}^2\theta_{k,\max}^2}{\alpha_{k,\min}\theta_{k,\min}} = c \quad \textit{with } c\in[0,1/2);$$

*(b)* $\sum_{k=0}^{\infty}\alpha_{k,i}=\infty$ *and* $\sum_{k=0}^{\infty}\alpha_{k,i}^2<\infty$ *for all i;*

*(c)* $\sum_{k=0}^{\infty}\left(\alpha_{k,\max}-\alpha_{k,\min}\right)<\infty$.

*(d)* $\sum_{k=0}^{\infty}\alpha_{k,\max}^2\mathbb{E}[\|w^k\|^2\mid\mathscr{F}_k]<\infty$ *almost surely.*

Later on, after our convergence results of this section, we will provide an example for the stepsizes and prox-parameters satisfying Assumption 3.

Our main result is given in the following proposition, where by using Assumption 3 we show almost sure convergence of the method. In addition, we assume that the mapping $F$ is strictly monotone over the set $K$, i.e.,

$$(F(x)-F(y))^T(x-y)>0 \qquad \text{for all } x,y\in K \text{ with } x\neq y.$$

We note that when $F$ is strictly monotone and the variational inequality VI$(K,F)$ has a solution, then the solution must be unique (see Theorem 2.3.3. in [16]).

**Proposition 3.** *Let Assumption 1 hold with $F$ being strictly monotone. Assume that SOL$(K,F)$ is nonempty. Also, let the steplengths and the prox-parameters satisfy Assumption 3. Then, the sequence $\{x^k\}$ generated by method (PITR) converges almost surely to the solution of VI$(K,F)$.*

*Proof:* By using $x_i^*=\Pi_{K_i}[x_i^*-\alpha_{k,i}F_i(x^*)]$ and the nonexpansivity of the Euclidean projection operator we observe that $\|x_i^{k+1}-x_i^*\|$ can be expressed as follows

$$\begin{aligned}\|x_i^{k+1}-x_i^*\|^2 &= \|\Pi_{K_i}[x_i^k-\alpha_{k,i}(F_i(x^k)+\theta_{k,i}(x_i^k-x_i^{k-1})+w_i^k)]-\Pi_{K_i}[x_i^*-\alpha_{k,i}F_i(x^*)]\|^2\\ &\le \left\|(x_i^k-x_i^*)-\alpha_{k,i}\left(F_i(x^k)-F_i(x^*)-\theta_{k,i}(x_i^k-x_i^{k-1})-w_i^k\right)\right\|^2.\end{aligned}$$

Further, the right hand side of preceding relation can be expanded as

$$\begin{aligned}\text{RHS} = {} & \|x_i^k-x_i^*\|^2+\alpha_{k,i}^2\|F_i(x^k)-F_i(x^*)\|^2+\alpha_{k,i}^2\|w_i^k\|^2+(\alpha_{k,i}\theta_{k,i})^2\|x_i^k-x_i^{k-1}\|^2\\ &-2\alpha_{k,i}(x_i^k-x_i^*)^T(F_i(x^k)-F_i(x^*))-2\alpha_{k,i}\theta_{k,i}(x_i^k-x_i^*)^T(x_i^k-x_i^{k-1})-2\alpha_{k,i}(x_i^k-x_i^*)^Tw_i^k\\ &+2\alpha_{k,i}^2\theta_{k,i}(F_i(x^k)-F_i(x^*))^T(x_i^k-x_i^{k-1})+2\alpha_{k,i}^2(F_i(x^k)-F_i(x^*))^Tw_i^k+2\alpha_{k,i}^2\theta_{k,i}(x_i^k-x_i^{k-1})^Tw_i^k.\end{aligned}$$

Taking expectation and using $\mathbb{E}[w_i^k\mid\mathcal{F}_k]=0$ (Assumption 1(c)), we obtain

$$\begin{aligned}\mathbb{E}[\|x_i^{k+1}-x_i^*\|^2\mid\mathcal{F}_k] \le {} & \|x_i^k-x_i^*\|^2+\alpha_{k,i}^2\|F_i(x^k)-F_i(x^*)\|^2+\alpha_{k,i}^2\|w_i^k\|^2+(\alpha_{k,i}\theta_{k,i})^2\|x_i^k-x_i^{k-1}\|^2\\ &-2\alpha_{k,i}(x_i^k-x_i^*)^T(F_i(x^k)-F_i(x^*))-2\alpha_{k,i}\theta_{k,i}(x_i^k-x_i^*)^T(x_i^k-x_i^{k-1})\\ &+2\alpha_{k,i}^2\theta_{k,i}(F_i(x^k)-F_i(x^*))^T(x_i^k-x_i^{k-1}).\end{aligned} \tag{23}$$

Let $\alpha_{k,\max} = \max_{1\le i\le N}\{\alpha_{k,i}\}$, $\alpha_{k,\min} = \min_{1\le i\le N}\{\alpha_{k,i}\}$, $\theta_{k,\max} = \max_{1\le i\le N}\{\theta_{k,i}\}$ and $\theta_{k,\min} = \min_{1\le i\le N}\{\theta_{k,i}\}$. Summing over all $i$ and using Lipschitz continuity of $F$ (Assumption 1(b)) we arrive at

$$\begin{aligned}\mathbb{E}[\|x^{k+1}-x^*\|^2 \mid \mathcal{F}_k] \le{}& (1+\alpha_{k,\max}^2L^2)\|x^k-x^*\|^2 + \alpha_{k,\max}^2\mathbb{E}[\|w^k\|^2 \mid \mathcal{F}_k] + (\alpha_{k,\max}\theta_{k,\max})^2\|x^k-x^{k-1}\|^2\\ &\underbrace{-2\sum_{i=1}^N \alpha_{k,i}(x_i^k-x_i^*)^T(F_i(x^k)-F_i(x^*))}_{\textbf{Term 1}} \underbrace{-2\sum_{i=1}^N \alpha_{k,i}\theta_k(x_i^k-x_i^*)^T(x_i^k-x_i^{k-1})}_{\textbf{Term 2}}\\ &\underbrace{+2\sum_{i=1}^N \alpha_{k,i}^2\theta_k(F_i(x^k)-F_i(x^*))^T(x_i^k-x_i^{k-1})}_{\textbf{Term 3}}. \end{aligned}\tag{24}$$

By adding and subtracting $2\alpha_{k,\min}(x_i^k-x_i^*)^T(F_i(x^k)-F_i(x^*))$ to the each term of Term 1 we see that

$$\begin{aligned}\textbf{Term 1} &= -2\alpha_{k,\min}\sum_{i=1}^N(x_i^k-x_i^*)^T(F_i(x^k)-F_i(x^*)) - 2\sum_{i=1}^N(\alpha_{k,i}-\alpha_{k,\min})(x_i^k-x_i^*)^T(F_i(x^k)-F_i(x^*))\\ &\le -2\alpha_{k,\min}(x^k-x^*)^T(F_i(x^k)-F_i(x^*)) + 2(\alpha_{k,\max}-\alpha_{k,\min})\sum_{i=1}^N\|x_i^k-x_i^*\|\|F_i(x^k)-F_i(x^*)\|\\ &\le -2\alpha_{k,\min}(x^k-x^*)^T(F_i(x^k)-F_i(x^*)) + 2(\alpha_{k,\max}-\alpha_{k,\min})\|x^k-x^*\|\|F(x^k)-F(x^*)\|\\ &\le -2\alpha_{k,\min}(x^k-x^*)^T(F_i(x^k)-F_i(x^*)) + 2(\alpha_{k,\max}-\alpha_{k,\min})L\|x^k-x^*\|^2, \end{aligned}\tag{25}$$

where we use Hölder's inequality and Lipschitz continuity of $F$.

We next estimate Term 2. Since $2(x-y)^T(x-z) = \|x-y\|^2+\|x-z\|^2-\|y-z\|^2$, it follows that

$$\begin{aligned}\textbf{Term 2} &= -\sum_{i=1}^N\alpha_{k,i}\theta_{k,i}\left[\|x_i^k-x_i^*\|^2+\|x_i^k-x_i^{k-1}\|^2-\|x_i^{k-1}-x_i^*\|^2\right]\\ &\le -\alpha_{k,\min}\theta_{k,\min}\sum_{i=1}^N\left[\|x_i^k-x_i^*\|^2+\|x_i^k-x_i^{k-1}\|^2\right] + \alpha_{k,\max}\theta_{k,\max}\sum_{i=1}^N\|x_i^{k-1}-x_i^*\|^2\\ &= -\alpha_{k,\min}\theta_{k,\min}\left[\|x^k-x^*\|^2+\|x^k-x^{k-1}\|^2\right] + \alpha_{k,\max}\theta_{k,\max}\|x^{k-1}-x^*\|^2. \end{aligned}\tag{26}$$

We now consider Term 3. Using $2x^Ty \le \|x\|^2+\|y\|^2$ and Lipschitz continuity of $F$, we obtain

$$\begin{aligned}\textbf{Term 3} &\le \sum_{i=1}^N\alpha_{k,i}^2\left(\|F_i(x^k)-F_i(x^*)\|^2+\theta_{k,i}^2\|x_i^k-x_i^{k-1}\|^2\right)\\ &\le \alpha_{k,\max}^2\sum_{i=1}^N\left(\|F_i(x^k)-F_i(x^*)\|^2+\theta_{k,\max}^2\|x_i^k-x_i^{k-1}\|^2\right)\\ &\le \alpha_{k,\max}^2\left(\|F(x^k)-F(x^*)\|^2+\theta_{k,\max}^2\|x^k-x^{k-1}\|^2\right)\\ &\le \alpha_{k,\max}^2\left(L^2\|x^k-x^*\|^2+\theta_{k,\max}^2\|x^k-x^{k-1}\|^2\right). \end{aligned}\tag{27}$$

Combining (24) with (25), (26) and (27) we obtain

$$\mathbb{E}[\|x^{k+1}-x^*\|^2 \mid \mathcal{F}_k] \le \left(1+2\alpha_{k,\max}^2 L^2 + 2(\alpha_{k,\max}-\alpha_{k,\min})L\right)\|x^k-x^*\|^2 + \alpha_{k,\max}\theta_{k,\max}\|x^{k-1}-x^*\|^2$$
$$-\alpha_{k,\min}\theta_{k,\min}\|x^k-x^*\|^2 - \alpha_{k,\min}\theta_{k,\min}\left(1-\frac{2\alpha_{k,\max}^2\theta_{k,\max}^2}{\alpha_{k,\min}\theta_{k,\min}}\right)\|x^k-x^{k-1}\|^2$$
$$-2\alpha_{k,\min}(x^k-x^*)^T(F(x^k)-F(x^*)) + \alpha_{k,\max}^2\mathbb{E}[\|w^k\|^2 \mid \mathcal{F}_k], \tag{28}$$

By Assumption 3(a) we have

$$\alpha_{k,\max}\theta_{k,\max} \le \left(1+2\alpha_{k,\max}^2 L^2\right)\alpha_{k-1,\min}\theta_{k-1} \le \left(1+2\alpha_{k,\max}^2 L^2 + 2(\alpha_{k,\max}-\alpha_{k,\min})L\right)\alpha_{k-1,\min}\theta_{k-1}.$$

Using this, moving the term $-\alpha_{k,\min}\theta_{k,\min}\|x^k-x^*\|^2$ on the other side of inequality (28), and noting that

$$\frac{2\alpha_{k,\max}^2\theta_{k,\max}^2}{\alpha_{k,\min}\theta_{k,\min}} \le d \qquad \text{for some } d\in(0,1) \text{ and for } k\ge K,$$

with sufficiently large $K$ (since $\frac{2\alpha_{k,\max}^2\theta_{k,\max}^2}{\alpha_{k,\min}\theta_{k,\min}} \to 2c$ with $2c<1$ by Assumption 3(a)), we further see that for $k \ge K$,

$$\mathbb{E}[\|x^{k+1}-x^*\|^2 \mid \mathcal{F}_k] + \alpha_{k,\min}\theta_{k,\min}\|x^k-x^*\|^2$$
$$\le \left(1+2\alpha_{k,\max}^2 L^2 + 2(\alpha_{k,\max}-\alpha_{k,\min})L\right)\left(\|x^k-x^*\|^2 + \alpha_{k-1,\min}\theta_{k-1,\min}\|x^{k-1}-x^*\|^2\right)$$
$$-\alpha_{k,\min}\theta_{k,\min}(1-d)\|x^k-x^{k-1}\|^2 - 2\alpha_{k,\min}(x^k-x^*)^T(F(x^k)-F(x^*)) + \alpha_{k,\max}^2\mathbb{E}[\|w^k\|^2 \mid \mathcal{F}_k]. \tag{29}$$

It remains to show that the sequence $\{\|x^{k+1}-y^k\|\}$ converges to zero. This can be done by applying Lemma 2 to relation (29) with the following identification

$$V_k = \|x^k-x^*\|^2 + \alpha_{k-1,\min}\theta_{k-1,\min}\|x^k-x^*\|^2, \qquad u_k = 1+2\alpha_{k,\max}^2 L^2 + 2(\alpha_{k,\max}-\alpha_{k,\min})L,$$
$$\gamma_k = \alpha_{k,\min}\theta_{k,\min}(1-d)\|x^k-x^{k-1}\|^2 + 2\alpha_{k,\min}(x^k-x^*)^T(F(x^k)-F(x^*)), \qquad \beta_k = \alpha_{k,\max}^2\mathbb{E}[\|w^k\|^2 \mid \mathcal{F}_k].$$

To use the lemma, we need to verify that $\gamma_k \ge 0$, $\sum_{k=0}^\infty u_k < \infty$ and $\sum_{k=0}^\infty \beta_k < \infty$. Note that $\gamma_k > 0$ for all $k \ge K$ since $d \in (0,1)$ and $F$ is monotone. The condition $\sum_{k=0}^\infty u_k < \infty$ holds by our assumption that $\sum_{k=0}^\infty \alpha_{k,i}^2 < \infty$ for all $i$ (Assumption 3(b)), while $\sum_{k=0}^\infty \beta_k < \infty$ holds by Assumption 3(d). Thus, according to Lemma 2 (that holds for all $k$ large enough) we have for the solution $x^*$,

$$\{\|x^k-x^*\|^2 + \alpha_{k-1,\min}\theta_{k-1,\min}\|x^{k-1}-x^*\|^2\} \text{ converges almost surely,} \tag{30}$$
$$\sum_{k=0}^\infty \alpha_{k,\min}(x^k-x^*)^T(F(x^k)-F(x^*)) < \infty \text{ almost surely.} \tag{31}$$

Relation (30) implies that the sequence $\{x^k\}$ is almost surely bounded, so it has accumulation points almost surely. Since $K$ is closed and $\{x^k\} \subset K$, it follows that all the accumulation points of $\{x^k\}$ belong to $K$. By (31) and the relation $\sum_{k=0}^\infty \alpha_{k,\min} = \infty$ (see Assumption 3(b)) it follows that $(x^k-x^*)^T(F(x^k)-F(x^*)) \to 0$ along a subsequence almost surely. This and strict monotonicity of $F$ imply that $\{x^k\}$ has one accumulation point, say $\tilde{x}$, that must coincide with the solution $x^*$. By relation (30) it follows that the whole sequence must converge to the random point $\tilde{x}$ almost surely. ■

Consider now the case when we have uniformity in user stepsize and prox-parameter. Precisely, let each user $i$ implement the following update rule:

$$x_i^{k+1} = \Pi_{K_i}[x_i^k - \alpha_k(F_i(x^k) + \theta_k(x_i^k - x_i^{k-1}) + w_i^k)], \tag{IPP}$$

where $\theta_k > 0$ is the prox-parameter and $\alpha_k$ is a stepsize for all players at iteration $k$. We refer to this version of the method as Iterative Proximal Point (PPT) algorithm, to differentiate it from its partially-coordinated version (PITR) where the users have some freedom in selecting the parameters.

Almost sure convergence of the sequence $\{x^k\}$ generated using (IPP) can be obtained as a corollary of Proposition 3.

**Corollary 2.** *Let Assumption 1 hold, where $F$ is strictly monotone. Assume that $SOL(K,F)$ is nonempty. Also, let the steplengths and the prox-parameters satisfy Assumption 3 with $\alpha_{k,i} = \alpha_k$ and $\theta_{k,i} = \theta_k$ for all $i$. Then, the sequence $\{x^k\}$ generated by method (IPP) converges almost surely to the solution of $VI(K,F)$.*

Now, we note that the result of Proposition 3 holds if the stepsize and prox-parameter sequences satisfy Assumption 3 for all $k \geq K$ where $K$ is some positive integer. We next discuss some examples for choices of stepsize sequence $\{\alpha_{k,i}\}$ and prox-parameter sequence $\{\theta_{k,i}\}$ that satisfy Assumption 3(a)–(c) for sufficiently large indices $k$. Let

$$\alpha_{k,i} = (k+\eta_i)^{-a} \qquad \text{and} \qquad \theta_{k,i} = (k+\eta_i)^{-b},$$

for some scalars $a$ and $b$ such that $a \in (1/2, 1]$ and $a+b > 0$. The scalars $\eta_i$ are random with uniform distribution over an interval $[-\eta, \eta]$ for some $\eta > 0$. Then, Assumption 3(a) holds if

$$\frac{\alpha_{k,\max}\theta_{k,\max}}{\alpha_{k-1,\min}\theta_{k-1,\min}} \leq 1 + 2\alpha_{\max,k}^2 L^2. \tag{32}$$

Also letting $\eta_{\max} = \max_{1\leq i\leq N}\{\eta_i\}$ and $\eta_{\min} = \min_{1\leq i\leq N}\{\eta_i\}$, we have that

$$\frac{\alpha_{k,\max}\theta_{k,\max}}{\alpha_{k-1,\min}\theta_{k-1,\min}} = \frac{(k+\eta_{\min})^{-(a+b)}}{(k-1+\eta_{\max})^{-(a+b)}}.$$

Thus,

$$\lim_{k\to\infty} \frac{\alpha_{k,\max}\theta_{k,\max}}{\alpha_{k-1,\min}\theta_{k-1,\min}} = 1,$$

implying that relation (32) holds for some sufficiently large $k$.

We now consider the limit in the second part of Assumption 3(a). We have

$$\lim_{k\to\infty} \frac{\alpha_{k,\max}^2\theta_{k,\max}^2}{\alpha_{k,\min}\theta_{k,\min}} = \lim_{k\to\infty} \frac{(k+\eta_{\min})^{-2(a+b)}}{(k+\eta_{\max})^{-(a+b)}} = 0,$$

where the zero-limit follows by $a+b>0$. The conditions of Assumption 3(b) hold trivially for $a \in (1/2,1]$. For the condition of Assumption 3(c), we have

$$\begin{aligned}\alpha_{k,\max}-\alpha_{k,\min} &= (k+\eta_{\min})^{-a}-(k+\eta_{\max})^{-a} = (k+\eta_{\max})^{-a}\left(\frac{(k+\eta_{\min})^{-a}}{(k+\eta_{\max})^{-a}}-1\right)\\ &= (k+\eta_{\max})^{-a}\left(\left(1-\frac{\eta_{\max}-\eta_{\min}}{k+\eta_{\max}}\right)^{-a}-1\right)\\ &\approx (k+\eta_{\max})^{-a}\left(1+a\frac{\eta_{\max}-\eta_{\min}}{k+\eta_{\max}}+O(1/k^2)-1\right) = O(1/k^{1+a}),\end{aligned}$$

which is summable for $a \geq 0$.

## V. Case Study

In this section, we examine the sensitivity performance of the proposed algorithms to algorithm parameters. We also compare their performance to the performance of their standard counterparts (involving two-nested processes) and sample-average approximation schemes for a stochastic noncooperative rate allocation game. In Section V-A, we describe the player payoffs and strategy sets as well as the network constraints employed in the case study. In Section V-B, we discuss the sensitivity of the schemes to changes in algorithm parameters. Sections V-C and V-D compare the performance of our algorithms with their standard (two-loop) counterparts as well as sample-average approximation methods.

In what follows, we use ITR to refer to the iterative Tikhonov algorithm (PITR) where the users implement coordinated stepsizes and the regularization terms (i.e., $\alpha_{k,i}=\alpha_k$ and $\varepsilon_{k,i}=\varepsilon_k$ for all $i$).

### A. Network and user data

We consider a spatial network as given in Fig. 1. Suppose there are $N$ selfish users that compete over the network. Each user is characterized by a user-specific utility and faces a congestion cost that is a function of the aggregate flow in a link. Such a problem captures traffic and communication networks where the congestion cost may manifest itself through link-specific delays [8], [12]. The $i$th user's cost function $f_i(x_i,\xi_i,\omega_i)$ is a function of flow decisions $x_i$ and is parameterized by the uncertainty, denoted by $(\xi_i,\omega_i)$. It is defined as

$$f_i(x_i,\xi_i,\omega_i) \triangleq -\xi_i \log(1+x_i+\omega_i). \tag{33}$$

Each user selects an origin-destination pair of nodes on this network and faces congestion based on the links traversed along the prescribed path connecting the selected origin-destination nodes. We assume that the network links are indexed by an index set $\mathscr{L}$ and we consider the congestion cost of the form:

$$c(x,\varsigma) = \varsigma \sum_{i=1}^{N}\sum_{l\in\mathscr{L}} x_{li}\left(\sum_{j=1}^{N} x_{lj}\right), \tag{34}$$

where $x_{lj}$ denotes the flow of user $j$ on link $l$ and $\varsigma$ is a random congestion scaling parameter. The total cost of the network users is given by

$$f(x,\xi,\omega,\varsigma) = \sum_{i=1}^{N} f_i(x_i,\xi_i,\omega_i)+c(x,\varsigma) = \sum_{i=1}^{N} -\xi_i\log(1+x_i+\omega_i)+\varsigma\sum_{i=1}^{N}\sum_{l\in\mathscr{L}} x_{li}\left(\sum_{j=1}^{N} x_{lj}\right),$$

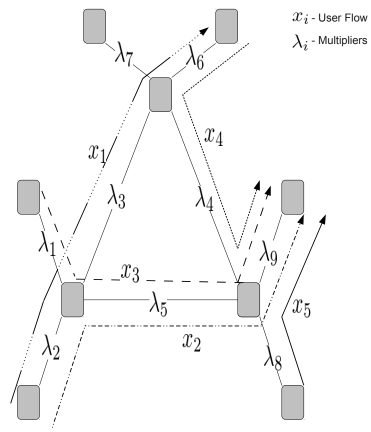


Fig. 1. A network with 5 users and 9 links.

where $\xi = (\xi_1, \ldots, \xi_N)^T$ and $\omega = (\omega_1, \ldots, \omega_N)^T$. Let $A$ denote the adjacency matrix that specifies the set of links traversed by the traffic generated by a particular user. More precisely, for every link $l \in \mathscr{L}$ and user $i$, we have $A_{li} = 1$ if link $l$ carries flow of user $i$, and $A_{li} = 0$ otherwise. Throughout this section, we consider the network with 9 links and 5 users, as given in Fig. 1. The simulation results are reported in tables, where we use $U(t, \tau)$ to denote the uniform distribution over an interval $[t, \tau]$ for $t < \tau$.

Table I summarizes the traffic in the network as generated by the users and provides the uniform distribution for the parameters $k_i(\xi_i)$ and noise $\omega_i$ of the user objectives. In addition we assume that congestion scaling parameter is $\varsigma \sim U(1/2, 1)$.

TABLE I
NETWORK AND USER DATA

| User($i$) | Links traversed | $\xi_i$ | $\omega_i$ |
|---|---|---|---|
| 1 | L2, L3, L6 | $U(0,10)$ | $U(0,1)$ |
| 2 | L2, L5, L9 | $U(0,10)$ | $U(0,1)$ |
| 3 | L1, L5, L9 | $U(0,10)$ | $U(0,1)$ |
| 4 | L6, L4, L9 | $U(0,10)$ | $U(0,1)$ |
| 5 | L8, L9 | $U(0,10)$ | $U(0,1)$ |

The user traffic rates are coupled through an expected-value constraint of the form $\sum_{i=1}^N A_{li} x_i \leq \mathbb{E}[C_l(\zeta_l)]$ for all $l \in \mathscr{L}$ where $C_l(\zeta_l)$ is the random aggregate traffic through link $l$. The constraint can be compactly written as $Ax \leq \mathbb{E}[C(\zeta)]$, where $\zeta = (\zeta_1, \ldots, \zeta_9)^T$ and $C(\zeta)$ is the random link capacity vector with $C(\zeta) \sim U(\bar{C} - 1, \bar{C} + 1)$ and $\bar{C} = (10, 15, 20, 10, 15, 20, 20, 15, 25)$.

We now articulate the stochastic variational inequality corresponding to this stochastic rate allocation game. Since, the strategy sets are coupled by a set of shared constraints, the associated game is a generalized Nash game with shared constraints. While the entire set of equilibria of this game are captured by a quasi-variational inequality, a subset of equilibria corresponding to a equilibria with common Lagrange multipliers associated with the shared constraint is given by a solution to a variational inequality. Given a price tuple $\lambda \triangleq (\lambda_l)_{l=1}^L$ such that $\lambda \in \mathbb{R}_+^L$, the payoff function of player $i$ is defined as

$$\mathbb{E}[f_i(x_i,\xi_i,\omega_i)+c(x,\varsigma)-\lambda^T(A_{\cdot,i}x_i-C(\zeta)],$$

where the expectation is with respect to $\xi_i$, $\omega_i$, $\varsigma$ and $\zeta$, and $A_{\cdot,i}$ denotes the $i$th column of the matrix $A$. **Uday - how is $C(\zeta)$ distributed among users??** Thus, the preceding can be viewed as a Lagrangian function $\mathbb{L}(x,\lambda)$ defined as:

$$\mathbb{L}(x,\lambda) \triangleq \mathbb{E}[\mathcal{L}(x,\lambda,\varphi)] \triangleq \left(\mathbb{E}[f_i(x_i,\xi_i,\omega_i)+c(x,\varsigma)-\lambda^T(A_{\cdot,i}x_i-C(\zeta))]\right)_{i=1}^N,$$

where $\varphi:\Omega\to\mathbb{R}^{2N+1+L}$ accounts for all the uncertainties. If we let $z=(x,\lambda)$ then a pair $z^{\rm NE}=(x^{\rm NE},\lambda^{\rm NE})$ solves the Nash equilibrium problem if $x^{\rm NE}\in \mathrm{SOL}(K\times\mathbb{R}_+^L,\Phi)$ where

$$K=K_1\times\cdots\times K_N \quad \text{with} \quad K_i=\mathbb{R}_+^{n_i} \quad \text{for each } i=1,\ldots,N$$

and the mapping $\Phi$ is defined as:

$$\Phi(z)=\Phi(x,\lambda)\triangleq(\nabla_x\mathbb{E}[\mathcal{L}(x,\lambda,\varphi)],-\nabla_\lambda\mathbb{E}[\mathcal{L}(x,\lambda,\varphi)]).$$

We now describe our experimental setup. Unless mentioned otherwise, we terminate each of our simulations after 10,000 iterations and obtain 95% confidence intervals by using 100 sample-paths. We report the confidence intervals at a 95% level for the normed error between the terminating iterate and optimal solution, i.e., the confidence interval for $\|x^k-x^*\|$ where $k=10,000$. For the ITR scheme, the update rule for the regularization parameter is taken to be of the form $\varepsilon_k=(1000+k)^{-a}$ where $k\geq 1$ is the current iterate. For the IPP scheme, the proximal parameter is updated using $\theta_k=(1000+k)^c$ where $k\geq 1$. The steplength $\alpha_k$ updated as $\alpha_k=(1000+k)^{-b}$, is chosen to be the same for both ITR and IPP schemes. In both partially coordinated schemes, we let $\alpha_{k,i}=(1000+k+\delta_i)^{-b}$ with $\varepsilon_{k,i}=(1000+k+\delta_i)^{-a}$ for PITR and $\theta_{k,i}=(1000+k+\delta_i)^{-b}$ for PIPP, where $\delta_i\sim U(-500,500)$. Finally, the optimal value $x^*$ of the network problem of minimizing $\mathbb{E}[f(x,\xi,\omega,\varsigma)]$ subject to $x\geq 0$ and $Ax\leq\mathbb{E}[C(\zeta)]$ is computed by solving a sample-average problem using the nonlinear programming solver `knitro` [45] on Matlab 7. For a sample size of 200, with 2000 replication for each sample the norm of the optimal value of $x^*$ was found to be $\|x^*\|=0.808$. We now briefly summarize our numerical algorithm for users iterate update.

1) For each sample path, at the beginning of each iteration, we draw a random sample of $\xi$, $\omega$, $\varsigma$ and $C(\zeta)$ from their respective distributions.
2) Using this random sample and its current iterate, each user gets a sample of a gradient.
3) The next iterate is generated using the proposed algorithm with the sampled gradient combined with appropriate stepsize and regularization parameter.
4) Steps 1–3 are repeated for $K$ iterations at the end of which the error from the optimal $\|x^K-x^*\|$ is recorded for a particular sample.
5) Mean error and a 95% confidence interval is reported for $S$ samples.

### *B. Sensitivity to parameters*

We considered cases when the regularization parameter sequence is driven to zero at different rates. Specifically, we choose $\epsilon_k$ as $(1000+k)^{-a}$ for ITR and for $i = 1,\ldots,N$, $\epsilon_{k,i}$ as $(1000+k+\delta_i)^{-a}$ where $\delta_i \sim U(-500,500)$ for PITR. The user stepsizes $\alpha_k$ are set to $(1000+k)^{-0.54}$ for ITR and $\alpha_{k,i} = (1000+k+\delta_i)^{-0.54}$ for PITR. Table II compares the 95% confidence interval for normed error $\|x^k - x^*\|$ of ITR method to that of PITR method, as a function of the parameter $a$ of the regularization stepsizes $\epsilon_k$ and $\epsilon_{k,i}$. It can be seen that as $a$ increases, the confidence intervals tend to be tighter upon termination. Furthermore, we also observe that when users choose their steplengths independently, the resulting confidence intervals appear to be slightly better.

TABLE II
VARYING $a$ IN REGULARIZATION TERM OF THE FORM $k^{-a}$ FOR A FIXED CHOICE FOR STEPSIZES

| Width of Confidence Intervals | | |
|---|---|---|
| $a$ | ITR | PITR |
| 0.25 | 1.29e − 02 | 1.09e − 02 |
| 0.30 | 1.19e − 02 | 1.08e − 02 |
| 0.35 | 1.15e − 02 | 1.07e − 02 |
| 0.40 | 1.14e − 02 | 1.07e − 02 |
| 0.45 | 1.10e − 02 | 1.06e − 02 |

We next examine the performance of iterative proximal-point methods. Table III compares the performance of IPP and PIPP methods when the rate of decay or growth of the prox parameter is varied under the assumption that users steplength update rule is fixed. Specifically, we let $\theta_k = (1000+k)^c$ with $\alpha_k = (1000+k)^{-0.54}$ for IPP, and $\theta_{k,i} = (1000+k+\delta_i)^c$ with $\alpha_k = (1000+k+\delta_i)^{-0.54}$ and $\delta_i \sim U(-500,500)$ for PIPP. Note that when $c > 0$ we have $\theta \nearrow \infty$. No clear relationship can be observed between $c$ (rate control parameter) and the recorded accuracy upon termination though it may seem that letting $c > 0$ results in a slightly better accuracy. Furthermore, we note that limited coordination has minimal impact on the obtained confidence intervals.

TABLE III
VARYING $c$ IN PROX-PARAMETER OF THE FORM $k^{-c}$ FOR A FIXED CHOICE OF THE STEPSIZES

| Width of Confidence Intervals | | |
|---|---|---|
| $c$ | IPP | PIPP |
| -0.35 | 1.23e − 02 | 1.13e − 02 |
| -0.15 | 1.18e − 02 | 1.04e − 02 |
| 0 | 1.16e − 02 | 1.17e − 02 |
| 0.15 | 9.46e − 03 | 1.07e − 02 |
| 0.35 | 1.04e − 02 | 9.62e − 02 |

Now, we examine the behavior of ITR and IPP methods by changing parameters that are common to both methods.

1) *Impact of steplength, regularization and proximal-parameter sequences:* We begin by examining the impact of varying the rate at which stepsize $\alpha_k$ decays to zero in both ITR and IPP schemes in terms of width of confidence interval and also report the computational time to achieve the desired level of accuracy. The table to the right in Table IV compares the width of confidence interval of both schemes. It can be observed that as the decay rates increase, IPP performs slightly better than ITR in terms of accuracy upon termination. Since there is relatively limited impact on ITR schemes from changing the decay rate of the regularization parameter (see Table II), given that $a+b<1$ choosing $b$ as close to 1 appears to be advantageous. Note the slight change in the update rule of $\varepsilon_k$ so as to accommodate a bigger range of variability for $b$. In Table IV (right), we list the computation times required by ITR and IPP to achieve the corresponding level of accuracy of right table in Table IV. Combining the results of Table IV with that of Table II it can be concluded that for iterative Tikhonov schemes it might be advantageous to chose a faster decay rate for $\alpha_k$.

TABLE IV
ITR VS. IPP: VARYING $b$ IN $\alpha_k = k^{-b}$ FOR FIXED $\varepsilon_k = (1000+k)^{-0.25}$ AND $\theta_k = (1000+k)^{0.35}$.

| Width of Confidence Intervals | | |
|---|---|---|
| $b$ | ITR | IPP |
| 0.54 | 1.16e − 02 | 1.03e − 02 |
| 0.59 | 8.89e − 03 | 8.88e − 03 |
| 0.64 | 7.20e − 03 | 7.36e − 03 |
| 0.69 | 7.08e − 03 | 5.53e − 03 |
| 0.74 | 4.84e − 03 | 4.48e − 03 |

| Computational Time in Seconds | | |
|---|---|---|
| $b$ | ITR | IPP |
| 0.54 | 164.71 | 128.76 |
| 0.59 | 165.98 | 129.47 |
| 0.64 | 165.95 | 129.62 |
| 0.69 | 165.85 | 129.67 |
| 0.74 | 165.94 | 129.54 |

2) *Impact of number of iterations:* Next, we examined the impact of changing the number of projection steps in each sample path. When conducting these tests, we choose the stepsize sequence to be $\alpha_k = (1000+k)^{-0.54}$ and $\alpha_{k,i} = (1000+k+\delta_i)^{-0.54}$ where $\delta_i \sim U(-500,500)$, in the fully and partially coordinated regimes, respectively. The regularization sequence $\varepsilon_k$ is specified as $(1000+k)^{-0.35}$ for ITR and $(1000+k+\delta_i)^{-0.35}$ for PITR while the prox parameter sequence $\theta_k$ is updated as $(1000+k)^{0.35}$ in the IPP and $(1000+k+\delta_i)^{0.35}$ in the PIPP schemes. Table V compares the performance of the ITR, PITR, IPP and PIPP. It is apparent that ITR performs slighty better than PITR while PIPP performs slightly better than IPP when the total number of iterations is smaller.

3) *Varying coordination requirements:* A worthwhile question in examining limited coordination generalizations is the extent to which disparity in steplength and parameter sequences impacts the overall confidence width. In Table VI, we tabulate the performance of varying the coordination amongst the user by changing the deviation in their individual steplengths and parameters which is controlled by changing the size of the support of the uniform distribution governing parameter $\delta_i$. Notably, our tests appear to show that within the range of testing conducted, there is relatively minor impact

TABLE V

PERFORMANCE OF VARIOUS SCHEMES FOR DIFFERENT LIMIT ON THE NUMBER OF ITERATIONS

| Number of | 95 % Confidence Interval | | | |
|---|---|---|---|---|
| Iterations | ITR | PITR | IPP | PIPP |
| 1000 | 1.67e-02 | 1.79e-02 | 1.69e-02 | 1.56e-02 |
| 2000 | 1.47e-02 | 1.71e-02 | 1.51e-02 | 1.33e-02 |
| 5000 | 1.42e-02 | 1.29e-02 | 1.15e-02 | 1.27e-02 |
| 10,000 | 1.02e-02 | 1.12e-02 | 1.01e-02 | 1.11e-02 |
| 20,000 | 8.41e-03 | 9.22e-03 | 7.90e-03 | 9.44e-03 |

associated with limited coordination.

TABLE VI

PERFORMANCE OF PITR AND PIPP FOR VARIOUS LEVEL OF COORDINATION $\delta_i$ WITH $\alpha_{k,i} = (1000+k+\delta_i)^{-0.54}$, $\varepsilon_{k,i} = (1000+k+\delta_i)^{-0.35}$ AND $\theta_{k,i} = (1000+k+\delta_i)^{0.35}$.

| Delta | 95 % Confidence Interval | |
|---|---|---|
| $\delta_i$ | PITR | PIPP |
| $U(-50,50)$ | 1.19e-02 | 9.08e-03 |
| $U(-100,100)$ | 1.08e-02 | 9.49e-03 |
| $U(-200,200)$ | 9.74e-03 | 1.07e-02 |
| $U(-500,500)$ | 1.14e-02 | 8.25e-03 |

### *C. Comparison with standard Tikhonov and proximal-point schemes*

Our schemes are motivated by the observation that regularization-based algorithms that rely on obtaining increasingly accurate solutions to a sequence of problems and such techniques cannot be easily extended to regimes where the subproblems are stochastic, particularly when relying on simulation-based methods. In particular, to improve the accuracy of an estimator by a factor of 10, requires a growth in replication size by a factor of 100. Naturally, getting solutions of increasing accuracy requires increasing replication lengths at a much faster rate, making such approaches computationally impractical. Iterative regularization schemes obviate this challenge by requiring a replication in which the regularization parameter is updated during the replication.

In this subsection, we detail the insights drawn from a rudimentary bounded complexity implementation of the standard Tikhonov and proximal-point schemes (with two-nested loops). In effect, the replication lengths are of fixed and unincreasing lengths and require *bounded complexity.* In effect, we obtain solutions of fixed accuracy and not increasing accuracy. More specifically, we examined the behavior of a Tikhonov regularization method where a sequence of subproblems was solved and the scheme was initiated with $\varepsilon = 1$ and was terminated when the regularization parameter $\varepsilon$ dropped below $(11000)^{-0.35}$ (terminating value of the base case of corresponding iterative Tikhonov method). Note that the regularized subproblem was solved via a simulation scheme in which the averaged solution over a fixed number of sample-paths was employed (100 in this case), each of which required $10,000$ steps. In the context of the proximal-point method, we terminate the scheme when the normed error for numerically obtained solution and actual solution drops below 1e-2. Note that both of these are heuristic termination criteria and other parameter setting were intended to get a flavor for the behavior of such schemes, as well as to make the comparison as fair as possible; more detailed termination criteria and experimental setup require using statistical validation techniques and remain a source of future research. It is worth commenting that these implementation of Tikhonov and proximal point differ from the exact theoretical implementation of Tikhonov and proximal point methods in the sense that theoretical methods requires *increasingly accurate* solutions to the subproblem while here we terminate after a fixed number of iterations. Table VII tabulates the performance for both methods in terms of width of confidence interval (table on the left) and the computational time required to reach the desired level of accuracy (table on the right). On comparing with corresponding data of iterative schemes in Table IV we notice that the both Tikhonov and proximal-point schemes display almost identical performance in terms of level of accuracy upon termination but require significant effort to do so especially Tikhonov methods.

TABLE VII

TIKHONOV V/S PROXIMAL POINT METHOD: VARYING $b$ IN $\alpha_k = k^{-b}$.

| Width of Confidence Intervals | | |
|---|---|---|
| $b$ | Tikhonov | Proximal Point |
| 0.54 | 1.58e − 02 | 1.62e − 02 |
| 0.59 | 1.30e − 02 | 1.38e − 02 |
| 0.64 | 1.15e − 02 | 1.03e − 02 |
| 0.69 | 8.57e − 03 | 8.10e − 03 |
| 0.74 | 6.85e − 03 | 7.27e − 03 |

| Computational Time in Seconds | | |
|---|---|---|
| $b$ | Tikhonov | Proximal Point |
| 0.54 | 297.62 | 183.08 |
| 0.59 | 360.30 | 138.10 |
| 0.64 | 360.32 | 137.62 |
| 0.69 | 360.36 | 137.56 |
| 0.74 | 360.26 | 137.46 |

### *D. Comparison with sample average approximation (SAA) techniques*

In this subsection, we investigate the performance of SAA methods with one of the candidate schemes namely, PIPP. We report a 95% confidence interval SAA obtained for a sample of size 100 for various replication levels to approximate the problem. For each sample, the replicate averaged proble is solved using `knitro` while PIPP is implemented with $\alpha_{k.i} = (1000 + k + \delta_i)^{-0.54}$ and $\theta_{k.i} = (1000 + k + \delta_i)^{0.35}$ where $\delta_i \sim U(-500, 500)$. Table VIII (Left) demonstrates the performance of SAA with accuracy level reached for various number of replication while Table VIII (Right) compares the computational effort required measured in time (seconds) by SAA to reach the accuracy level to that of PIPP. On comparing the width of confidence interval from Table V for PIPP it is apparent that SAA outperforms PIPP however the computational effort to reach such an accuracy is enormous. **Uday: I do not understand this last statement- Perhaps Table number is wrong.** Though the results of Table IV are for fully coordinated iterative methods a similar performance is expected for partially coordinated methods indicating that an almost similar performance to SAA can be achieved by increasing the decay rate of stepsize but at a much lower computation cost.

TABLE VIII
COMPARISON OF PERFORMACE FOR SAA AND PIPP

| Width of Confidence Intervals | |
|---|---|
| Number of Replications | SAA |
| 1000 | $3.86\mathrm{e}-03$ |
| 2000 | $2.32\mathrm{e}-03$ |
| 5000 | $1.65\mathrm{e}-03$ |

| Number of Replications or Iterations | Computational Time in Seconds | |
|---|---|---|
| | SAA | PIPP |
| 1000 | 1438.93 | 20.65 |
| 2000 | 3598.07 | 41.32 |
| 5000 | 14,645.32 | 103.27 |

**Also: what is the first column for table to the right- How do we interpret the column values**.

## VI. CONCLUDING REMARKS

This paper has proposed and investigated some algorithms for computing solutions to stochastic variational inequalities when the mappings are not necessarily strongly monotone. The paper is related to the past work by Jiang and Xu [19] who considered how stochastic approximation procedures could address stochastic variational inequalities with strongly monotone mappings. yet, these schemes cannot easily contend with weaker requirements (such as strict monotonicity or merely monotonicity) while retaining the single itera structure. Instead, a simple regularization-based extension leads to a two-level method, that is generally harder to implement in networked settings.

Accordingly, this paper makes the following contributions. First, we present *single-iterative* counterparts of standard Tikhonov and proximal-point methods that obviate the need to solve a sequence of subproblems. Instead, we present a stochastic iterative Tikhonov regularization scheme and a stochastic iterative proximal-point method in which the regularization parameter in the former and the centering parameter in the latter are updated at *every* iteration. Suitable conditions on the parameter sequences are established for guaranteeing the almost-sure convergence of the resulting schemes. Notably, the iterative proximal-point scheme also allows for raising the proximal-parameter at every step.

The paper concludes with a detailed study of the computational performance of these schemes on a networked monotone stochastic rate-allocation game. Through this case-study, we observe that the schemes perform better when the steplength sequences are driven to zero at a faster rate but are less sensitive to changing the regularization and proximal parameter sequences. Notably, partial coordination of steplength choices has minimal impact on the accuracy of the solution. Finally, naive implementations of standard Tikhonov and proximal-point schemes prove illuminating; while Tikhonov schemes provide accurate solutions at significant computational expense, proximal schemes seem to be less successful.

## REFERENCES


[1] J. Koshal, A. Nedić, and U. V. Shanbhag, “Single timescale regularized stochastic approximation schemes for monotone nash games under uncertainty,” *Proceedings of the IEEE Conference on Decision and Control (CDC)*, 2010.

[2] F.P. Kelly, A.K. Maulloo, and D.K. Tan, “Rate control for communication networks: shadow prices, proportional fairness, and stability,” *Journal of the Operational Research Society*, vol. 49, pp. 237–252, 1998.

[3] R. Srikant, *Mathematics of Internet Congestion Control*, Birkhauser, 2004.

[4] S. Low and D.E. Lapsley, “Optimization flow control, I: Basic algorithm and convergence,” *IEEE/ACM Transactions on Networking*, vol. 7, no. 6, pp. 861–874, 1999.

[5] E. Altman, T. Basar, T. Jiménez, and N. Shimkin, “Competitive routing in networks with polynomial cost,” in *INFOCOM (3)*, 2000, pp. 1586–1593.

[6] T. Alpcan and T. Başar, “Distributed algorithms for Nash equilibria of flow control games,” in *Advances in Dynamic Games*. 2003, vol. 7 of *Annals of the International Society of Dynamic Games*, pp. 473–498, Birkhäuser Boston.

[7] D. Dutta, A. Goel, and J. Heidemann, “Oblivious aqm and nash equilibria,” *INFOCOM 2003. Twenty-Second Annual Joint Conference of the IEEE Computer and Communications Societies. IEEE*, vol. 1, pp. 106–113 vol.1, March-3 April 2003.

[8] T. Alpcan and T. Başar, “A game-theoretic framework for congestion control in general topology networks,” in *41th IEEE Conf. Decision and Control*, Las Vegas, NV, December 2002.

[9] M. Bloem, T. Alpcan, and T. Başar, “A stackelberg game for power control and channel allocation in cognitive radio networks,” in *ValueTools ’07: Proceedings of the 2nd international conference on Performance evaluation methodologies and tools*, ICST, Brussels, Belgium, Belgium, 2007, pp. 1–9, ICST (Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering).

[10] L. Pavel, “A noncooperative game approach to OSNR optimization in optical networks,” *IEEE Transactions on Automatic Control*, vol. 51, no. 5, pp. 848–852, 2006.

[11] Y. Pan and L. Pavel, “Games with coupled propagated constraints in optical network with multi-link topologies,” *Automatica*, vol. 45, pp. 871–880, 2009.

[12] H. Yin, U. V. Shanbhag, and P. G. Mehta, “Nash equilibrium problems with congestion costs and shared constraints,” *To appear in IEEE Transactions of Automatic Control (July 2011)*, 2009.

[13] G. Scutari, F. Facchinei, J-S. Pang, and D.P. Palomar, “Monotone games in communications: Theory, algorithms, and models,” *submitted to IEEE Transactions on Information Theory*, April 2010.

[14] A. Rantzer, “Using game theory for distributed control engineering,” Tech. Rep. ISRN LUTFD2/TFRT--7620--SE, Department of Automatic Control, Lund University, Sweden, July 2008, Presented at Games 2008, 3rd World Congress of the Game Theory Society.

[15] L. Pavel, “An extension of duality to a game-theoretic framework,” *Automatica*, vol. 43, pp. 226–237, 2007.

[16] F. Facchinei and J.-S. Pang, *Finite-Dimensional Variational Inequalities and Complementarity Problems*, vol. 1 and 2, Springer-Verlag Inc., New York, 2003.

[17] C. H. Papadimitriou and M. Yannakakis, “On bounded rationality and computational complexity,” Tech. Rep., Indiana University, 1994.

[18] H. A. Simon, *The sciences of the artificial (3rd ed.)*, MIT Press, Cambridge, MA, USA, 1996.

[19] H. Jiang and H. Xu, “Stochastic approximation approaches to the stochastic variational inequality problem,” *IEEE Transactions on Automatic Control,*, vol. 53, no. 6, pp. 1462–1475, 2008.

[20] H. Robbins and S. Monro, “A stochastic approximation method,” *The Annals of Mathematical Statistics*, vol. 22, no. 3, pp. 400–407, 1951.

[21] Y. Ermoliev, *Stochastic programming methods*, Nauka, Moscow, 1976.

[22] Y. Ermoliev, “Stochastic quasi-gradient methods and their application to system optimization,” *Stochastics*, vol. 9, no. 1, pp. 1–36, 1983.

[23] Y. Ermoliev, “Stochastic quazigradient methods,” in *Numerical Techniques for Stochastic Optimization*, pp. 141–186. Springer-Verlag, N.Y., 1988.

[24] D.P. Bertsekas and J.N. Tsitsiklis, “Gradient convergence in gradient methods,” *SIAM Journal on Optimization*, vol. 10, no. 3, pp. 627–642, 2000.

[25] V.S. Borkar, *Stochastic Approximation: A Dynamical Systems Viewpoint*, Cambridge University Press, 2008.

[26] S.S. Ram, A. Nedić, and V.V. Veeravalli, “Incremental stochastic sub-gradient algorithms for convex optimization,” *SIAM Journal on Optimization*, vol. 20, no. 2, pp. 691–717, 2009.

[27] A. Shapiro, “Monte carlo sampling methods,” in *Handbook in Operations Research and Management Science*, vol. 10, pp. 353–426. Elsevier Science, Amsterdam, 2003.

[28] A. Shapiro and T. Homem de Mello, “A simulation-based approach to two-stage stochastic programming with recourse,” *Mathematical Programming: Series A*, vol. 81, no. 3, pp. 301–325, 1998.

[29] J. Linderoth, A. Shapiro, and S.J. Wright, “The empirical behavior of sampling methods for stochastic programming,” *Ann. Oper. Res.*, vol. 142, pp. 215–241, 2006.

[30] H. Xu, “Sample average approximation methods for a class of stochastic variational inequality problems,” *Asia-Pacific Journal of Operational Research (APJOR)*, vol. 27, no. 01, pp. 103–119, 2010.

[31] A.J. King and R.T. Rockafellar, “Asymptotic theory for solutions in statistical estimation and stochastic programming,” *Mathematics of Operations Research*, 1993.

[32] W.K. Mak, D.P. Morton, and R.K. Wood, “Monte carlo bounding techniques for determining solution quality in stochastic programs,” *Operations Research Letters*, vol. 24, pp. 47–56, 1999.

[33] G. Gurkan, A.Y. Ozge, and S.M. Robinson, “Sample-path solution of stochastic variational inequalities,” *Mathematical Programming*, vol. 84, pp. 313–333, 1999.

[34] U. V. Shanbhag, G. Infanger, and P. W. Glynn, “A complementarity framework for forward contracting under uncertainty,” *To appear in Operations Research*.

[35] A.N. Tikhonov and V. Arsénine, *Méthodes de resolution de problèmes mal posés*, Éditions Mir, Moscow, 1976, Traduit du russe par Vladimir Kotliar.

[36] A.N. Tikhonov, “On the solution of incorrectly put problems and the regularisation method,” in *Outlines Joint Sympos. Partial Differential Equations (Novosibirsk, 1963)*, pp. 261–265. Acad. Sci. USSR Siberian Branch, Moscow, 1963.

[37] B. Martinet, “Régularisation d’inéquations variationnelles par approximations successives,” *Rev. Française Informat. Recherche Opérationnelle*, vol. 4, no. Ser. R-3, pp. 154–158, 1970.

[38] B.T. Polyak, *Introduction to Optimisation*, Optimization Software, Inc., New York, 1987.

[39] R.T. Rockafellar, “Monotone operators and the proximal point algorithm,” *SIAM Journal on Control and Optimization*, vol. 14, no. 5, pp. 877–898, 1976.

[40] Yurii Nesterov, “Dual extrapolation and its applications to solving variational inequalities and related problems,” *Mathematical Programming*, vol. 109, no. 2, pp. 319–344, Mar. 2007.

[41] A. Nemirovski, “Prox-method with rate of convergence O(1/T) for variational inequalities with lipschitz continuous monotone operators and smooth convex-concave saddle point problems,” *SIAM Journal on Optimization*, vol. 15, pp. 229–251, 2004.

[42] I.V. Konnov, *Equilibrium Models and Variational Inequalities*, Elsevier, 2007.

[43] A. Kannan and U.V. Shanbhag, “Distributed iterative regularization algorithms for monotone Nash games,” *Proceedings of the IEEE Conference on Decision and Control (CDC)*, 2010.

[44] R.T. Rockafellar, “Augmented lagrangians and applications of the proximal point algorithm in convex programming,” *Mathematics of Operations Research*, vol. 1, no. 2, pp. 97–116, 1976.

[45] R. H. Byrd, M. E. Hribar, and J. Nocedal, “An interior point algorithm for large-scale nonlinear programming,” *SIAM Journal on Optimization*, vol. 9, no. 4, pp. 877–900 (electronic), 1999, Dedicated to John E. Dennis, Jr., on his 60th birthday.