



Stochastic Approximation for Estimating the Price of Stability in Stochastic Nash Games

AFROOZ JALILZADEH, University of Arizona, USA

FARZAD YOUSEFIAN, Rutgers University, USA

MOHAMMADJAVAD EBRAHIMI, Rutgers University, USA

The goal in this paper is to approximate the Price of Stability (PoS) in stochastic Nash games using stochastic approximation (SA) schemes. PoS is amongst the most popular metrics in game theory and provides an avenue for estimating the efficiency of Nash games. In particular, knowing the value of PoS can help with designing efficient networked systems, including transportation networks and power market mechanisms. Motivated by the absence of efficient methods for computing the PoS, first we consider stochastic optimization problems with a nonsmooth and merely convex objective function and a merely monotone stochastic variational inequality (SVI) constraint. This problem appears in the numerator of the PoS ratio. We develop a randomized block-coordinate stochastic extra-(sub)gradient method where we employ a novel iterative penalization scheme to account for the mapping of the SVI in each of the two gradient updates of the algorithm. We obtain an iteration complexity of the order ϵ^{-4} that appears to be best known result for this class of constrained stochastic optimization problems, where ϵ denotes an arbitrary bound on suitably defined infeasibility and suboptimality metrics. Second, we develop an SA-based scheme for approximating the PoS and derive lower and upper bounds on the approximation error. To validate the theoretical findings, we provide preliminary simulation results on a networked stochastic Nash Cournot competition.

CCS Concepts: • **Mathematics of computing** → **Continuous optimization**; • **Computing methodologies** → **Stochastic games**.

Additional Key Words and Phrases: stochastic optimization, variational inequality, Nash equilibrium, price of stability

1 INTRODUCTION

The goal in this paper lies in the development of a stochastic approximation method, equipped with performance guarantees, for computing the price of stability (PoS) ratio in monotone stochastic Nash games. Nash equilibrium (NE) is a fundamental concept in game theory and captures a wide range of phenomena in engineering, economics, and finance [12]. Consider a stochastic Nash game with N players, each associated with a strategy set $X_i \subseteq \mathbb{R}^{n_i}$ and a cost function f_i . Player i 's objective is to determine, for any collection of arbitrary strategies of the other players, denoted by $x^{(-i)}$, an optimal strategy $x^{(i)}$ that solves the stochastic minimization problem

$$\begin{aligned} & \underset{x^{(i)}}{\text{minimize}} && \mathbb{E} \left[f_i \left(\left(x^{(i)}; x^{(-i)} \right), \xi \right) \right], && (P_i(x^{(-i)})) \\ & \text{subject to} && x^{(i)} \in X_i, \end{aligned}$$

Authors' addresses: Afroz Jalilzadeh, afroz@arizona.edu, University of Arizona, Tucson, Arizona, USA, 85721; Farzad Yousefian, farzad.yousefian@rutgers.edu, Rutgers University, Piscataway, New Jersey, USA, 08854; Mohammadjavad Ebrahimi, me586@scarletmail.rutgers.edu, Rutgers University, Piscataway, New Jersey, USA, 08854.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

© 2023 Copyright held by the owner/author(s). Publication rights licensed to ACM.

1049-3301/2023/11-ART \$15.00

<https://doi.org/10.1145/3632525>

where $f_i((x^{(i)}; x^{(-i)}), \xi)$ denotes a random cost function associated with the i th player that is parameterized in terms of the action of the player $x^{(i)}$, actions of other players denoted by $x^{(-i)}$, and a random variable ξ , where $\xi : \Omega \rightarrow \mathbb{R}^d$ denotes a random variable associated with the probability space $(\Omega, \mathcal{F}, \mathbb{P})$.

Remark 1. Throughout, similar to [22, 31, 42], we focus on settings where the stochasticity is only present in the objective function of the players. In particular, we assume that the strategy sets are deterministic.

An NE is described as a collection of specific strategies chosen by all the players, denoted by the tuple $x \triangleq (x^{(1)}, \dots, x^{(N)})$ where no player can reduce her cost by unilaterally changing her strategy within her feasible strategy set. Mathematically, NE can be described as a vector x that satisfies, for all $i = 1, \dots, N$, the inequality given as

$$\mathbb{E} \left[f_i \left((x^{(i)}; x^{(-i)}), \xi \right) \right] \leq \mathbb{E} \left[f_i \left((y^{(i)}; x^{(-i)}), \xi \right) \right], \quad \text{for all } y^{(i)} \in X_i. \quad (1)$$

Suppose n denotes the total number of dimensions associated with an NE, i.e., $n \triangleq \sum_{i=1}^N n_i$. Let us define the set $X \subseteq \mathbb{R}^n$ as the Cartesian product of the players' strategy sets, i.e., $X \triangleq \prod_{i=1}^N X_i$. Also, under a differentiability assumption, define the stochastic mapping $F : \mathbb{R}^n \times \mathbb{R}^d \rightarrow \mathbb{R}^n$ and its deterministic counterpart $F : \mathbb{R}^n \rightarrow \mathbb{R}^n$ as the collection of players' gradient mappings as

$$F(x) \triangleq \mathbb{E}[F(x, \xi)], \quad \text{where } F(x, \xi) \triangleq (\nabla_{x^{(1)}} f_1(x, \xi), \dots, \nabla_{x^{(N)}} f_N(x, \xi)).$$

Note that for expository ease, we use F in naming both deterministic and stochastic mappings. Then, under the convexity of the players' objective functions, the problem of seeking an NE to the game characterized by problems $(P_i(x^{(-i)}))$ for $i = 1, \dots, N$, can be compactly formulated as a stochastic variational inequalities (VI) problem, denoted by $\text{VI}(X, F)$. Recall that a vector $x^* \in X$ solves $\text{VI}(X, F)$ if $(y - x^*)^T F(x^*) \geq 0$, for all $y \in X$. Indeed, it can be observed that the inequality above compactly captures the optimality conditions of the convex programs (1) written for all $i = 1, \dots, N$. To this end, computing a solution to $\text{VI}(X, F)$ leads to finding an NE to the described stochastic Nash game. Generally, a VI problem may admit multiple solutions leading to a collection of NEs. Throughout, we let $\text{SOL}(X, F)$ denote the solution set of the $\text{VI}(X, F)$. In this paper, our aim is to develop a provably convergent scheme for estimating the efficiency in stochastic Nash games with monotone mappings. The notion of efficiency in Nash games is a storied area of research and dates back to the celebrated Prisoner's Dilemma. In fact, Nash equilibrium is provably known to be inefficient [11], in the sense that the competition among the players often leads to a degradation of the overall performance of the system of players. In view of this, understanding the efficiency of an NE has received much attention in game theory. Among, the popular measures of the efficiency of NE is a metric called *price of stability* (PoS) [35]. Given an arbitrary cost metric for quantifying the overall performance of the system, PoS is defined as the ratio between the following two quantities: (1) the minimal cost attained by the best Nash equilibrium (among possibly many NEs); (2) the optimal cost when the competition among the players is (hypothetically) suppressed. Let stochastic function $f : \mathbb{R}^n \times \mathbb{R}^d \rightarrow \mathbb{R}$ denote the system's overall performance metric. Mathematically and following our notation, PoS can be formulated as

$$\text{PoS} \triangleq \frac{\min_{x \in \text{SOL}(X, \mathbb{E}[F(\bullet, \xi)])} \mathbb{E}[f(x, \xi)]}{\min_{x \in X} \mathbb{E}[f(x, \xi)]}. \quad (2)$$

Remark 2. We note that the function f may or may not relate to the individual objective functions of the players denoted by f_i . In the literature [1, 21], different choices have been considered. Two common examples include the *utilitarian* approach where f is defined as the summation of all players' objectives, and the *egalitarian* approach where f is defined as the maximum of the individual objective functions.

Evaluating the PoS ratio, even in deterministic problems, is a computationally challenging task. To elaborate on this, we provide a simple example in the following.

Example (PoS in saddle-point problems). The problem of seeking a saddle-point in minmax optimization is an important class of equilibrium problems that has received considerable attention in game theory [12, 27, 30, 31] and more recently, in adversarial learning [13], fairness in machine learning [38], and distributionally robust federated learning [10]. In fact, the canonical minmax problem can be viewed as a subclass of two-person zero-sum games. The existence of equilibrium in such a game was established by the von Neumann's minmax theorem in 1928 [37]. To elaborate, consider a minmax problem given as

$$\min_{11 \leq x_1 \leq 60} \max_{10 \leq x_2 \leq 50} \phi(x_1, x_2) \triangleq 20 - 0.1x_1x_2 + x_1. \quad (3)$$

Figure 1 shows the saddle-shaped function ϕ . Associated with problem (3), we can consider a pair of optimization problems as

$$\begin{cases} \text{minimize}_{x_1} & f_1(x_1, x_2) \triangleq 20 - 0.1x_1x_2 + x_1 \\ \text{subject to} & x_1 \in X_1 \triangleq [11, 60], \end{cases} \quad (4) \quad \begin{cases} \text{minimize}_{x_2} & f_2(x_1, x_2) \triangleq -20 + 0.1x_1x_2 - x_1 \\ \text{subject to} & x_2 \in X_2 \triangleq [10, 50]. \end{cases} \quad (5)$$

Problems (4) and (5) together construct a two-person zero-sum Nash game. From [12, 1.4.2 Proposition], the set of saddle-points are the solutions to the variational inequality problem $\text{VI}(X, F)$ where we define

$$F(x_1, x_2) \triangleq (\nabla_{x_1} f_1(x), \nabla_{x_2} f_2(x)) = (-0.1x_2 + 1, 0.1x_1) \quad \text{and} \quad X \triangleq X_1 \times X_2.$$

Note that the mapping F is merely monotone, in view of $(F(x) - F(y))^T(x - y) = 0$ for all $x \in \mathbb{R}^2$ and $y \in \mathbb{R}^2$. We observe that the set of all the saddle-points is given by $\text{SOL}(X, F) = \{(x_1, x_2) \mid x_1 \in [11, 60], x_2 = 10\}$, implying that there are infinitely many Nash equilibria to this game characterized by the convex set $\text{SOL}(X, F)$. To measure the PoS, let us consider the global metric defined as $f(x_1, x_2) \triangleq 20 + |x_1 - x_2|$ for instance. This implies that the numerator of the PoS in (2) is equal to 21, while its denominator is equal to 20. As such, we obtain $\text{PoS} = 1.05$, implying that the competition in the game leads to an %5 loss in the metric f . Although in this simple example, we are able to evaluate the PoS, in practice, we often encounter several challenges that may make this impossible. Two main challenges are explained as follows: (i) The solution set of the VI is often unknown. Even in deterministic settings, it is often impossible to determine the entire set $\text{SOL}(X, F)$; (ii) Nash games might be afflicted by the presence of uncertainty which motivates the need for leveraging Monte Carlo sampling schemes, such as stochastic approximation, for contenting with stochasticity and the large-scale of the problem. For example, in distributionally robust federated learning [10], the problem is cast a stochastic minmax problem where the stochasticity emerges from the probability distribution of the local data sets, privately maintained by the clients.

To estimate the PoS with guarantees, first, we need to solve the numerator of the right-hand side of (2) that is characterized as a stochastic optimization with a stochastic VI constraint. Naturally, addressing the presence of VI constraints is a challenging task in optimization. This is mainly because VI constraints do not appear to lend themselves to standard Lagrangian relaxation schemes. In this work, this challenge is exacerbated due to the presence of uncertainty in the mapping of the VI constraint. To this end, our goal is to employ stochastic approximation (SA) schemes. SA is an iterative scheme that has been widely employed for solving problems in which the objective function is corrupted by a random noise. In the context of optimization problems, the

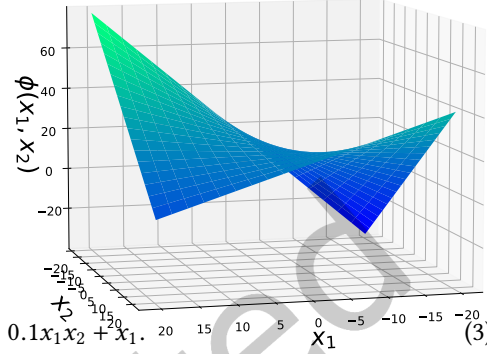


Fig. 1. Function ϕ in problem (3)

function values and/or higher-order information are estimated from noisy samples in a Monte Carlo simulation procedure [4]. The SA scheme, first introduced by Robbins and Monro [34], has been studied extensively in recent years for addressing stochastic optimization and stochastic variational inequality problems [22, 28, 33, 39].

In addressing constrained stochastic formulations, the majority of the SA schemes in the existing literature address the standard cases where the constraints are in the form of functional inequalities, equalities, or easy-to-project sets. However, motivated by the need for efficiency estimation in stochastic Nash games, we aim at devising a provably convergent SA method for estimation of the PoS. To this end, our primary interest lies in solving the following stochastic optimization problem whose constraint set is characterized as the solution set of a stochastic VI problem. This optimization problem is given as

$$\begin{aligned} & \text{minimize} && \mathbb{E}[f(x, \xi)] \\ & \text{subject to} && x \in \text{SOL}(X, \mathbb{E}[F(\bullet, \xi)]), \end{aligned} \tag{6}$$

where $f : \mathbb{R}^n \times \mathbb{R}^d \rightarrow \mathbb{R}$ is a convex function, $X \subseteq \mathbb{R}^n$ is the Cartesian product of the component sets $X_i \subseteq \mathbb{R}^{n_i}$ where $\sum_{i=1}^N n_i = n$, i.e., $X \triangleq \prod_{i=1}^N X_i$. We let the i th block-coordinate of the mapping $F(\bullet, \xi)$ be denoted by $F_i : \mathbb{R}^n \times \mathbb{R}^d \rightarrow \mathbb{R}^{n_i}$ for any $i \in [N] \triangleq \{1, \dots, N\}$. As noted earlier, for the ease of presentation, throughout we define $f(x) \triangleq \mathbb{E}[f(x, \xi)]$ and $F(x) \triangleq \mathbb{E}[F(x, \xi)]$.

Existing literature on VIs. The variational inequality problem has been extensively studied in the literature due to its versatility in capturing a wide range of problems including optimization, equilibrium and complementarity problems, amongst others [12]. The extra-gradient method, initially proposed by Korpelevich [27] and its extensions [5–7, 16, 22, 41, 43], is a classical method for solving VI problems which requires weaker assumptions than standard gradient schemes [2, 36]. In stochastic problems, amongst the earliest schemes for resolving stochastic variational inequalities via stochastic approximation was presented by Jiang and Xu [20] under the strong monotonicity and smoothness assumptions of the mapping. Regularized variants of SA schemes were developed by Koshal et al. [28] for addressing stochastic VIs with merely monotone mappings. Further, smoothness requirements were weakened by leveraging randomized smoothing in [40, 42]. In the absence of strong monotonicity, extra-gradient approaches that rely on two projections per iteration provide an avenue for resolving merely monotone problems [17]. The per-iteration complexity can be reduced to a single projection via projected reflected gradient and splitting techniques as examined in [8, 9] (also see [14]). When the assumption on the mapping is weakened to pseudomonotonicity and its variants, rate statements have been provided in [15, 24, 25] via a stochastic extra-gradient framework.

Gap in the literature. Despite these advances in addressing VIs and their stochastic variants, solving problem (6) remains challenging. In fact, we are unaware of any provably convergent stochastic approximation method for solving problem (6) that appears to be essential in estimating the PoS, defined as (2). One main approach to solve (6), when the constraint set is the solution set of a deterministic VI and the objective function is also deterministic, is the sequential regularization (SR) approach which is a two-loop framework (see [12, Chapter 12]). In each iteration of the SR scheme, a regularized VI is required to be solved and convergence has been shown under the monotonicity of the mapping F and closedness and convexity of the set X . However, the iteration complexity of the SR algorithm is unknown and it requires solving a series of increasingly more difficult VI problems. To resolve these shortcomings, recently, Kaushik and Yousefian [26] developed a more efficient first-order method called averaging randomized block iteratively regularized gradient. Non-asymptotic suboptimality and infeasibility convergence rates of $O(1/K^{0.25})$ have been obtained where K is the total number of iterations. Here, we consider a more general problem with a stochastic objective function and a stochastic VI constraint. Employing a novel iterative penalization technique, we propose an extra-(sub)gradient-based SA method and we derive convergence

results in expectation, of the same order of magnitude as in [26], despite the presence of stochasticity in the both levels of the problem.

Main contributions. In this paper, we study a stochastic optimization problem with a nonsmooth and merely convex objective function and a constraint set characterized as the solution set of a stochastic variational inequality problem. Motivated by the absence of efficient and scalable SA methods for addressing this class of constrained stochastic optimization problems, we develop a single-timescale first-order stochastic approximation method with block-coordinate updates, called Averaging Randomized Iteratively Penalized Stochastic Extra-Gradient Method (aR-IP-SeG). We derive convergence rates in terms of suitably defined metrics for suboptimality and infeasibility. In particular, in Theorem 1, we obtain an iteration complexity of the order of ϵ^{-4} where ϵ denotes a user-specified bound on both the objective function's error and a suitably defined infeasibility metric (i.e., dual gap function). This iteration complexity appears to be best known result for this class of constrained stochastic optimization problems. Moreover, utilizing the proposed extra-(sub)gradient-based method, we derive lower and upper bounds, both of the order $1/K^{0.25}$, for approximating the price of stability. Such guarantees appear to be new in computing the PoS.

Outline of the paper. Next, we introduce the notation that we use throughout the paper. In the next section, we precisely state the main definitions and assumptions that we need for the convergence analysis. In Section 2, we describe the aR-IP-SeG algorithm to solve problem (6) and the complexity analysis is provided in Section 4. Additionally, in Section 5, we propose a scheme to approximate the price of stability in (2) with guarantees. Finally, some empirical experiments are presented in Section 6 for addressing a stochastic Nash Cournot competition over a network where we compare our proposed scheme with the few existing schemes that can be employed for estimating the PoS.

Notation. Throughout, we often use column vectors. For a convex function $h : \mathbb{R}^n \rightarrow \mathbb{R}$ with the domain $\text{dom}(h)$ and any $x \in \text{dom}(h)$, a vector $\tilde{\nabla}h(x) \in \mathbb{R}^n$ is called a subgradient of h at x if $h(x) + \tilde{\nabla}h(x)^T(y - x) \leq h(y)$ for all $y \in \text{dom}(h)$. We let $\partial h(x)$ denote the subdifferential set of function h at x . Given a vector $x \in \mathbb{R}^n$, we use $x^{(i)} \in \mathbb{R}^{n_i}$ to denote its i th block-coordinate. We let $\tilde{\nabla}_i h(x)$ denote the i th block-coordinate of $\tilde{\nabla}h(x)$. We use similar notation for referring to the i th block-coordinate of mappings. We let $\mathbb{E}[\bullet]$ denote the expectation with respect to the all probability distributions under study. We use filtration to take conditional expectations with respect to a subgroup of probability distributions. We denote the optimal objective value of the problem (6) by f^* . The Euclidean projection of vector x onto a convex set X is denoted by $\mathcal{P}_X(x)$, where $\mathcal{P}_X(x) \triangleq \text{argmin}_{y \in X} \|y - x\|^2$. Throughout the paper, unless specified otherwise, k denotes the iteration counter while K represents the total number of steps employed in the proposed methods. Moreover, we define $\text{dist}(x, X) \triangleq \min_{y \in X} \|y - x\|$.

2 ALGORITHM OUTLINE

Our goal in this section is to devise an SA scheme for solving problem (6). To this end, we develop a method, called Averaging Randomized Iteratively Penalized Stochastic Extra-Gradient (aR-IP-SeG) presented by Algorithm 1. Compared with standard extra-gradient methods, a key novelty in the design of aR-IP-SeG lies in how we iteratively penalize the stochastic mapping of the VI using the parameter ρ_k . Intuitively, this is done to penalize the infeasibility of the generated iterate in terms of the stochastic VI constraint in problem (6). At each iteration k , we select indices i_k and $\tilde{i}_k$ uniformly at random and update only the corresponding blocks of the variables y_k and x_k by taking a step in a negative direction of the partial sample subgradient $\tilde{\nabla}_i f(\bullet, \xi_k)$ and sample map $F_i(\bullet, \xi_k)$ for $i = i_k$ and $\tilde{i}_k$. Then, we compute the projection onto sets X_{i_k} and $X_{\tilde{i}_k}$. Note that each player is associated with multi-dimensional strategies, denoted by n_i for $i = 1, \dots, N$, where $\sum_{i=1}^N n_i = n$. Also, at each iteration, a player is randomly chosen to update her/his full block of strategy. Also, γ_k and ρ_k denote the stepsize and the penalty parameter, respectively. Finally, the output of the proposed algorithm is a weighted average of the generated

sequence $\{y_k\}$. This is done in a novel way through incorporating both the stepsize and the penalty parameter into averaging weights.

Algorithm 1 Averaging Randomized Iteratively Penalized Stochastic Extra-Gradient Method (aR-IP-SeG)

- 1: **initialization:** Set random initial points $x_0, y_0 \in X$, an initial stepsize $\gamma_0 > 0$, an initial penalty parameter $\rho_0 > 0$ a scalar $r < 1$, $\bar{y}_0 = y_0$, and $\Gamma_0 = 0$.
- 2: **for** $k = 0, 1, \dots, K - 1$ **do**
- 3: Generate i_k and $\tilde{i}_k$ uniformly from $\{1, \dots, N\}$.
- 4: Generate ξ_k and $\tilde{\xi}_k$ as realizations of the random vector ξ .
- 5: Update the variables y_k and x_k as

$$y_{k+1}^{(i)} := \begin{cases} \mathcal{P}_{X_i} \left(x_k^{(i)} - \gamma_k (\tilde{\nabla}_i f(x_k, \tilde{\xi}_k) + \rho_k F_i(x_k, \tilde{\xi}_k)) \right) & \text{if } i = \tilde{i}_k, \\ x_k^{(i)} & \text{if } i \neq \tilde{i}_k, \end{cases} \quad (7)$$

$$x_{k+1}^{(i)} := \begin{cases} \mathcal{P}_{X_i} \left(x_k^{(i)} - \gamma_k (\tilde{\nabla}_i f(y_{k+1}, \xi_k) + \rho_k F_i(y_{k+1}, \xi_k)) \right) & \text{if } i = i_k, \\ x_k^{(i)} & \text{if } i \neq i_k. \end{cases} \quad (8)$$

- 6: Update Γ_k and $\bar{y}_k$ using the following recursions:

$$\Gamma_{k+1} := \Gamma_k + (\gamma_k \rho_k)^r, \quad (9)$$

$$\bar{y}_{k+1} := \frac{\Gamma_k \bar{y}_k + (\gamma_k \rho_k)^r y_{k+1}}{\Gamma_{k+1}}. \quad (10)$$

- 7: **end for**

- 8: Return $\bar{y}_K$.
-

Throughout, we consider the following assumptions on map F , objective function f and set X in (6).

ASSUMPTION 1 (PROBLEM PROPERTIES). Consider problem (6). Let the following holds.

- (i) Mapping $F(\bullet) : \mathbb{R}^n \rightarrow \mathbb{R}^n$ is vector-valued, continuous, and merely monotone on its domain, i.e., for all $x, y \in \text{dom}(F)$, $(F(x) - F(y))^T(x - y) \geq 0$.
- (ii) Function $f(\bullet) : \mathbb{R}^n \rightarrow \mathbb{R}$ is closed, proper, and merely convex on its domain.
- (iii) Set $X \subseteq \text{int}(\text{dom}(F) \cap \text{dom}(f))$ is nonempty, compact, and convex.

Remark 3. In view of Assumption 1, the subdifferential set $\partial f(x)$ is nonempty for all $x \in \text{int}(\text{dom}(f))$. Also, f has bounded subgradients over X . Throughout, we let scalars D_X and D_f be defined as $D_X \triangleq \sup_{x \in X} \|x\|$ and $D_f \triangleq \sup_{x \in X} |f(x)|$, respectively. Also, we let $C_F > 0$ and $C_f > 0$ be scalars such that $\|F(x)\| \leq C_F$, and $\|\tilde{\nabla} f(x)\| \leq C_f$ for all $\tilde{\nabla} f(x) \in \partial f(x)$, for all $x \in X$.

Next, we impose some standard conditions on the conditional bias and the conditional second moment on the sampled subgradient $\tilde{\nabla} f(\bullet, \xi)$ and sampled map $F(\bullet, \xi)$ produced by the oracle.

ASSUMPTION 2 (RANDOM SAMPLES). (a) The random samples $\tilde{\xi}_k$ and ξ_k are i.i.d., and $\tilde{i}_k$ and i_k are i.i.d. from the range $\{1, \dots, N\}$. Also, all these random variables are independent from each other.

- (b) For all $k \geq 0$ the stochastic mappings $F(\bullet, \tilde{\xi}_k)$ and $F(\bullet, \xi_k)$ are both unbiased estimators of $F(\bullet)$. Similarly, $\tilde{\nabla} f(\bullet, \tilde{\xi}_k)$ and $\tilde{\nabla} f(\bullet, \xi_k)$ are both unbiased estimators of $\tilde{\nabla} f(\bullet)$.
- (c) For all $x \in X$, $\mathbb{E}[\|F(x, \xi) - F(x)\|^2 \mid x] \leq v_F^2$ and $\mathbb{E}[\|\tilde{\nabla} f(x, \xi) - \tilde{\nabla} f(x)\|^2 \mid x] \leq v_f^2$, for some $v_F, v_f > 0$.

Remark 4. Under Assumption 3, we can write $\mathbb{E}[\|F(x, \xi)\|^2 \mid x] = \mathbb{E}[\|F(x, \xi) - F(x)\|^2 \mid x] + \|F(x)\|^2 \leq v_F^2 + C_F^2$, where we use Remark 3. Similarly, we have that $\mathbb{E}[\|\tilde{\nabla} f(x, \xi)\|^2 \mid x] \leq v_f^2 + C_f^2$.

Remark 5. In the case when the stochastic VI represents a Nash game, we assume that each player has access to stochastic gradient of its objective as well as stochastic gradient of the global function f .

3 PRELIMINARIES AND BACKGROUND

DEFINITION 1. We denote the history of the method by $\mathcal{F}_k$ for $k \geq 0$ defined as

$$\mathcal{F}_k \triangleq \cup_{t=0}^k \{\tilde{\xi}_t, \tilde{i}_t, \xi_t, i_t\} \cup \{x_0, y_0\}.$$

Next, we define the errors for stochastic approximation of objective function f and operator F , and block-coordinate sampling. We use the terms $w_{\bullet, k}$ and $\tilde{w}_{\bullet, k}$ to denote the errors of stochastic approximation involved at iteration k and similarly, the terms $e_{\bullet, k}$ and $\tilde{e}_{\bullet, k}$ for the errors of block-coordinate sampling.

DEFINITION 2 (STOCHASTIC ERRORS). For all $k \geq 0$ we define

$$\begin{aligned} \tilde{w}_{F, k} &\triangleq F(x_k, \tilde{\xi}_k) - F(x_k), & \tilde{w}_{f, k} &\triangleq \tilde{\nabla} f(x_k, \tilde{\xi}_k) - \tilde{\nabla} f(x_k), \\ w_{F, k} &\triangleq F(y_{k+1}, \xi_k) - F(y_{k+1}), & w_{f, k} &\triangleq \tilde{\nabla} f(y_{k+1}, \xi_k) - \tilde{\nabla} f(y_{k+1}), \\ \tilde{e}_{F, k} &\triangleq NU_{i_k} F_{i_k}(x_k, \tilde{\xi}_k) - F(x_k, \tilde{\xi}_k), & \tilde{e}_{f, k} &\triangleq NU_{i_k} \tilde{\nabla}_{i_k} f(x_k, \tilde{\xi}_k) - \tilde{\nabla} f(x_k, \tilde{\xi}_k), \\ e_{F, k} &\triangleq NU_{i_k} F_{i_k}(y_{k+1}, \xi_k) - F(y_{k+1}, \xi_k), & e_{f, k} &\triangleq NU_{i_k} \tilde{\nabla}_{i_k} f(y_{k+1}, \xi_k) - \tilde{\nabla} f(y_{k+1}, \xi_k). \end{aligned}$$

where $U_\ell \in \mathbb{R}^{n \times n_\ell}$ for $\ell \in [N]$ such that $[U_1, \dots, U_N] = I_n$ where I_n denotes the $n \times n$ identity matrix.

Based on the above definitions, we state some standard properties of the errors. The proof of the following result can be found in the extended version of the paper [19].

LEMMA 1 (PROPERTIES OF STOCHASTIC APPROXIMATION AND RANDOM BLOCKS). Consider $\tilde{e}_{F, k}$, $\tilde{e}_{f, k}$, $e_{F, k}$, and $e_{f, k}$ given by Definition 2. Let Assumption 2 hold. Then, the following statements hold almost surely for all $k \geq 0$:

- | | |
|--|---|
| (a-i) $\mathbb{E}[\tilde{w}_{F, k} \mid \mathcal{F}_{k-1}] = 0$, | (c-i) $\mathbb{E}[\tilde{e}_{F, k} \mid \mathcal{F}_{k-1} \cup \{\tilde{\xi}_k\}] = 0$, |
| (a-ii) $\mathbb{E}[\tilde{w}_{f, k} \mid \mathcal{F}_{k-1}] = 0$, | (c-ii) $\mathbb{E}[\tilde{e}_{f, k} \mid \mathcal{F}_{k-1} \cup \{\tilde{\xi}_k\}] = 0$, |
| (a-iii) $\mathbb{E}[w_{F, k} \mid \mathcal{F}_{k-1} \cup \{\tilde{\xi}_k, \tilde{i}_k\}] = 0$, | (c-iii) $\mathbb{E}[e_{F, k} \mid \mathcal{F}_{k-1} \cup \{\tilde{\xi}_k, \tilde{i}_k, \xi_k\}] = 0$, |
| (a-iv) $\mathbb{E}[w_{f, k} \mid \mathcal{F}_{k-1} \cup \{\tilde{\xi}_k, \tilde{i}_k\}] = 0$. | (c-iv) $\mathbb{E}[e_{f, k} \mid \mathcal{F}_{k-1} \cup \{\tilde{\xi}_k, \tilde{i}_k, \xi_k\}] = 0$. |
| (b-i) $\mathbb{E}[\ \tilde{w}_{F, k}\ ^2 \mid \mathcal{F}_{k-1}] \leq v_F^2$, | (d-i) $\mathbb{E}[\ \tilde{e}_{F, k}\ ^2 \mid \mathcal{F}_{k-1} \cup \{\tilde{\xi}_k\}] = (N-1)\ F(x_k, \tilde{\xi}_k)\ ^2$, |
| (b-ii) $\mathbb{E}[\ \tilde{w}_{f, k}\ ^2 \mid \mathcal{F}_{k-1}] \leq v_f^2$, | (d-ii) $\mathbb{E}[\ \tilde{e}_{f, k}\ ^2 \mid \mathcal{F}_{k-1} \cup \{\tilde{\xi}_k\}] = (N-1)\ \tilde{\nabla} f(x_k, \tilde{\xi}_k)\ ^2$, |
| (b-iii) $\mathbb{E}[\ w_{F, k}\ ^2 \mid \mathcal{F}_{k-1} \cup \{\tilde{\xi}_k, \tilde{i}_k\}] \leq v_F^2$, | (d-iii) $\mathbb{E}[\ e_{F, k}\ ^2 \mid \mathcal{F}_{k-1} \cup \{\tilde{\xi}_k, \tilde{i}_k, \xi_k\}] = (N-1)\ F(y_{k+1}, \xi_k)\ ^2$, |
| (b-iv) $\mathbb{E}[\ w_{f, k}\ ^2 \mid \mathcal{F}_{k-1} \cup \{\tilde{\xi}_k, \tilde{i}_k\}] \leq v_f^2$. | (d-iv) $\mathbb{E}[\ e_{f, k}\ ^2 \mid \mathcal{F}_{k-1} \cup \{\tilde{\xi}_k, \tilde{i}_k, \xi_k\}] = (N-1)\ \tilde{\nabla} f(y_{k+1}, \xi_k)\ ^2$. |

COROLLARY 1. Consider $\tilde{e}_{F, k}$, $\tilde{e}_{f, k}$, $e_{F, k}$, and $e_{f, k}$ given by Definition 2. Let Assumption 2 hold. Then, the following statements hold almost surely for all $k \geq 0$:

- | | |
|---|---|
| (a) $\mathbb{E}[\tilde{w}_{F, k}] = \mathbb{E}[\tilde{w}_{f, k}] = \mathbb{E}[w_{F, k}] = \mathbb{E}[w_{f, k}] = 0$, | (c) $\mathbb{E}[\tilde{e}_{F, k}] = \mathbb{E}[\tilde{e}_{f, k}] = \mathbb{E}[e_{F, k}] = \mathbb{E}[e_{f, k}] = 0$, |
| (b-i) $\mathbb{E}[\ \tilde{w}_{F, k}\ ^2] \leq v_F^2$, | (d-i) $\mathbb{E}[\ \tilde{e}_{F, k}\ ^2] \leq (N-1)(v_F^2 + C_F^2)$, |
| (b-ii) $\mathbb{E}[\ \tilde{w}_{f, k}\ ^2] \leq v_f^2$, | (d-ii) $\mathbb{E}[\ \tilde{e}_{f, k}\ ^2] \leq (N-1)(v_f^2 + C_f^2)$, |
| (b-iii) $\mathbb{E}[\ w_{F, k}\ ^2] \leq v_F^2$, | (d-iii) $\mathbb{E}[\ e_{F, k}\ ^2] \leq (N-1)(v_F^2 + C_F^2)$, |
| (b-iv) $\mathbb{E}[\ w_{f, k}\ ^2] \leq v_f^2$. | (d-iv) $\mathbb{E}[\ e_{f, k}\ ^2] \leq (N-1)(v_f^2 + C_f^2)$. |

PROOF. The relations (a-c) follow from taking expectations on both sides of the results in parts (a-c) of Lemma 1 and invoking the law of total expectation. We can show (d-i) as follows: (i) taking expectations with respect to $\tilde{\xi}_k$ on both sides of (d-i) in Lemma 1; (ii) applying Remark 4; (iii) lastly, taking expectations with respect to $\mathcal{F}_{k-1}$ on both sides of the resulting inequality in (ii). This will complete the proof of (d-i) in Corollary 1. Similarly, we can show (d-ii), (d-iii), and (d-iv) in Corollary 1. $\square$

In the following lemma, we show that the update rules (7) and (8) in Algorithm 1 can be written compactly in terms of the full subgradient $\tilde{\nabla}f$ and map F following the terms introduced in Definition 2.

LEMMA 2 (COMPACT REPRESENTATION OF THE SCHEME). Consider Algorithm 1. The update rules (7) and (8) can be compactly written as

$$\begin{aligned} y_{k+1} &= \mathcal{P}_X \left(x_k - N^{-1} \gamma_k \left(\tilde{\nabla}f(x_k) + \tilde{w}_{f,k} + \tilde{e}_{f,k} + \rho_k F(x_k) + \rho_k \tilde{w}_{F,k} + \rho_k \tilde{e}_{F,k} \right) \right) \\ x_{k+1} &= \mathcal{P}_X \left(x_k - N^{-1} \gamma_k \left(\tilde{\nabla}f(y_{k+1}) + w_{f,k} + e_{f,k} + \rho_k F(y_{k+1}) + \rho_k w_{F,k} + \rho_k e_{F,k} \right) \right). \end{aligned}$$

PROOF. Note that in view of $X = \prod_{i=1}^N X_i$, using the definition of the Euclidean projection operator, we have that $\mathcal{P}_X(\bullet) = (\mathcal{P}_{X_1}(\bullet), \dots, \mathcal{P}_{X_N}(\bullet))$, then update rule (7) can be written as

$$y_{k+1} = \mathcal{P}_X \left(x_k - \gamma_k (\mathbf{U}_i \tilde{\nabla}_i f(x_k, \tilde{\xi}_k) + \rho_k F_i(x_k, \tilde{\xi}_k)) \right), \quad i = \tilde{i}_k.$$

The result follows using Definition 2. Similarly, one can obtain the compact form of the update rule (8). $\square$

In our analysis, we use the following properties of projection map.

LEMMA 3 (PROPERTIES OF PROJECTION MAPPING [3]). Let $X \subseteq \mathbb{R}^n$ be a nonempty closed convex set.

- (a) $\|\mathcal{P}_X(u) - \mathcal{P}_X(v)\| \leq \|u - v\|$ for all $u, v \in \mathbb{R}^n$.
- (b) $(\mathcal{P}_X(u) - u)^T (x - \mathcal{P}_X(u)) \geq 0$ for all $u \in \mathbb{R}^n$ and $x \in X$.

We will adopt the following error function to measure the quality of solution generated by Algorithm 1 in terms of infeasibility.

DEFINITION 3 (THE DUAL GAP FUNCTION [29]). Let $X \subseteq \mathbb{R}^n$ be a nonempty, closed, and convex set and $F : X \rightarrow \mathbb{R}^n$ be a vector-valued mapping. Then, for any $x \in X$, the dual gap function $\text{Gap}^* : X \rightarrow \mathbb{R} \cup \{+\infty\}$ is defined as $\text{Gap}^*(x) \triangleq \sup_{y \in X} F(y)^T (x - y)$.

Remark 6. Notably when $X \neq \emptyset$, the dual gap function is nonnegative over X . Also, when F is continuous and monotone and X is closed and convex, $\text{Gap}^*(x^*) = 0$ if and only if $x^* \in \text{SOL}(X, F)$ (cf. [22]).

LEMMA 4 (BOUNDS ON THE HARMONIC SERIES [26]). Let $0 \leq \alpha < 1$ be a given scalar. Then, for any integer $K \geq 2^{\frac{1}{1-\alpha}}$, we have $\frac{K^{1-\alpha}}{2(1-\alpha)} \leq \sum_{k=0}^{K-1} \frac{1}{(k+1)^\alpha} \leq \frac{K^{1-\alpha}}{1-\alpha}$.

4 PERFORMANCE ANALYSIS

In this section, we develop a rate and complexity analysis for Algorithm 1. We begin with showing that $\bar{y}_k$ generated by Algorithm 1 is a well-defined weighted average.

LEMMA 5 (WEIGHTED AVERAGING). Let $\{\bar{y}_k\}$ be generated by Algorithm 1. Let us define the weights $\lambda_{k,K} \triangleq \frac{(y_k \rho_k)^r}{\sum_{j=0}^{K-1} (y_j \rho_j)^r}$ for $k \in \{0, \dots, K-1\}$ and $K \geq 1$. Then, for any $K \geq 1$, we have $\bar{y}_K = \sum_{k=0}^{K-1} \lambda_{k,K} y_{k+1}$. Also, when X is a convex set, we have $\bar{y}_K \in X$.

PROOF. We employ induction to show $\bar{y}_K = \sum_{k=0}^{K-1} \lambda_{k,K} y_{k+1}$ for any $K \geq 1$. For $K = 1$ we have $\sum_{k=0}^0 \lambda_{k,1} y_{k+1} = \lambda_{0,1} y_1 = y_1$, where we used $\lambda_{0,1} = 1$. Also, from the equations (9)–(10) and the initialization $\Gamma_0 = 0$, we have

$$\bar{y}_1 := \frac{\Gamma_0 \bar{y}_0 + (\gamma_0 \rho_0)^r y_1}{\Gamma_1} = \frac{0 + (\gamma_0 \rho_0)^r y_1}{\Gamma_0 + \gamma_0^r} = y_1.$$

The preceding two relations imply that the hypothesis statement holds for $K = 1$. Next, suppose the relation holds for some $K \geq 1$. From the hypothesis, equations (9)–(10), and that $\Gamma_K = \sum_{k=0}^{K-1} \gamma_k^r$ for all $K \geq 1$, we have

$$\begin{aligned} \bar{y}_{K+1} &= \frac{\Gamma_K \bar{y}_K + (\gamma_K \rho_K)^r y_{K+1}}{\Gamma_{K+1}} = \frac{\left(\sum_{k=0}^{K-1} (\gamma_k \rho_k)^r \right) \sum_{k=0}^{K-1} \lambda_{k,K} y_{k+1} + (\gamma_K \rho_K)^r y_{K+1}}{\Gamma_{K+1}} \\ &= \frac{\sum_{k=0}^K (\gamma_k \rho_k)^r y_{k+1}}{\sum_{j=0}^K (\gamma_j \rho_j)^r} = \sum_{k=0}^K \left(\frac{(\gamma_k \rho_k)^r}{\sum_{j=0}^K (\gamma_j \rho_j)^r} \right) y_{k+1} = \sum_{k=0}^K \lambda_{k,K+1} y_{k+1}, \end{aligned}$$

implying that the induction hypothesis holds for $K + 1$. Thus, we conclude that the averaging formula holds for all $K \geq 1$. Note that since $\sum_{k=0}^{K-1} \lambda_{k,K} = 1$, under the convexity of the set X , we have $\bar{y}_K \in X$. This completes the proof. $\square$

Next, we prove a one-step lemma to obtain an upper bound for $F(y)^T (y_{k+1} - y) + \rho_k^{-1} (f(y_{k+1}) - f(y))$ in terms of consecutive iterates and error terms. this result will later help us obtain upper bounds for both the suboptimality of the objective function and the dual gap function in Proposition 1. The proof of the following lemma can be found in the extended version of the paper [19].

LEMMA 6 (AN ERROR BOUND). Consider Algorithm 1 for solving problem (6). Let Assumptions 1 and 2 hold. Let the auxiliary stochastic sequence $\{u_k\}$ be defined recursively as

$$u_{k+1} \triangleq \mathcal{P}_X \left(u_k + N^{-1} \gamma_k (w_{f,k} + e_{f,k} + \rho_k w_{F,k} + \rho_k e_{F,k}) \right), \quad (11)$$

where $u_0 := x_0$. Then for any arbitrary $y \in X$ and $k \geq 0$ we have

$$\begin{aligned} & (\gamma_k \rho_k)^r F(y)^T (y_{k+1} - y) + (\gamma_k \rho_k)^r \rho_k^{-1} (f(y_{k+1}) - f(y)) \\ & \leq 0.5N(\gamma_k \rho_k)^{r-1} (\|x_k - y\|^2 - \|x_{k+1} - y\|^2 + \|u_k - y\|^2 - \|u_{k+1} - y\|^2) \\ & \quad + 2N^{-1}(\gamma_k \rho_k)^{r+1} \rho_k^{-2} \left(6C_f^2 + 3\|\tilde{w}_{f,k}\|^2 + 3\|\tilde{e}_{f,k}\|^2 + 4\|w_{f,k}\|^2 + 4\|e_{f,k}\|^2 \right) \\ & \quad + 2N^{-1}(\gamma_k \rho_k)^{r+1} \left(6C_F^2 + 3\|\tilde{w}_{F,k}\|^2 + 3\|\tilde{e}_{F,k}\|^2 + 4\|w_{F,k}\|^2 + 4\|e_{F,k}\|^2 \right) \\ & \quad + \gamma_k^r \rho_k^{r-1} (w_{f,k} + e_{f,k} + \rho_k w_{F,k} + \rho_k e_{F,k})^T (u_k - y_{k+1}). \end{aligned} \quad (12)$$

In the following result, we show that one of the error terms that appear in the inequality (12) has a zero mean. This result will help us with obtaining the convergence rates for Algorithm 1.

LEMMA 7. Consider the auxiliary sequence defined by (11). Let Assumptions 1 and 2 hold. Then for any $k \geq 0$ we have $\mathbb{E} \left[(w_{f,k} + e_{f,k} + \rho_k w_{F,k} + \rho_k e_{F,k})^T (u_k - y_{k+1}) \right] = 0$.

PROOF. Consider $\{u_k\}$ defined by (11). From this definition and Algorithm 1 we observe that u_k is $\mathcal{F}_{k-1}$ -measurable. Also, note that y_{k+1} is $\mathcal{F}_{k-1} \cup \{\tilde{\xi}_k, \tilde{i}_k\}$ -measurable. We can write

$$\begin{aligned} & \mathbb{E} \left[(w_{f,k} + e_{f,k} + \rho_k w_{F,k} + \rho_k e_{F,k})^T (u_k - y_{k+1}) \mid \mathcal{F}_{k-1} \cup \{\tilde{\xi}_k, \tilde{i}_k\} \right] \\ & = \mathbb{E} \left[(w_{f,k} + e_{f,k} + \rho_k w_{F,k} + \rho_k e_{F,k}) \mid \mathcal{F}_{k-1} \cup \{\tilde{\xi}_k, \tilde{i}_k\} \right]^T (u_k - y_{k+1}). \end{aligned} \quad (13)$$

Note that from Lemma 1 (a) we have

$$\mathbb{E}[w_{f,k} + \rho_k w_{F,k} \mid \mathcal{F}_{k-1} \cup \{\tilde{\xi}_k, \tilde{i}_k\}] = 0. \quad (14)$$

We also have from Lemma 1 (c) that $\mathbb{E}[e_{f,k} + \rho_k e_{F,k} \mid \mathcal{F}_{k-1} \cup \{\tilde{\xi}_k, \tilde{i}_k, \xi_k\}] = 0$. Taking conditional expectations with respect to ξ_k on both sides of the preceding equation, we obtain $\mathbb{E}[e_{f,k} + \rho_k e_{F,k} \mid \mathcal{F}_{k-1} \cup \{\tilde{\xi}_k, \tilde{i}_k\}] = 0$. Combining the preceding relation with (13) and (14), we have that

$$\mathbb{E} \left[(w_{f,k} + e_{f,k} + \rho_k w_{F,k} + \rho_k e_{F,k})^T (u_k - y_{k+1}) \mid \mathcal{F}_{k-1} \cup \{\tilde{\xi}_k, \tilde{i}_k\} \right] = 0.$$

Taking conditional expectations with respect to $\mathcal{F}_{k-1} \cup \{\tilde{\xi}_k, \tilde{i}_k\}$ on the both sides, we obtain the result. $\square$

In the following, we employ the results of Lemmas 6 and 7 to obtain upper bounds on the suboptimality of the objective function and the dual gap function associated with the stochastic VI constraint in problem (6). This will prepare us to analyze the convergence speed of Algorithm 1 later in Theorem 1.

PROPOSITION 1 (ERROR BOUNDS). Consider Algorithm 1 for solving problem (6). Let Assumptions 1 and 2 hold. Suppose $\{\gamma_k \rho_k\}$ is nonincreasing, $\{\rho_k\}$ is nondecreasing, and $0 \leq r < 1$ is a scalar. The following results hold for all $K \geq 2$

$$\mathbb{E}[f(\bar{y}_K)] - f^* \leq \frac{4ND_X^2(\gamma_{K-1}\rho_{K-1})^{r-1}\rho_{K-1} + 2N^{-1}\sum_{k=0}^{K-1}(\gamma_k\rho_k)^{1+r}\rho_k \left(\theta_F + \theta_f\rho_k^{-2}\right)}{\sum_{k=0}^{K-1}(\gamma_k\rho_k)^r}, \quad (15)$$

$$\mathbb{E}[\text{Gap}^*(\bar{y}_K)] \leq \frac{4ND_X^2(\gamma_{K-1}\rho_{K-1})^{r-1} + 2N^{-1}\sum_{k=0}^{K-1}(\gamma_k\rho_k)^r \left(\theta_F\gamma_k\rho_k + \theta_f\gamma_k\rho_k^{-1} + 2ND_f\rho_k^{-1}\right)}{\sum_{k=0}^{K-1}(\gamma_k\rho_k)^r}, \quad (16)$$

where $\theta_F \triangleq (7N-1)C_F^2 + 7Nv_F^2$ and $\theta_f \triangleq (7N-1)C_f^2 + 7Nv_f^2$.

PROOF. First we show the relation (15). Consider the inequality (12). Let $y := x^*$ where $x^* \in X$ is an optimal solution to the problem (6). This implies that $x^* \in \text{SOL}(X, \mathbb{E}[F(\bullet, \xi)])$ or equivalently, $F(x^*)^T(y_{k+1} - x^*) \geq 0$. We obtain

$$\begin{aligned} (\gamma_k\rho_k)^r \rho_k^{-1} (f(y_{k+1}) - f^*) &\leq 0.5N(\gamma_k\rho_k)^{r-1} (\|x_k - x^*\|^2 - \|x_{k+1} - x^*\|^2 + \|u_k - x^*\|^2 - \|u_{k+1} - x^*\|^2) \\ &\quad + 2N^{-1}(\gamma_k\rho_k)^{r+1}\rho_k^{-2} \left(6C_f^2 + 3\|\tilde{w}_{f,k}\|^2 + 3\|\tilde{e}_{f,k}\|^2 + 4\|w_{f,k}\|^2 + 4\|e_{f,k}\|^2\right) \\ &\quad + 2N^{-1}(\gamma_k\rho_k)^{r+1} \left(6C_F^2 + 3\|\tilde{w}_{F,k}\|^2 + 3\|\tilde{e}_{F,k}\|^2 + 4\|w_{F,k}\|^2 + 4\|e_{F,k}\|^2\right) \\ &\quad + \gamma_k^r \rho_k^{r-1} (w_{f,k} + e_{f,k} + \rho_k w_{F,k} + \rho_k e_{F,k})^T (u_k - y_{k+1}). \end{aligned} \quad (17)$$

Multiplying the both sides by ρ_k and then, adding and subtracting the term

$$0.5N(\gamma_{k-1}\rho_{k-1})^{r-1}\rho_{k-1} (\|x_k - x^*\|^2 + \|u_k - x^*\|^2),$$

we have for all $k \geq 1$

$$\begin{aligned}
(\gamma_k \rho_k)^r (f(y_{k+1}) - f^*) &\leq 0.5N(\gamma_{k-1} \rho_{k-1})^{r-1} \rho_{k-1} (\|x_k - x^*\|^2 + \|u_k - x^*\|^2) \\
&\quad - 0.5N(\gamma_k \rho_k)^{r-1} \rho_k (\|x_{k+1} - x^*\|^2 + \|u_{k+1} - x^*\|^2) \\
&\quad + 0.5N((\gamma_k \rho_k)^{r-1} \rho_k - (\gamma_{k-1} \rho_{k-1})^{r-1} \rho_{k-1}) (\|x_k - x^*\|^2 + \|u_k - x^*\|^2) \\
&\quad + 2N^{-1}(\gamma_k \rho_k)^{r+1} \rho_k^{-1} \left(6C_f^2 + 3\|\tilde{w}_{f,k}\|^2 + 3\|\tilde{e}_{f,k}\|^2 + 4\|w_{f,k}\|^2 + 4\|e_{f,k}\|^2 \right) \\
&\quad + 2N^{-1}(\gamma_k \rho_k)^{r+1} \rho_k (6C_F^2 + 3\|\tilde{w}_{F,k}\|^2 + 3\|\tilde{e}_{F,k}\|^2 + 4\|w_{F,k}\|^2 + 4\|e_{F,k}\|^2) \\
&\quad + (\gamma_k \rho_k)^r (w_{f,k} + e_{f,k} + \rho_k w_{F,k} + \rho_k e_{F,k})^T (u_k - y_{k+1}). \tag{18}
\end{aligned}$$

Note that because $r < 1$ and that $\{\gamma_k \rho_k\}$ is nonincreasing and $\{\rho_k\}$ is nondecreasing, we have

$$\gamma_k^{r-1} \rho_k - \gamma_{k-1}^{r-1} \rho_{k-1} \geq 0.$$

Thus, in view of Remark 3 we have

$$\begin{aligned}
0.5N((\gamma_k \rho_k)^{r-1} \rho_k - (\gamma_{k-1} \rho_{k-1})^{r-1} \rho_{k-1}) (\|x_k - x^*\|^2 + \|u_k - x^*\|^2) \\
\leq 4ND_X^2 ((\gamma_k \rho_k)^{r-1} \rho_k - (\gamma_{k-1} \rho_{k-1})^{r-1} \rho_{k-1}).
\end{aligned}$$

Substituting the preceding bound in (19) and then, summing the resulting inequality for $k = 1, \dots, K-1$ we obtain

$$\begin{aligned}
\sum_{k=1}^{K-1} (\gamma_k \rho_k)^r (f(y_{k+1}) - f^*) &\leq 0.5N(\gamma_0 \rho_0)^{r-1} \rho_0 (\|x_1 - x^*\|^2 + \|u_1 - x^*\|^2) \\
&\quad + 4ND_X^2 ((\gamma_{K-1} \rho_{K-1})^{r-1} \rho_{K-1} - (\gamma_0 \rho_0)^{r-1} \rho_0) \\
&\quad + 2N^{-1} \sum_{k=1}^{K-1} (\gamma_k \rho_k)^{r+1} \rho_k^{-1} \left(6C_f^2 + 3\|\tilde{w}_{f,k}\|^2 + 3\|\tilde{e}_{f,k}\|^2 + 4\|w_{f,k}\|^2 + 4\|e_{f,k}\|^2 \right) \\
&\quad + 2N^{-1} \sum_{k=1}^{K-1} (\gamma_k \rho_k)^{r+1} \rho_k (6C_F^2 + 3\|\tilde{w}_{F,k}\|^2 + 3\|\tilde{e}_{F,k}\|^2 + 4\|w_{F,k}\|^2 + 4\|e_{F,k}\|^2) \\
&\quad + \sum_{k=1}^{K-1} (\gamma_k \rho_k)^r (w_{f,k} + e_{f,k} + \rho_k w_{F,k} + \rho_k e_{F,k})^T (u_k - y_{k+1}). \tag{19}
\end{aligned}$$

From (17) for $k = 0$ we have

$$\begin{aligned}
(\gamma_0 \rho_0)^r (f(y_1) - f^*) &\leq 0.5N(\gamma_0 \rho_0)^{r-1} \rho_0 (\|x_0 - x^*\|^2 - \|x_1 - x^*\|^2 + \|u_0 - x^*\|^2 - \|u_1 - x^*\|^2) \\
&\quad + 2N^{-1}(\gamma_0 \rho_0)^{1+r} \rho_0^{-1} \left(6C_f^2 + 3\|\tilde{w}_{f,0}\|^2 + 3\|\tilde{e}_{f,0}\|^2 + 4\|w_{f,0}\|^2 + 4\|e_{f,0}\|^2 \right) \\
&\quad + 2N^{-1}(\gamma_0 \rho_0)^{1+r} \rho_0 (6C_F^2 + 3\|\tilde{w}_{F,0}\|^2 + 3\|\tilde{e}_{F,0}\|^2 + 4\|w_{F,0}\|^2 + 4\|e_{F,0}\|^2) \\
&\quad + (\gamma_0 \rho_0)^r (w_{f,0} + e_{f,0} + \rho_k w_{F,0} + \rho_k e_{F,0})^T (u_0 - y_1). \tag{20}
\end{aligned}$$

Summing the preceding two relations we obtain

$$\begin{aligned}
\sum_{k=0}^{K-1} (\gamma_k \rho_k)^r (f(y_{k+1}) - f^*) &\leq 0.5N(\gamma_0 \rho_0)^{r-1} \rho_0 (\|x_0 - x^*\|^2 + \|u_0 - x^*\|^2) \\
&\quad + 4ND_X^2 ((\gamma_{K-1} \rho_{K-1})^{r-1} \rho_{K-1} - (\gamma_0 \rho_0)^{r-1} \rho_0) \\
&\quad + 2N^{-1} \sum_{k=0}^{K-1} (\gamma_k \rho_k)^{r+1} \rho_k^{-1} \left(6C_f^2 + 3\|\tilde{w}_{f,k}\|^2 + 3\|\tilde{e}_{f,k}\|^2 + 4\|w_{f,k}\|^2 + 4\|e_{f,k}\|^2 \right) \\
&\quad + 2N^{-1} \sum_{k=0}^{K-1} (\gamma_k \rho_k)^{r+1} \rho_k \left(6C_F^2 + 3\|\tilde{w}_{F,k}\|^2 + 3\|\tilde{e}_{F,k}\|^2 + 4\|w_{F,k}\|^2 + 4\|e_{F,k}\|^2 \right) \\
&\quad + \sum_{k=0}^{K-1} (\gamma_k \rho_k)^r (w_{f,k} + e_{f,k} + \rho_k w_{F,k} + \rho_k e_{F,k})^T (u_k - y_{k+1}). \tag{21}
\end{aligned}$$

Note that from the convexity of f and Lemma 5, we have

$$\frac{\sum_{k=0}^{K-1} (\gamma_k \rho_k)^r f(y_{k+1})}{\sum_{k=0}^{K-1} (\gamma_k \rho_k)^r} = \sum_{k=0}^{K-1} \left(\frac{(\gamma_k \rho_k)^r}{\sum_{j=0}^{K-1} (\gamma_j \rho_j)^r} \right) f(y_{k+1}) = \sum_{k=0}^{K-1} \lambda_{k,K} f(y_{k+1}) \geq f \left(\sum_{k=0}^{K-1} \lambda_{k,K} y_{k+1} \right) = f(\bar{y}_K).$$

Dividing the both sides of (21) by $\sum_{k=0}^{K-1} (\gamma_k \rho_k)^r$, using the preceding relation, and $\|x_0 - x^*\|^2 + \|u_0 - x^*\|^2 \leq 8D_X^2$, we obtain

$$\begin{aligned}
f(\bar{y}_K) - f^* &\leq \left(\sum_{k=0}^{K-1} (\gamma_k \rho_k)^r \right)^{-1} \left(4ND_X^2 (\gamma_0 \rho_0)^{r-1} \rho_0 + 4ND_X^2 ((\gamma_{K-1} \rho_{K-1})^{r-1} \rho_{K-1} - (\gamma_0 \rho_0)^{r-1} \rho_0) \right. \\
&\quad + 2N^{-1} \sum_{k=0}^{K-1} (\gamma_k \rho_k)^{r+1} \rho_k^{-1} \left(6C_f^2 + 3\|\tilde{w}_{f,k}\|^2 + 3\|\tilde{e}_{f,k}\|^2 + 4\|w_{f,k}\|^2 + 4\|e_{f,k}\|^2 \right) \\
&\quad + 2N^{-1} \sum_{k=0}^{K-1} (\gamma_k \rho_k)^{r+1} \rho_k \left(6C_F^2 + 3\|\tilde{w}_{F,k}\|^2 + 3\|\tilde{e}_{F,k}\|^2 + 4\|w_{F,k}\|^2 + 4\|e_{F,k}\|^2 \right) \\
&\quad \left. + \sum_{k=0}^{K-1} (\gamma_k \rho_k)^r (w_{f,k} + e_{f,k} + \rho_k w_{F,k} + \rho_k e_{F,k})^T (u_k - y_{k+1}) \right). \tag{22}
\end{aligned}$$

Taking expectations on the both sides and applying Corollary 1 and Lemma 7, we obtain

$$\begin{aligned}
\mathbb{E}[f(\bar{y}_K)] - f^* &\leq \left(\sum_{k=0}^{K-1} (\gamma_k \rho_k)^r \right)^{-1} \left(4ND_X^2 (\gamma_{K-1} \rho_{K-1})^{r-1} \rho_{K-1} \right. \\
&\quad + 2N^{-1} \sum_{k=0}^{K-1} (\gamma_k \rho_k)^{r+1} \rho_k^{-1} \left(6C_f^2 + 7v_f^2 + 7(N-1)(v_f^2 + C_f^2) \right) \\
&\quad \left. + 2N^{-1} \sum_{k=0}^{K-1} (\gamma_k \rho_k)^{r+1} \rho_k \left(6C_F^2 + 7v_F^2 + 7(N-1)(v_F^2 + C_F^2) \right) \right).
\end{aligned}$$

This implies that the inequality (15) holds for all $K \geq 2$. Next we show the inequality (16). Consider the inequality (12) again for an arbitrary $y \in X$. In view of Remark 3 we have $f(y_{k+1}) - f(y) \leq 2D_f$. Rearranging the terms in (12)

we obtain

$$\begin{aligned}
(\gamma_k \rho_k)^r F(y)^T (y_{k+1} - y) &\leq 0.5N(\gamma_k \rho_k)^{r-1} (\|x_k - y\|^2 - \|x_{k+1} - y\|^2 + \|u_k - y\|^2 - \|u_{k+1} - y\|^2) \\
&\quad + 2N^{-1}(\gamma_k \rho_k)^{r+1} \rho_k^{-2} \left(6C_f^2 + 3\|\tilde{w}_{f,k}\|^2 + 3\|\tilde{e}_{f,k}\|^2 + 4\|w_{f,k}\|^2 + 4\|e_{f,k}\|^2 \right) \\
&\quad + 2N^{-1}(\gamma_k \rho_k)^{r+1} \left(6C_F^2 + 3\|\tilde{w}_{F,k}\|^2 + 3\|\tilde{e}_{F,k}\|^2 + 4\|w_{F,k}\|^2 + 4\|e_{F,k}\|^2 \right) \\
&\quad + \gamma_k^r \rho_k^{r-1} (w_{f,k} + e_{f,k} + \rho_k w_{F,k} + \rho_k e_{F,k})^T (u_k - y_{k+1}) + 2(\gamma_k \rho_k)^r \rho_k^{-1} D_f. \quad (23)
\end{aligned}$$

Adding and subtracting $(\gamma_k \rho_k)^{r-1} (\|x_k - y\|^2 + \|u_k - y\|^2)$, for all $k \geq 1$ we have

$$\begin{aligned}
(\gamma_k \rho_k)^r F(y)^T (y_{k+1} - y) &\leq 0.5N(\gamma_{k-1} \rho_{k-1})^{r-1} (\|x_k - y\|^2 + \|u_k - y\|^2) \\
&\quad - 0.5N(\gamma_k \rho_k)^{r-1} (\|x_{k+1} - y\|^2 + \|u_{k+1} - y\|^2) \\
&\quad + 0.5N \left((\gamma_k \rho_k)^{r-1} - (\gamma_{k-1} \rho_{k-1})^{r-1} \right) (\|x_k - y\|^2 + \|u_k - y\|^2) \\
&\quad + 2N^{-1}(\gamma_k \rho_k)^{r+1} \rho_k^{-2} \left(6C_f^2 + 3\|\tilde{w}_{f,k}\|^2 + 3\|\tilde{e}_{f,k}\|^2 + 4\|w_{f,k}\|^2 + 4\|e_{f,k}\|^2 \right) \\
&\quad + 2N^{-1}(\gamma_k \rho_k)^{r+1} \left(6C_F^2 + 3\|\tilde{w}_{F,k}\|^2 + 3\|\tilde{e}_{F,k}\|^2 + 4\|w_{F,k}\|^2 + 4\|e_{F,k}\|^2 \right) \\
&\quad + \gamma_k^r \rho_k^{r-1} (w_{f,k} + e_{f,k} + \rho_k w_{F,k} + \rho_k e_{F,k})^T (u_k - y_{k+1}) + 2(\gamma_k \rho_k)^r \rho_k^{-1} D_f. \quad (24)
\end{aligned}$$

Note that because $r < 1$ and that $\{\gamma_k \rho_k\}$ is nonincreasing, we have $(\gamma_k \rho_k)^{r-1} - (\gamma_{k-1} \rho_{k-1})^{r-1} \geq 0$. Thus, in view of Remark 3 we have

$$0.5N \left((\gamma_k \rho_k)^{r-1} - (\gamma_{k-1} \rho_{k-1})^{r-1} \right) (\|x_k - x^*\|^2 + \|u_k - x^*\|^2) \leq 4ND_X^2 \left((\gamma_k \rho_k)^{r-1} - (\gamma_{k-1} \rho_{k-1})^{r-1} \right).$$

Substituting the preceding bound in (24) and then, summing the resulting inequality for $k = 1, \dots, K-1$ we obtain

$$\begin{aligned}
\sum_{k=1}^{K-1} (\gamma_k \rho_k)^r F(y)^T (y_{k+1} - y) &\leq 0.5N(\gamma_0 \rho_0)^{r-1} (\|x_1 - y\|^2 + \|u_1 - y\|^2) \\
&\quad + 4ND_X^2 \left((\gamma_{K-1} \rho_{K-1})^{r-1} - (\gamma_0 \rho_0)^{r-1} \right) \\
&\quad + 2N^{-1} \sum_{k=1}^{K-1} (\gamma_k \rho_k)^{r+1} \rho_k^{-2} \left(6C_f^2 + 3\|\tilde{w}_{f,k}\|^2 + 3\|\tilde{e}_{f,k}\|^2 + 4\|w_{f,k}\|^2 + 4\|e_{f,k}\|^2 \right) \\
&\quad + 2N^{-1} \sum_{k=1}^{K-1} (\gamma_k \rho_k)^{r+1} \left(6C_F^2 + 3\|\tilde{w}_{F,k}\|^2 + 3\|\tilde{e}_{F,k}\|^2 + 4\|w_{F,k}\|^2 + 4\|e_{F,k}\|^2 \right) \\
&\quad + \sum_{k=1}^{K-1} \gamma_k^r \rho_k^{r-1} (w_{f,k} + e_{f,k} + \rho_k w_{F,k} + \rho_k e_{F,k})^T (u_k - y_{k+1}) + 2D_f \sum_{k=1}^{K-1} (\gamma_k \rho_k)^r \rho_k^{-1}. \quad (25)
\end{aligned}$$

Consider (23) for $k = 0$. Summing that relation with (25) we have

$$\begin{aligned}
F(y)^T \left(\sum_{k=0}^{K-1} (\gamma_k \rho_k)^r y_{k+1} - y \right) &\leq 0.5N(\gamma_0 \rho_0)^{r-1} (\|x_0 - y\|^2 + \|u_0 - y\|^2) \\
&+ 4ND_X^2 ((\gamma_{K-1} \rho_{K-1})^{r-1} - (\gamma_0 \rho_0)^{r-1}) \\
&+ 2N^{-1} \sum_{k=0}^{K-1} (\gamma_k \rho_k)^{r+1} \rho_k^{-2} \left(6C_f^2 + 3\|\tilde{w}_{f,k}\|^2 + 3\|\tilde{e}_{f,k}\|^2 + 4\|w_{f,k}\|^2 + 4\|e_{f,k}\|^2 \right) \\
&+ 2N^{-1} \sum_{k=0}^{K-1} (\gamma_k \rho_k)^{r+1} (6C_F^2 + 3\|\tilde{w}_{F,k}\|^2 + 3\|\tilde{e}_{F,k}\|^2 + 4\|w_{F,k}\|^2 + 4\|e_{F,k}\|^2) \\
&+ \sum_{k=0}^{K-1} \gamma_k^r \rho_k^{r-1} (w_{f,k} + e_{f,k} + \rho_k w_{F,k} + \rho_k e_{F,k})^T (u_k - y_{k+1}) + 2D_f \sum_{k=0}^{K-1} (\gamma_k \rho_k)^r \rho_k^{-1}. \tag{26}
\end{aligned}$$

Dividing the both side of (26) by $\sum_{k=0}^{K-1} (\gamma_k \rho_k)^r$, invoking Lemma 5, and $\|x_0 - y\|^2 + \|u_0 - y\|^2 \leq 8D_X^2$, we obtain

$$\begin{aligned}
F(y)^T (\bar{y}_K - y) &\leq \left(\sum_{k=0}^{K-1} (\gamma_k \rho_k)^r \right)^{-1} (4ND_X^2 (\gamma_{K-1} \rho_{K-1})^{r-1} \\
&+ 2N^{-1} \sum_{k=0}^{K-1} (\gamma_k \rho_k)^{r+1} \rho_k^{-2} \left(6C_f^2 + 3\|\tilde{w}_{f,k}\|^2 + 3\|\tilde{e}_{f,k}\|^2 + 4\|w_{f,k}\|^2 + 4\|e_{f,k}\|^2 \right) \\
&+ 2N^{-1} \sum_{k=0}^{K-1} (\gamma_k \rho_k)^{r+1} (6C_F^2 + 3\|\tilde{w}_{F,k}\|^2 + 3\|\tilde{e}_{F,k}\|^2 + 4\|w_{F,k}\|^2 + 4\|e_{F,k}\|^2) \\
&+ \sum_{k=0}^{K-1} \gamma_k^r \rho_k^{r-1} (w_{f,k} + e_{f,k} + \rho_k w_{F,k} + \rho_k e_{F,k})^T (u_k - y_{k+1}) + 2D_f \sum_{k=0}^{K-1} (\gamma_k \rho_k)^r \rho_k^{-1}). \tag{27}
\end{aligned}$$

Taking the supremum on the both sides of (27) with respect to y over the set X and invoking Definition 3, we have

$$\begin{aligned}
Gap^*(\bar{y}_K) &\leq \left(\sum_{k=0}^{K-1} (\gamma_k \rho_k)^r \right)^{-1} (4ND_X^2 (\gamma_{K-1} \rho_{K-1})^{r-1} \\
&+ 2N^{-1} \sum_{k=0}^{K-1} (\gamma_k \rho_k)^{r+1} \rho_k^{-2} \left(6C_f^2 + 3\|\tilde{w}_{f,k}\|^2 + 3\|\tilde{e}_{f,k}\|^2 + 4\|w_{f,k}\|^2 + 4\|e_{f,k}\|^2 \right) \\
&+ 2N^{-1} \sum_{k=0}^{K-1} (\gamma_k \rho_k)^{r+1} (6C_F^2 + 3\|\tilde{w}_{F,k}\|^2 + 3\|\tilde{e}_{F,k}\|^2 + 4\|w_{F,k}\|^2 + 4\|e_{F,k}\|^2) \\
&+ \sum_{k=0}^{K-1} \gamma_k^r \rho_k^{r-1} (w_{f,k} + e_{f,k} + \rho_k w_{F,k} + \rho_k e_{F,k})^T (u_k - y_{k+1}) + 2D_f \sum_{k=0}^{K-1} (\gamma_k \rho_k)^r \rho_k^{-1}).
\end{aligned}$$

Taking expectations on the both sides and applying Corollary 1 and Lemma 7, we obtain

$$\begin{aligned}\mathbb{E}[\text{Gap}^*(\bar{y}_K)] &\leq \left(\sum_{k=0}^{K-1} (\gamma_k \rho_k)^r \right)^{-1} \left(4ND_X^2 (\gamma_{K-1} \rho_{K-1})^{r-1} \right. \\ &\quad + 2N^{-1} \sum_{k=0}^{K-1} (\gamma_k \rho_k)^{r+1} \rho_k^{-2} \left(6C_f^2 + 7v_f^2 + 7(N-1)(v_f^2 + C_f^2) \right) \\ &\quad \left. + 2N^{-1} \sum_{k=0}^{K-1} (\gamma_k \rho_k)^{r+1} (6C_F^2 + 7v_F^2 + 7(N-1)(v_F^2 + C_F^2)) + 2D_f \sum_{k=0}^{K-1} (\gamma_k \rho_k)^r \rho_k^{-1} \right).\end{aligned}$$

Hence, we obtain the infeasibility bound given by (16). $\square$

The main result of this section is presented in the following theorem where we obtain convergence rates for solving problem (6). In particular, we specify update rules for stepsize γ_k and penalty parameter ρ_k to guarantee this performance for Algorithm 1.

THEOREM 1 (RATE STATEMENTS AND ITERATION COMPLEXITY GUARANTEES). Consider Algorithm 1 applied to problem (6). Suppose $r \in [0, 1)$ is an arbitrary scalar. Let Assumptions 1 and 2 hold. Suppose, for any $k \geq 0$, the stepsize and the penalty sequence are given by

$$\gamma_k \triangleq \frac{\gamma_0}{\sqrt[3]{(k+1)^3}} \quad \text{and} \quad \rho_k \triangleq \rho_0 \sqrt[3]{k+1}.$$

Then, for all $K \geq 2^{\frac{2}{1-r}}$ the following statements hold.

(i) The convergence rate in terms of the suboptimality is given as

$$\mathbb{E}[f(\bar{y}_K)] - f^* \leq \left(\frac{D_X^2}{\gamma_0 \rho_0} + \frac{\gamma_0 \rho_0 \left((7-N^{-1})C_F^2 + 7v_F^2 + \frac{(7-N^{-1})C_f^2 + 7v_f^2}{\rho_0^2} \right)}{(1.5-r)N} \right) \frac{4\rho_0(2-r)N}{\sqrt[3]{K}}.$$

(ii) The convergence rate in terms of the infeasibility is given as

$$\mathbb{E}[\text{Gap}^*(\bar{y}_K)] \leq \left(\frac{D_X^2}{\gamma_0 \rho_0 \sqrt[3]{K}} + \frac{\gamma_0 \rho_0 \left((7-N^{-1})C_F^2 + 7v_F^2 + \frac{(7-N^{-1})C_f^2 + 7v_f^2}{\rho_0^2} \right)}{(1-r)N \sqrt[3]{K}} + \frac{D_f N^{-1}}{\rho_0(0.75-0.5r)} \right) \frac{4(2-r)N}{\sqrt[3]{K}}.$$

(iii) Given $\epsilon > 0$, let K_ϵ denote a deterministic integer to achieve $\mathbb{E}[f(\bar{y}_{K_\epsilon})] - f^* \leq \epsilon$ and $\mathbb{E}[\text{Gap}^*(\bar{y}_{K_\epsilon})] \leq \epsilon$. Then the total iteration complexity and also, the total sample complexity of Algorithm 1 are the same and are $O(N^4 \epsilon^{-4})$ where N denotes the number of blocks (In particular, in the Nash game, N denotes the number of players).

PROOF. (i) Substituting the update rules of γ_k and ρ_k in (15), we obtain

$$\begin{aligned}\mathbb{E}[f(\bar{y}_K)] - f^* &\leq \frac{4ND_X^2 (\gamma_{K-1} \rho_{K-1})^{r-1} \rho_{K-1} + 2N^{-1} \sum_{k=0}^{K-1} (\gamma_k \rho_k)^{1+r} \rho_k \left(\theta_F + \theta_f \rho_k^{-2} \right)}{\sum_{k=0}^{K-1} (\gamma_k \rho_k)^r} \\ &\leq \frac{4ND_X^2 \rho_0 (\gamma_0 \rho_0)^{r-1} K^{0.75-0.5r} + 2N^{-1} \rho_0 \left(\theta_F + \theta_f \rho_0^{-2} \right) (\gamma_0 \rho_0)^{1+r} \sum_{k=0}^{K-1} (k+1)^{-(0.25+0.5r)}}{(\gamma_0 \rho_0)^r \sum_{k=0}^{K-1} (k+1)^{-0.5r}}.\end{aligned}$$

Because $0 \leq r < 1$, note that both the terms $0.25 + 0.5r$ and $0.5r$ are nonnegative and smaller than 1. This implies that the conditions of Lemma 4 are met. Employing the bounds provided by Lemma 4, from the preceding inequality we have

$$\begin{aligned} \mathbb{E}[f(\bar{y}_K)] - f^* &\leq \frac{4ND_X^2 \rho_0 (\gamma_0 \rho_0)^{r-1} K^{0.75-0.5r} + 2N^{-1} \rho_0 (\theta_F + \theta_f \rho_0^{-2}) (\gamma_0 \rho_0)^{1+r} (0.75 - 0.5r)^{-1} K^{0.75-0.5r}}{0.5(1 - 0.5r)^{-1} (\gamma_0 \rho_0)^r K^{1-0.5r}} \\ &= \frac{(2 - r) (4ND_X^2 \rho_0 (\gamma_0 \rho_0)^{-1} + 2N^{-1} \rho_0 (\theta_F + \theta_f \rho_0^{-2}) (\gamma_0 \rho_0) (0.75 - 0.5r)^{-1})}{K^{0.25}}. \end{aligned}$$

Substituting θ_f and θ_F by their values and then, rearranging the terms we obtain the desired rate statement in (i). (ii) Next we derive the non-asymptotic rate statement in terms of the infeasibility. Substituting the update rules of γ_k and ρ_k in (16), and noting that γ_k and ρ_k^{-1} are nonincreasing, we obtain

$$\begin{aligned} \mathbb{E}[\text{Gap}^*(\bar{y}_K)] &\leq \frac{4ND_X^2 (\gamma_{K-1} \rho_{K-1})^{r-1} + 2N^{-1} \sum_{k=0}^{K-1} (\gamma_k \rho_k)^r (\theta_F \gamma_k \rho_k + \theta_f \gamma_k \rho_k^{-1} + 2ND_f \rho_k^{-1})}{\sum_{k=0}^{K-1} (\gamma_k \rho_k)^r} \\ &\leq \frac{4ND_X^2 (\gamma_{K-1} \rho_{K-1})^{r-1} + 2N^{-1} (\theta_F + \theta_f \rho_0^{-2}) \sum_{k=0}^{K-1} (\gamma_k \rho_k)^{r+1} + 4D_f \sum_{k=0}^{K-1} (\gamma_k \rho_k)^r \rho_k^{-1}}{\sum_{k=0}^{K-1} (\gamma_k \rho_k)^r} \\ &\leq \frac{4ND_X^2 (\gamma_0 \rho_0 K^{-0.5})^{r-1} + 2N^{-1} (\theta_F + \theta_f \rho_0^{-2}) (\gamma_0 \rho_0)^{1+r} \sum_{k=0}^{K-1} (k+1)^{-0.5(1+r)}}{(\gamma_0 \rho_0)^r \sum_{k=0}^{K-1} (k+1)^{-0.5r}} \\ &\quad + \frac{4D_f (\gamma_0 \rho_0)^r \rho_0^{-1} \sum_{k=0}^{K-1} (k+1)^{-0.5r-0.25}}{(\gamma_0 \rho_0)^r \sum_{k=0}^{K-1} (k+1)^{-0.5r}}. \end{aligned}$$

Employing the bounds provided by Lemma 4, from the preceding inequality we have

$$\begin{aligned} \mathbb{E}[\text{Gap}^*(\bar{y}_K)] &\leq \frac{4ND_X^2 (\gamma_0 \rho_0)^{-1} K^{-0.5(r-1)} + 2N^{-1} (\theta_F + \theta_f \rho_0^{-2}) (\gamma_0 \rho_0) (1 - 0.5(1+r))^{-1} K^{1-0.5(1+r)}}{0.5(1 - 0.5r)^{-1} K^{1-0.5r}} \\ &\quad + \frac{4D_f \rho_0^{-1} (1 - 0.5r - 0.25)^{-1} K^{1-0.5r-0.25}}{0.5(1 - 0.5r)^{-1} K^{1-0.5r}} \\ &\leq (2 - r) \frac{4ND_X^2 (\gamma_0 \rho_0)^{-1} + 4N^{-1} (\theta_F + \theta_f \rho_0^{-2}) (\gamma_0 \rho_0) (1 - r)^{-1}}{K^{0.5}} \\ &\quad + (2 - r) \frac{4D_f \rho_0^{-1} (0.75 - 0.5r)^{-1}}{K^{0.25}}. \end{aligned}$$

The rate statement in (ii) can be obtained by substituting θ_f and θ_F by their values and then, rearranging the terms.

(iii) The result of part (iii) holds directly from the rate statements in parts (i) and (ii). $\square$

5 APPROXIMATING THE PRICE OF STABILITY

Our goal in this section lies in devising a stochastic scheme for approximating the price of stability, defined by (2), in monotone stochastic Nash games. The proposed scheme includes three main steps described as follows:

(i) Employing Algorithm 1 for approximating a solution to the optimization problem (6).

(ii) Employing a stochastic approximation method for approximating a solution to the nonsmooth stochastic optimization problem $\min_{x \in X} \mathbb{E}[f(x, \xi)]$. This can be done through a host of well-known methods including the stochastic subgradient [32, 39] and its accelerated smoothed variants [18]. Another avenue for solving this class of problems is stochastic extra-subgradient methods [15, 22, 31, 41, 43].

(iii) Lastly, given the two approximate optimal solutions in (i) and (ii), we estimate the objective function value $\mathbb{E}[f(x, \xi)]$ at each solution. The PoS is then approximated by dividing the sample average approximation of optimal objective value of problem (6) by that of $\min_{x \in X} \mathbb{E}[f(x, \xi)]$.

An example of this scheme is presented by Algorithm 2. Here, vectors $y_{k,1}$ and $x_{k,1}$ are generated by Algorithm 1, while $y_{k,2}$ and $x_{k,2}$ are generated by a standard stochastic extra-subgradient method for solving $\min_{x \in X} \mathbb{E}[f(x, \xi)]$. We provide the following remark to make clarifications about this scheme.

Remark 7. As mentioned earlier, we do have several options in employing a method for solving the canonical nonsmooth stochastic optimization problem $\min_{x \in X} \mathbb{E}[f(x, \xi)]$. Here, we use the stochastic extra-subgradient method that is known to achieve the convergence rate of the order $\frac{1}{\sqrt{K}}$ when employing a suitable weighted averaging scheme specified by (35) (cf. [43]). We also note that Algorithm 2 can be compactly presented by the two extra-subgradient schemes, separately. However, we note that there are different groups of random samples generated in Algorithm 2 and the analysis of the scheme relies on what assumptions we make on these samples, presented in the following.

ASSUMPTION 3. Let the following statements hold.

- (i) The random samples $\{\xi_{k,1}\}_{k=0}^{K-1}$, $\{\tilde{\xi}_{k,1}\}_{k=0}^{K-1}$, $\{\xi_{k,2}\}_{k=0}^{K-1}$, $\{\tilde{\xi}_{k,2}\}_{k=0}^{K-1}$, and $\{\zeta_t\}_{t=0}^{M-1}$ are i.i.d. associated with the probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Also, $\{\tilde{i}_{k,1}\}_{k=0}^{K-1}$, $\{\tilde{i}_{k,1}\}_{k=0}^{K-1}$, $\{i_{k,2}\}_{k=0}^{K-1}$, and $\{\tilde{i}_{k,2}\}_{k=0}^{K-1}$ are i.i.d. uniformly distributed within the range $\{1, \dots, N\}$. Additionally, all the aforementioned random variables are independent from each other.
- (ii) $f(\bullet, \xi)$ is an unbiased estimator of the deterministic function $f(\bullet)$.

To approximate the PoS, we need upper and lower bounds for suboptimality of problem (6). We established the upper bound in Theorem 1. Now we obtain the lower bound considering the following weak sharpness assumption.

ASSUMPTION 4 (WEAK SHARPNESS [8]). The variational inequality problem $\text{VI}(X, F)$ satisfies the weak sharpness property implying that there exists an $\alpha > 0$ such that $(x - x^*)^T F(x^*) \geq \alpha \text{dist}(x, X^*)$ for any $x \in X^*$, where X^* denotes the solution set of $\text{VI}(X, F)$.

COROLLARY 2. Under the premises of Theorem 1 and considering Assumption 4, we have for all $K \geq 2$

$$-\frac{O(N)}{\sqrt[4]{K}} \leq \mathbb{E}[f(\bar{y}_K) - f^*] \leq \frac{O(N)}{\sqrt[4]{K}}.$$

PROOF. From Assumption 4, we know that there exists $\alpha > 0$ such that $\mathbb{E}[\text{dist}(\bar{y}_K, X^*)] \leq \frac{1}{\alpha} \mathbb{E}[\text{Gap}^*(\bar{y}_K)]$. Moreover, since X^* is a compact set, there exists $\hat{y}^* \in X^*$ such that $\text{dist}(\bar{y}_K, X^*) = \min_{y \in X^*} \|y - \bar{y}_K\| = \|\hat{y}^* - \bar{y}_K\|$. Therefore, using the result of Theorem 1, we have

$$\mathbb{E}[\|\hat{y}^* - \bar{y}_K\|] \leq \frac{1}{\alpha} \mathbb{E}[\text{Gap}^*(\bar{y}_K)] \leq \frac{O(N)}{\sqrt[4]{K}}. \quad (28)$$

Moreover, using convexity of f and Cauchy-Schwartz inequality, we conclude that

$$\mathbb{E}[f(\bar{y}_K)] - f^* \geq \mathbb{E}[f(\bar{y}_K)] - f(\hat{y}^*) \geq \mathbb{E}[\nabla f(\hat{y}^*)^T (\bar{y}_K - \hat{y}^*)] \geq -\|\nabla f(\hat{y}^*)\| \mathbb{E}[\|\bar{y}_K - \hat{y}^*\|] \geq -\frac{O(N)}{\sqrt[4]{K}},$$

where in the first inequality we used the fact that $f^* \leq f(\hat{y}^*)$ and the last inequality follows from (28) and the fact that the gradient is bounded. $\square$

The main result in this section is presented in the following

LEMMA 8 (ERROR BOUNDS IN APPROXIMATING THE PoS). Consider Algorithm 2. Let Assumptions 1, 2, 3, and 4 hold. Suppose, $r_1, r_2 \in [0, 1)$ be fixed scalars and for any $k \geq 0$, let us define

$$\gamma_{k,1} \triangleq \frac{\gamma_{0,1}}{\sqrt[4]{(k+1)^3}}, \quad \rho_k \triangleq \rho_0 \sqrt[4]{k+1}, \quad \gamma_{k,2} \triangleq \frac{\gamma_{0,2}}{\sqrt{k+1}}.$$

Then the following holds for all $K \geq 1$.

$$-O\left(\frac{1}{\sqrt[4]{K}}\right) \leq \frac{\mathbb{E}[\hat{f}(\bar{y}_{K,1})]}{\mathbb{E}[\hat{f}(\bar{y}_{K,2})]} - \text{PoS} \leq O\left(\frac{1}{\sqrt[4]{K}}\right). \quad (29)$$

Algorithm 2 Approximating PoS using randomized stochastic extra-gradient schemes

- 1: **initialization:** Set random initial points $x_{0,1}, x_{0,2}, y_{0,1}, y_{0,2} \in X$, initial stepsizes $\gamma_{0,1}, \gamma_{0,2} > 0$, scalar $0 \leq r_1, r_2 < 1$, $\bar{y}_{0,1} = \bar{y}_{0,2} := y_0$, $\Gamma_{0,1} = \Gamma_{0,2} := 0$, $S_{0,1} = S_{0,2} := 0$.
- 2: **for** $k = 0, 1, \dots, K-1$ **do**
- 3: Generate $i_{k,1}, \tilde{i}_{k,1}, i_{k,2}$, and $\tilde{i}_{k,2}$ uniformly from $\{1, \dots, N\}$.
- 4: Generate $\xi_{k,1}, \tilde{\xi}_{k,1}, \xi_{k,2}$, and $\tilde{\xi}_{k,2}$ as realizations of the random vector ξ .
- 5: Update the variables $y_{k,1}, x_{k,1}, y_{k,2}$, and $x_{k,2}$ as

$$y_{k+1,1}^{(i)} := \begin{cases} \mathcal{P}_{X_i} \left(x_{k,1}^{(i)} - \gamma_{k,1} (\tilde{\nabla}_i f(x_{k,1}, \tilde{\xi}_{k,1}) + \rho_k F_i(x_{k,1}, \tilde{\xi}_{k,1})) \right) & \text{if } i = \tilde{i}_{k,1}, \\ x_{k,1}^{(i)} & \text{if } i \neq \tilde{i}_{k,1}, \end{cases} \quad (30)$$

$$x_{k+1,1}^{(i)} := \begin{cases} \mathcal{P}_{X_i} \left(x_{k,1}^{(i)} - \gamma_{k,1} (\tilde{\nabla}_i f(y_{k+1,1}, \xi_{k,1}) + \rho_k F_i(y_{k+1,1}, \xi_{k,1})) \right) & \text{if } i = i_{k,1}, \\ x_{k,1}^{(i)} & \text{if } i \neq i_{k,1}, \end{cases} \quad (31)$$

$$y_{k+1,2}^{(i)} := \begin{cases} \mathcal{P}_{X_i} \left(x_{k,2}^{(i)} - \gamma_{k,2} \tilde{\nabla}_i f(x_{k,2}, \tilde{\xi}_{k,2}) \right) & \text{if } i = \tilde{i}_{k,2}, \\ x_{k,2}^{(i)} & \text{if } i \neq \tilde{i}_{k,2}, \end{cases} \quad (32)$$

$$x_{k+1,2}^{(i)} := \begin{cases} \mathcal{P}_{X_i} \left(x_{k,2}^{(i)} - \gamma_{k,2} \tilde{\nabla}_i f(y_{k+1,2}, \xi_{k,2}) \right) & \text{if } i = i_{k,2}, \\ x_{k,2}^{(i)} & \text{if } i \neq i_{k,2}. \end{cases} \quad (33)$$

- 6: Update $\Gamma_{k,1}, \Gamma_{k,2}, \bar{y}_{k,1}$, and $\bar{y}_{k,2}$ using the following recursions.

$$\Gamma_{k+1,1} := \Gamma_{k,1} + (\gamma_{k,1} \rho_k)^{r_1}, \quad \bar{y}_{k+1,1} := \frac{\Gamma_{k,1} \bar{y}_{k,1} + (\gamma_{k,1} \rho_k)^{r_1} y_{k+1,1}}{\Gamma_{k+1,1}}, \quad (34)$$

$$\Gamma_{k+1,2} := \Gamma_{k,2} + \gamma_{k,2}^{r_2}, \quad \bar{y}_{k+1,2} := \frac{\Gamma_{k,2} \bar{y}_{k,2} + \gamma_{k,2}^{r_2} y_{k+1,2}}{\Gamma_{k+1,2}}. \quad (35)$$

- 7: **end for**
- 8: Generate the batch of samples $\{\zeta_t\}$ as i.i.d realizations of ξ , for $t = 0, \dots, M-1$
- 9: Evaluate sample average approximations $\hat{f}_M(\bar{y}_{K,1}) := \frac{1}{M} \sum_{t=0}^{M-1} f(\bar{y}_{K,1}, \zeta_t)$ and $\hat{f}_M(\bar{y}_{K,2}) := \frac{1}{M} \sum_{t=0}^{M-1} f(\bar{y}_{K,2}, \zeta_t)$
- 10: Return $\frac{\hat{f}_M(\bar{y}_{K,1})}{\hat{f}_M(\bar{y}_{K,2})}$.

PROOF. We utilize the following notation in the proof.

$$\begin{aligned} \mathcal{F}_{k,1} &\triangleq \cup_{t=0}^k \{\tilde{\xi}_{t,1}, \tilde{i}_{t,1}, \xi_{t,1}, i_{t,1}\} \cup \{x_{0,1}, y_{0,1}\}, & \text{for all } k \in \{0, \dots, K-1\}, \\ \mathcal{F}_{k,2} &\triangleq \cup_{t=0}^k \{\tilde{\xi}_{t,2}, \tilde{i}_{t,2}, \xi_{t,2}, i_{t,2}\} \cup \{x_{0,2}, y_{0,2}\}, & \text{for all } k \in \{0, \dots, K-1\}. \end{aligned}$$

Recall the definitions $\hat{f}_M(\bar{y}_{K,1}) := \frac{1}{M} \sum_{t=0}^{M-1} f(\bar{y}_{K,1}, \xi_t)$ and $\hat{f}_M(\bar{y}_{K,2}) := \frac{1}{M} \sum_{t=0}^{M-1} f(\bar{y}_{K,2}, \xi_t)$. Then, we may write

$$\mathbb{E}[\hat{f}_M(\bar{y}_{K,1})] = \mathbb{E}\left[\mathbb{E}\left[\hat{f}_M(\bar{y}_{K,1}) \mid \mathcal{F}_{K-1,1}\right]\right] = \mathbb{E}\left[\mathbb{E}\left[\frac{1}{M} \sum_{t=0}^{M-1} f(\bar{y}_{K,1}, \xi_t) \mid \mathcal{F}_{K-1,1}\right]\right] = \mathbb{E}[f(\bar{y}_{K,1})].$$

From the preceding relation and Theorem 1 we have

$$-\frac{O(N)}{\sqrt[4]{K}} \leq \mathbb{E}[\hat{f}_M(\bar{y}_{K,1})] - f^* \leq \frac{O(N)}{\sqrt[4]{K}},$$

where f^* denotes the optimal objective value of problem (6). Let us define $f_{Opt}^* \triangleq \min_{x \in X} \mathbb{E}[f(x, \xi)]$. Similarly,

$$\mathbb{E}[\hat{f}_M(\bar{y}_{K,2})] = \mathbb{E}\left[\mathbb{E}\left[\hat{f}_M(\bar{y}_{K,2}) \mid \mathcal{F}_{K-1,2}\right]\right] = \mathbb{E}\left[\mathbb{E}\left[\frac{1}{M} \sum_{t=0}^{M-1} f(\bar{y}_{K,2}, \xi_t) \mid \mathcal{F}_{K-1,2}\right]\right] = \mathbb{E}[f(\bar{y}_{K,2})].$$

and we also have that

$$0 \leq \mathbb{E}[\hat{f}_M(\bar{y}_{K,2})] - f_{Opt}^* \leq \frac{O(N)}{\sqrt[4]{K}}.$$

We show the result holds when $f^*, f_{Opt}^* \geq 0$ and one can verify that the result also holds for other cases. From the definition of PoS given by (2) and the two preceding inequalities, we may write

$$\frac{\mathbb{E}[\hat{f}_M(\bar{y}_{K,1})]}{\mathbb{E}[\hat{f}_M(\bar{y}_{K,2})]} \leq \frac{f^* + \frac{O(N)}{\sqrt[4]{K}}}{f_{Opt}^*} = \frac{f^*}{f_{Opt}^*} + \frac{O(N)}{\sqrt[4]{K}} = PoS + \frac{O(N)}{\sqrt[4]{K}}.$$

We can also write

$$\frac{\mathbb{E}[\hat{f}_M(\bar{y}_{K,1})]}{\mathbb{E}[\hat{f}_M(\bar{y}_{K,2})]} \geq \frac{f^* - \frac{O(N)}{\sqrt[4]{K}}}{f_{Opt}^* + \frac{O(N)}{\sqrt[4]{K}}} = \left(\frac{1 - \frac{O(N)}{\sqrt[4]{K}}}{1 + \frac{O(N)}{\sqrt[4]{K}}} \right) \times PoS \implies \frac{\mathbb{E}[\hat{f}_M(\bar{y}_{K,1})]}{\mathbb{E}[\hat{f}_M(\bar{y}_{K,2})]} - PoS \geq -\frac{O(N)}{\sqrt[4]{K}}.$$

Thus, in view of the two preceding inequalities, the result holds. $\square$

Remark 8. We note that in Algorithm 2, in using the extra-gradient method employed for solving $\min_{x \in X} \mathbb{E}[f(x, \xi)]$, we do not use any penalization. However, in solving $\min_{x \in \text{SOL}(X, \mathbb{E}[F(\bullet, \xi)])} \mathbb{E}[f(x, \xi)]$, we employ Algorithm 1 where we utilize iterative penalization. Intuitively speaking, problem $\min_{x \in X} \mathbb{E}[f(x, \xi)]$ can be viewed as a special case of $\min_{x \in \text{SOL}(X, \mathbb{E}[F(\bullet, \xi)])} \mathbb{E}[f(x, \xi)]$ where the mapping $F(x)$ is zero for all x . As such, we suppress the penalization in solving $\min_{x \in X} \mathbb{E}[f(x, \xi)]$. This allows us to use larger stepsizes in solving $\min_{x \in X} \mathbb{E}[f(x, \xi)]$ and obtain faster convergence for the optimality metric.

Moreover, in Algorithm 2, in solving $\min_{x \in X} \mathbb{E}[f(x, \xi)]$, we employ the averaging weights $\frac{(y_{k,2})^r}{\sum_{j=0}^{K-1} (y_{j,2})^r}$. However, in solving $\min_{x \in \text{SOL}(X, \mathbb{E}[F(\bullet, \xi)])} \mathbb{E}[f(x, \xi)]$, we use the averaging weights $\frac{(y_{k,1}\rho_k)^r}{\sum_{j=0}^{K-1} (y_{j,1}\rho_j)^r}$. We note that in view of the choices of the stepsizes and penalty parameter in Lemma 8, the averaging weights of the two schemes are indeed almost identical. This is because in Lemma 8, assuming that $\gamma_{0,1}\rho_0 = \gamma_{0,2}$, we have $\gamma_{k,1}\rho_k = \gamma_{k,2}$ for all k .

6 NUMERICAL EXPERIMENTS

In this section we present the performance of the proposed schemes in estimating the price of stability for a stochastic Nash Cournot competition over a network. Cournot game is one of the most popular and amongst the first economic models for formulating the competition among multiple firms (see [12, 21] for the applications of Cournot models in imperfectly competitive power markets and also, rate control in communication networks). The Cournot model is described as follows. Consider a collection of N firms who compete over a network with J nodes to sell a product. The strategy of firm $i \in \{1, \dots, N\}$ is characterized by the decision variables y_{ij} and s_{ij} , denoting the generation and sales of firm i at the node j , respectively. Compactly, the decision variables of the i^{th} firm is denoted by $x^{(i)} \triangleq (y_i, s_i) \in \mathbb{R}^{2J}$ where we assume that $y_i \triangleq (y_{i1}, \dots, y_{iJ})$ and $s_i \triangleq (s_{i1}, \dots, s_{iJ})$. The goal of the i^{th} firm lies in minimizing the expected value of a net cost function $f_i(x^{(i)}, x^{(-i)}, \xi)$ over the network over the strategy set X_i . This optimization problem for the firm i is defined as

$$\begin{aligned} \text{minimize} \quad & \mathbb{E} \left[f_i \left(x^{(i)}, x^{(-i)}, \xi \right) \right] \triangleq \mathbb{E} \left[\sum_{j=1}^J c_{ij}(y_{ij}) - \sum_{j=1}^J s_{ij} p_j(\bar{s}_j, \xi) \right] \\ \text{Subject to.} \quad & x^{(i)} \in X_i \triangleq \left\{ (y_i, s_i) \mid y_{ij} \leq \mathcal{B}_{ij}, \sum_{j=1}^J y_{ij} = \sum_{j=1}^J s_{ij}, y_{ij}, s_{ij} \geq 0, \text{ for all } j = 1, \dots, J \right\}. \end{aligned}$$

Here, $\bar{s}_j \triangleq \sum_{i=1}^N s_{ij}$ denotes the aggregate sales from all the firms at node j , $p_j : \mathbb{R} \times \Omega \rightarrow \mathbb{R}$ denotes the price function characterized in terms of the aggregate sales at the node j and a random variable ξ , and $c_{ij} : \mathbb{R} \rightarrow \mathbb{R}$ denotes the production cost function of firm i at node j . The price functions are given as $p_j(\bar{s}_j, \xi) \triangleq \alpha_j(\xi) - \beta_j(\bar{s}_j)^\sigma$, where $\alpha_j(\xi)$ is a random positive variable, β_j is a positive scalar, and $\sigma \geq 1$. We assume that cost functions are linear and the transportation costs are zero. The constraint $y_{ij} \leq \mathcal{B}_{ij}$ states that the generation is capacitated where $\mathcal{B}_{ij}$ is a positive scalar for all i and j . Similar to [26], in defining a global objective function for the price of stability, we consider the Marshallian aggregate surplus function defined as

$$\mathbb{E}[f(x, \xi)] \triangleq \sum_{i=1}^N \mathbb{E} \left[f_i \left(x^{(i)}, x^{(-i)}, \xi \right) \right].$$

It has been shown [23] that when $\sigma \geq 1$, f is convex and also, when either $\sigma = 1$ or $1 < \sigma \leq 3$ and $N \leq \frac{3\sigma-1}{\sigma-1}$, the mapping associated with the Cournot game, i.e., $F(x) \triangleq (\nabla_{x^{(1)}} \mathbb{E}[f_1(x, \xi)], \dots, \nabla_{x^{(N)}} \mathbb{E}[f_N(x, \xi)])$ is merely monotone.

Experiments and set-up. We compare the performance of Algorithm 1 with that of the two existing methods, namely aRB-IRG in [26] and the sequential regularization (SR) scheme (cf. [12, 26]). Note that both the SR scheme and aRB-IRG can only use deterministic gradients. To apply these two methods, we use a sample average approximation scheme by assuming that the deterministic gradient is approximated using a batch size of 1000 random samples. In Algorithm 1, however, we can use stochastic gradients (using a single sample ξ). In both Algorithm 1 and aRB-IRG, we employ a randomized block-coordinate scheme with N number of blocks, where N is the number of firms. We consider four different settings in our simulation results, where they differ in terms of the choices of the initial stepsize, the initial regularization parameter used in aRB-IRG, and the initial penalty parameter. For each setting, we implement the three methods on two different Cournot games, one with 4 players over a network with 5 nodes, and another one with 10 players over a network with 2 nodes. We assume that $\alpha_j(\xi)$ is uniformly distributed for all the agents. To compare the simulation results, we generate 15 independent sample-paths for any of the schemes that are stochastic and/or randomized.

Table 1. The four settings for the algorithm parameters

Algorithm	Parameter(s)	Setting 1	Setting 2	Setting 3	Setting 4
SR scheme	γ_0	0.1	0.1	1	1
aRB-IRG	(γ_0, η_0)	(0.1,0.1)	(0.1,1)	(1,0.1)	(1,1)
aR-IP-SeG	(γ_0, ρ_0)	(0.01,10)	(0.1,1)	(0.1,10)	(1,1)

Results and insights. The simulation results are presented in Figures 3-4, and 5. Note that the legend for Figures 3-4 is presented in Figure 2. Several observations can be made: (i) As it can be seen in Figures 3-4, Algorithm 1 outperforms the other two methods in almost all the scenarios. We note that a smaller gap function value implies a smaller infeasibility for the solution iterate. However, because the solution iterate may be infeasible during the implementation of aRB-IRG and aR-IP-SeG, a smaller objective value may not necessarily imply a better solution. Instead, when comparing the objective function metric in the figures, it is important to observe how fast the objective value of each method reaches to a stable value. (ii) Although both Algorithm 1 and aRB-IRG are equipped with the same convergence speeds, Algorithm 1 enjoys a better performance with respect to the run-time. This is because it uses stochastic gradients that are cheaper to compute in contrast with the sample average gradients used in aRB-IRG. (iii) We do observe that as the size of the problem increases in terms of the number of players and the size of the network, the performance of all the schemes is downgraded. However, Algorithm 1 seems to stay robust across most settings and often outperforms the other two methods. (vi) In estimating the PoS in Figure 5, the methods seem to converge to a PoS smaller than one. This is because in this numerical experiment, we have considered the minimization of the negative of the profit function. As such, the optimal objective values of the minimization problems become negative. Consequently, the PoS is theoretically less than or equal to one. This is indeed aligned with the findings in Figure 5. Supplementary numerical experiments can be found in the extended version of the paper [19].

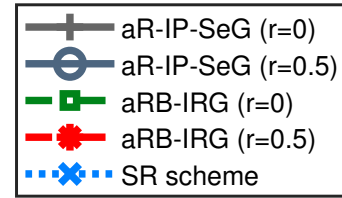


Fig. 2. The figure legend used in the numerical experiments in Figures 3-4

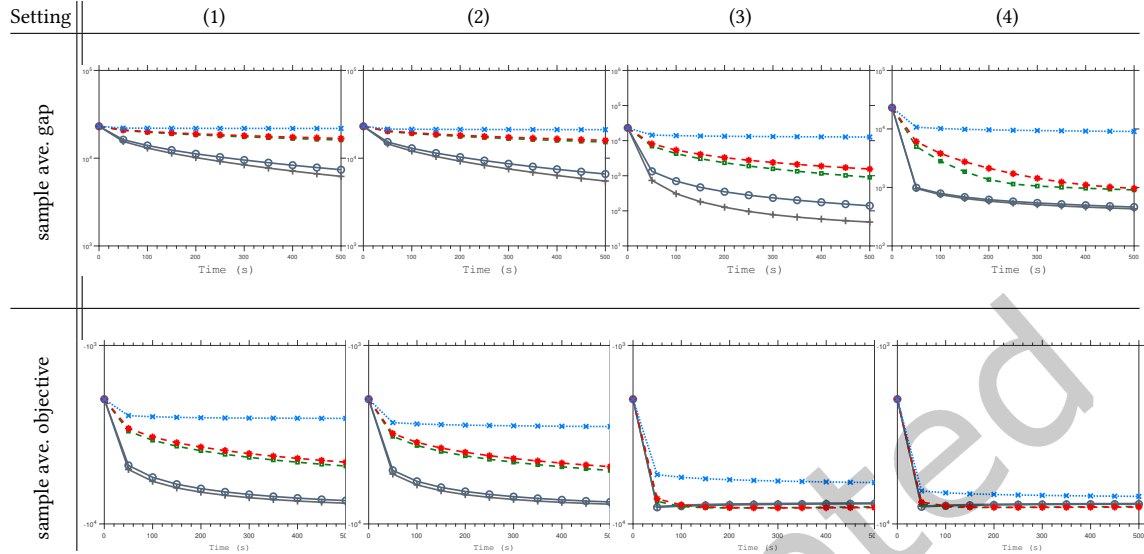


Fig. 3. Simulation results for a stochastic Nash Cournot game with 4 players over a network with 5 nodes, comparing Algorithm (1) with other existing methods for solving problem (6).

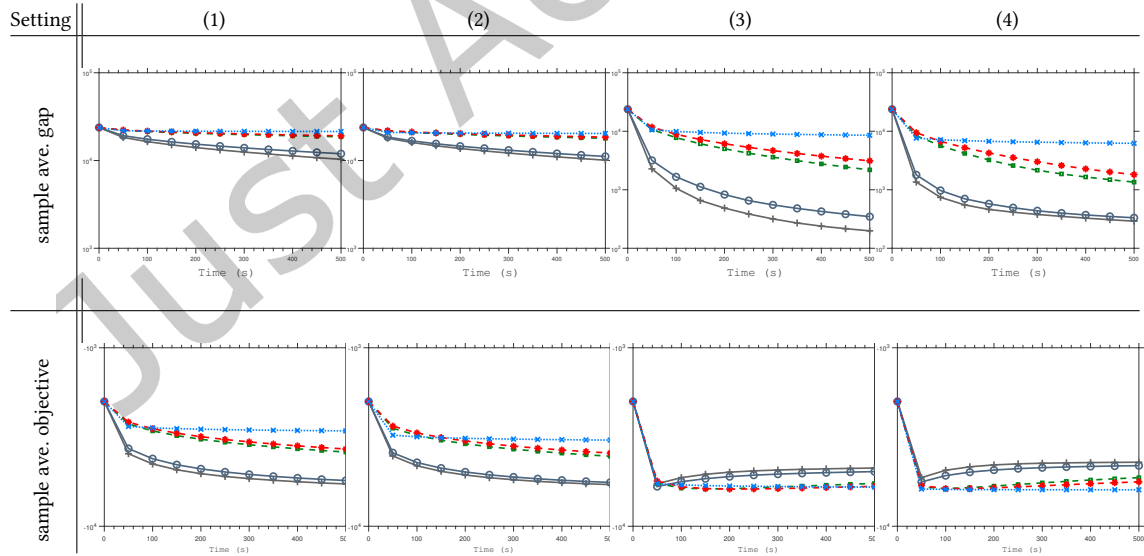
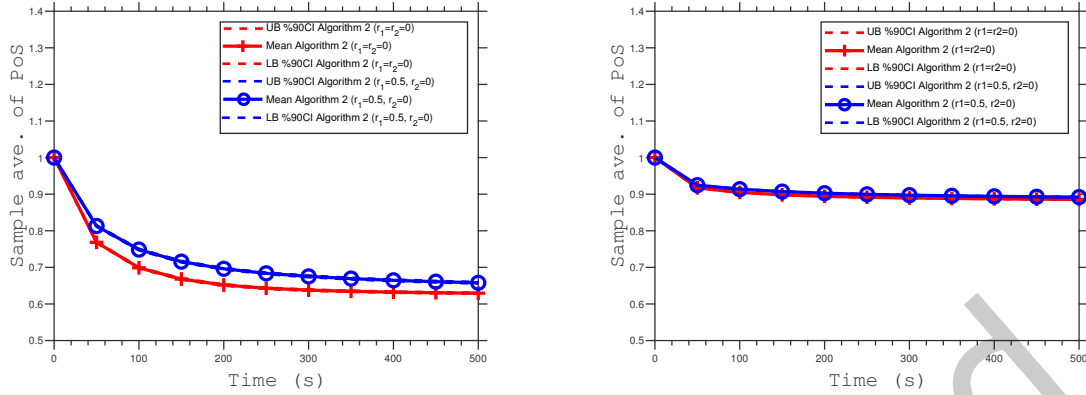


Fig. 4. Simulation results for a stochastic Nash Cournot game with 10 players over a network with 2 nodes, comparing Algorithm (1) with other existing methods for solving problem (6).



(a) Cournot game with 10 players and network 2 nodes

(b) Cournot game with 4 players and network 5 nodes

Fig. 5. Performance of Algorithm 2 in estimating PoS. 90% confidence intervals become tighter as the scheme proceeds.

7 ACKNOWLEDGMENTS

This work is supported by the NSF CAREER grant ECCS-1944500, NSF grant ECCS-2231863, and ONR grant N00014-22-1-2757.

REFERENCES

- [1] E. Anshelevich, A. Dasgupta, J. Kleinberg, E. Tardos, T. Wexler, and T. Roughgarden, *The price of stability for network design with fair cost allocation*, SIAM Journal on Computing **38** (2008), no. 4, 1602–1623.
- [2] D. P. Bertsekas, *Nonlinear programming*, Journal of the Operational Research Society **48** (1997), no. 3, 334–334.
- [3] D. P. Bertsekas, A. Nedić, and A. Ozdaglar, *Convex analysis and optimization*, vol. 1, Athena Scientific, 2003.
- [4] M. Broadie, D. M. Cicek, and A. Zeevi, *Multidimensional stochastic approximation: Adaptive algorithms and applications*, ACM Transactions on Modeling and Computer Simulation (TOMACS) **24** (2014), no. 1, 1–28.
- [5] Y. Censor, A. Gibali, and S. Reich, *The subgradient extragradient method for solving variational inequalities in hilbert space*, Journal of Optimization Theory and Applications **148** (2011), no. 2, 318–335.
- [6] ———, *Extensions of Korpelevich’s extragradient method for the variational inequality problem in Euclidean space*, Optimization **61** (2012), no. 9, 1119–1132.
- [7] Y. Chen, G. Lan, and Y. Ouyang, *Accelerated schemes for a class of variational inequalities*, Mathematical Programming **165** (2017), no. 1, 113–149.
- [8] S. Cui and U. V. Shanbhag, *On the analysis of reflected gradient and splitting methods for monotone stochastic variational inequality problems*, 2016 IEEE 55th Conference on Decision and Control (CDC), IEEE, 2016, pp. 4510–4515.
- [9] ———, *On the analysis of variance-reduced and randomized projection variants of single projection schemes for monotone stochastic variational inequality problems*, Set-Valued and Variational Analysis **29** (2021), no. 2, 453–499.
- [10] Y. Deng, M. M. Kamani, and M. Mahdavi, *Distributionally robust federated averaging*, Advances in Neural Information Processing Systems (H. Larochelle, M. Ranzato, R. Hadsell, M.F. Balcan, and H. Lin, eds.), vol. 33, Curran Associates, Inc., 2020, pp. 15111–15122.
- [11] P. Dubey, *Inefficiency of Nash equilibria*, Mathematics of Operations Research **11** (1986), no. 1, 1–8.
- [12] F. Facchinei and J.-S. Pang, *Finite-dimensional variational inequalities and complementarity problems. Vols. I,II*, Springer Series in Operations Research, Springer-Verlag, New York, 2003.
- [13] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, *Generative adversarial nets*, Advances in Neural Information Processing Systems (Z. Ghahramani, M. Welling, C. Cortes, N. Lawrence, and K.Q. Weinberger, eds.), vol. 27, Curran Associates, Inc., 2014.
- [14] Y.-G. Hsieh, F. Iutzeler, J. Malick, and P. Mertikopoulos, *On the convergence of single-call stochastic extra-gradient methods*, Advances in Neural Information Processing Systems **32** (2019).
- [15] A. N. Iusem, A. Jofré, R. I. Oliveira, and P. Thompson, *Extragradient method with variance reduction for stochastic variational inequalities*, SIAM Journal on Optimization **27** (2017), no. 2, 686–724.

- [16] A. N. Iusem and M. Nasri, *Korpelevich's method for variational inequality problems in banach spaces*, Journal of Global Optimization **50** (2011), no. 1, 59–76.
- [17] A. Jalilzadeh and U. V. Shanbhag, *A proximal-point algorithm with variable sample-sizes (ppawss) for monotone stochastic variational inequality problems*, 2019 Winter Simulation Conference (WSC), IEEE, 2019, pp. 3551–3562.
- [18] A. Jalilzadeh, U. V. Shanbhag, J. H. Blanchet, and P. W. Glynn, *Smoothed variable sample-size accelerated proximal methods for nonsmooth stochastic convex programs*, Stochastic Systems **12** (2022), no. 4, 373–410.
- [19] A. Jalilzadeh, F. Yousefian, and M. Ebrahimi, *Stochastic approximation for estimating the price of stability in stochastic Nash games*, arXiv preprint: <https://arxiv.org/abs/2203.01271>.
- [20] H. Jiang and H. Xu, *Stochastic approximation approaches to the stochastic variational inequality problem*, IEEE Transactions on Automatic Control **53** (2008), no. 6, 1462–1475.
- [21] R. Johari, *Efficiency loss in market mechanisms for resource allocation*, Ph.D. thesis, MIT, 2004.
- [22] A. Juditsky, A. Nemirovski, and C. Tauvel, *Solving variational inequalities with stochastic mirror-prox algorithm*, Stochastic Systems **1** (2011), no. 1, 17–58.
- [23] A. Kannan and U. V. Shanbhag, *Distributed computation of equilibria in monotone Nash games via iterative regularization techniques*, SIAM Journal on Optimization **22** (2012), no. 4, 1177–1205.
- [24] A. Kannan and U. V. Shanbhag, *The pseudomonotone stochastic variational inequality problem: Analytical statements and stochastic extragradient schemes*, 2014 American Control Conference, IEEE, 2014, pp. 2930–2935.
- [25] ———, *Optimal stochastic extragradient schemes for pseudomonotone stochastic variational inequality problems and their variants*, Computational Optimization and Applications **74** (2019), no. 3, 779–820.
- [26] H. D. Kaushik and F. Yousefian, *A method with convergence rates for optimization problems with variational inequality constraints*, SIAM Journal on Optimization **31** (2021), no. 3, 2171–2198.
- [27] G. M. Korpelevich, *The extragradient method for finding saddle points and other problems*, Matecon **12** (1976), 747–756.
- [28] J. Koshal, A. Nedić, and U. V. Shanbhag, *Regularized iterative stochastic approximation methods for variational inequality problems*, IEEE Transactions on Automatic Control **58**(3) (2013), 594–609.
- [29] P. Marcotte and D. Zhu, *Weak sharp solutions of variational inequalities*, SIAM Journal on Optimization **9** (1998), no. 1, 179–189.
- [30] A. Nedić and A. Ozdaglar, *Subgradient methods for saddle-point problems*, Journal of Optimization Theory and Applications **142** (2009), 205–228.
- [31] 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 **15** (2004), no. 1, 229–251.
- [32] A. Nemirovski, A. Juditsky, G. Lan, and A. Shapiro, *Robust stochastic approximation approach to stochastic programming*, SIAM Journal on Optimization **19** (2009), no. 4, 1574–1609.
- [33] B. T. Polyak and A. B. Juditsky, *Acceleration of stochastic approximation by averaging*, SIAM J. Control Optim. **30** (1992), no. 4, 838–855. MR 1167814 (93g:62110)
- [34] H. Robbins and S. Monro, *A stochastic approximation method*, Ann. Math. Statistics **22** (1951), 400–407.
- [35] T. Roughgarden, *Algorithmic game theory*, Communications of the ACM **53** (2010), no. 7, 78–86.
- [36] M. Sibony, *Méthodes itératives pour les équations et inéquations aux dérivées partielles non linéaires de type monotone*, Calcolo **7** (1970), no. 1, 65–183.
- [37] v. Neumann, *Zur theorie der gesellschaftsspiele*, Mathematische Annalen **19** (1928), no. 2, 295–320.
- [38] D. Xu, S. Yuan, L. Zhang, and X. Wu, *Fairgan: Fairness-aware generative adversarial networks*, 2018 IEEE International Conference on Big Data (Big Data), IEEE, 2018, pp. 570–575.
- [39] F. Yousefian, A. Nedić, and U. V. Shanbhag, *On stochastic gradient and subgradient methods with adaptive steplength sequences*, Automatica **48** (2012), no. 1, 56–67, An extended version of the paper available at: <http://arxiv.org/abs/1105.4549>.
- [40] ———, *A regularized smoothing stochastic approximation (RSSA) algorithm for stochastic variational inequality problems*, 2013 Winter Simulations Conference (WSC), IEEE, 2013, pp. 933–944.
- [41] ———, *Optimal robust smoothing extragradient algorithms for stochastic variational inequality problems*, 53rd IEEE Conference on Decision and Control, IEEE, 2014, pp. 5831–5836.
- [42] ———, *On smoothing, regularization, and averaging in stochastic approximation methods for stochastic variational inequality problems*, Mathematical Programming **165** (2017), no. 1, 391–431.
- [43] ———, *On stochastic mirror-prox algorithms for stochastic cartesian variational inequalities: Randomized block coordinate and optimal averaging schemes*, Set-Valued and Variational Analysis **26** (2018), no. 4, 789–819.