SIAM J. OPTIM.
Vol. 31, No. 3, pp. 2171–2198

A METHOD WITH CONVERGENCE RATES FOR OPTIMIZATION PROBLEMS WITH VARIATIONAL INEQUALITY CONSTRAINTS*

HARSHAL D. KAUSHIK† AND FARZAD YOUSEFIAN†

Abstract. We consider a class of optimization problems with Cartesian variational inequality (CVI) constraints, where the objective function is convex and the CVI is associated with a monotone mapping and a convex Cartesian product set. This mathematical formulation captures a wide range of optimization problems including those complicated by the presence of equilibrium constraints, complementarity constraints, or an inner-level large scale optimization problem. In particular, an important motivating application arises from the notion of efficiency estimation of equilibria in multi-agent networks, e.g., communication networks and power systems. In the literature, the iteration complexity of the existing solution methods for optimization problems with CVI constraints appears to be unknown. To address this shortcoming, we develop a first-order method called averaging randomized block iteratively regularized gradient (aRB-IRG). The main contributions include the following: (i) In the case where the associated set of the CVI is bounded and the objective function is nondifferentiable and convex, we derive new nonasymptotic suboptimality and infeasibility convergence rate statements in an ergodic sense. We also obtain deterministic variants of the convergence rates when we suppress the randomized block-coordinate scheme. Importantly, this paper appears to be the first work to provide these rate guarantees for this class of problems. (ii) In the case where the CVI set is unbounded and the objective function is smooth and strongly convex, utilizing the properties of the Tikhonov trajectory, we establish the global convergence of aRB-IRG in an almost sure and a mean sense. We provide the numerical experiments for computing the best Nash equilibrium in a networked Cournot competition model.

Key words. first-order methods, variational inequalities, complexity analysis, efficiency of Nash equilibria, iterative regularization, randomized block-coordinate, bilevel optimization

AMS subject classifications. 65K15, 49J40, 90C33, 91A10, 90C06

DOI. 10.1137/20M1357378

1. Introduction.

Traditionally, the mathematical models and algorithms for constrained optimization have been much focused on the cases where the functional constraints are in the form of inequalities, equalities, or easy-to-project sets. However, in the breadth of emerging applications in control theory and economics, the system constraints are too complex to be characterized in those forms. This may arise in several network application domains where the optimization model is complicated by the presence of equilibrium constraints, complementarity constraints, or an inner-level large scale optimization problem. Accordingly, the goal in this paper lies in addressing the following constrained optimization problem

$$(\mathrm{P}^f_{\mathrm{VI}}) \qquad \begin{array}{ll} \text{minimize} & f(x) \\ \text{subject to} & x \in \mathrm{SOL}(X, F). \end{array}$$

Here, $f : \mathbb{R}^n \to \mathbb{R}$ is a convex function and $X \subseteq \mathbb{R}^n$ is given as a Cartesian product, i.e., $X \triangleq \prod_{i=1}^d X_i$, where $X_i \subseteq \mathbb{R}^{n_i}$ is convex for all $i = 1, \ldots, d$ and $\sum_{i=1}^d n_i = n$. We consider $F : X \to \mathbb{R}^n$ to be a monotone mapping. The term $\mathrm{SOL}(X, F)$

*Received by the editors August 3, 2020; accepted for publication (in revised form) June 1, 2021; published electronically August 31, 2021. A very preliminary version of this work appeared in *Proceedings of the American Control Conference*, 2019, pp. 3420–3425, https://doi.org/10.23919/ACC.2019.8815256.
https://doi.org/10.1137/20M1357378
Funding: The second author received support from the NSF through CAREER grant ECCS-1944500.
†School of Industrial Engineering & Management, Oklahoma State University, Stillwater, OK 74078 USA (harshal.kaushik@okstate.edu, farzad.yousefian@okstate.edu).

denotes the solution set of the variational inequality $\mathrm{VI}(X, F)$ defined as follows. A vector $x \in X$ is said to be a solution to $\mathrm{VI}(X, F)$ if for any $y \in X$, we have $F(x)^T(y - x) \geq 0$. Variational inequalities (VIs), first introduced in late 1950s, are an immensely powerful mathematical tool that can serve as a unifying framework for capturing a wide range of applications arising in operations research, finance, and economics (cf. [13, 39, 11, 49]). Importantly, as will be discussed shortly, the block structure of the set X allows for addressing Nash games as well as high-dimensional optimization problems. We note that the problem ($\mathrm{P}^f_{\mathrm{VI}}$) can represent a variety of standard problems in optimization and VI regimes. For example, when $F(x) := 0_n$, the problem ($\mathrm{P}^f_{\mathrm{VI}}$) is equivalent to the canonical optimization problem $\min_{x\in X} f(x)$. Also, when $f(x) := 0$, the problem ($\mathrm{P}^f_{\mathrm{VI}}$) is equivalent to solving $\mathrm{VI}(X, F)$. More detailed examples that can be reformulated as the problem ($\mathrm{P}^f_{\mathrm{VI}}$) are presented below.

1.1. Motivating examples.

Example 1.1 (efficiency estimation of equilibria). The main motivating application arises from the notion of *efficiency of equilibria* in multiagent networks, including communication networks and power systems. In the noncooperative regimes, the system behavior is governed by a collection of decisions (i.e., equilibrium) made by a set of independent and self-interested agents. As a result of this noncooperative behavior (i.e., *game*) among the agents, the global performance of the system may become worse than the case where the agents cooperatively seek an *optimal* decision. A well-known example is the prisoners' dilemma, where the costs of the players incurred by the Nash equilibrium are superior to their costs when they cooperate [36]. Indeed, it has been well-received in economics and computer science communities that Nash equilibria of a game may not attain full efficiency. This perception has led to a surge of research for understanding the quality of an equilibrium in noncooperative games. In particular, addressing this question becomes imperative for network design [4] in the areas of routing [10] and load balancing [40]. In such networks, a protocol designer seeks *the best equilibrium* with respect to a global performance measure, i.e., the function f in ($\mathrm{P}^f_{\mathrm{VI}}$). To this end, the notion of *price of stability* is defined as the ratio of the best objective function value over the set of equilibria to the best objective function value under no competition [35]. In regard to the choice of the objective function f in ($\mathrm{P}^f_{\mathrm{VI}}$), different approaches have been considered, including the *utilitarian* function and the *egalitarian* function [35]. In the utilitarian approach, f is defined as the summation of the individual objective functions of the agents, while in the egalitarian approach, the maximum of the individual cost functions is considered. In the context of network resource allocation where a monetary value is measured, the utilitarian approach is also referred to as Marshallian aggregate surplus (e.g., see [19]). In the following, we describe the details for the problem of selecting the best equilibrium in Nash games, where we employ the utilitarian approach. Consider a canonical Nash game among d players where the ith player is associated with a strategy $x^{(i)} \in X_i \subseteq \mathbb{R}^{n_i}$ and a cost function $g_i(x^{(i)}; x^{(-i)})$, where $x^{(-i)}$ denotes the collection of actions of other players. Nash games arise in a wide range of problems including communication networks [1, 2, 51], cognitive radio networks [48, 44, 28], and power markets [22, 23, 46]. The game is defined formally as the following collection of problems for all $i = 1, \ldots, d$,

$$\mathrm{P}_i\left(x^{(-i)}\right) \qquad \begin{aligned} &\text{minimize}_{x^{(i)}} \quad g_i\left(x^{(i)}; x^{(-i)}\right) \\ &\text{subject to} \quad x^{(i)} \in X_i. \end{aligned}$$

A Nash equilibrium is a tuple of strategies $x^* \triangleq (x^{*(1)}; x^{*(2)}; \ldots; x^{*(d)})$ where no player can obtain a lower cost by deviating from his own strategy, given that the strategies of the other players remain unchanged. It is known that (cf. Proposition 1.4.2 in [13]) when for all i, X_i is a closed convex set and g_i is a differentiable convex function with respect to $x^{(i)}$, the resulting equilibrium conditions of the Nash game given by $(\mathrm{P}_i\left(x^{(-i)}\right))$ are compactly captured by a Cartesian VI(X, F) where $X \triangleq \prod_{i=1}^d X_i$ and $F(x) \triangleq (F_1(x); \ldots; F_d(x))$ with $F_i(x) \triangleq \nabla_{x^{(i)}} g_i(x^{(i)}; x^{(-i)})$. The set SOL$(X, F)$ will then represent the set of Nash equilibria to the game $(\mathrm{P}_i\left(x^{(-i)}\right))$. The best Nash equilibrium problem employing the utilitarian approach is formulated as follows

$$\begin{aligned} &\text{minimize} \quad \sum_{i=1}^d g_i\left(x^{(i)}; x^{(-i)}\right) \\ &\text{subject to} \quad x \in \mathrm{SOL}\left(\prod_{i=1}^d X_i, (\nabla_{x^{(1)}} g_1; \ldots; \nabla_{x^{(d)}} g_d)\right). \end{aligned} \tag{1.1}$$

In section 5, we solve the model (1.1) for a class of networked Nash–Cournot games.

Example 1.2 (high-dimensional constrained convex optimization). Another class of problems that can be captured by the model $(\mathrm{P}^f_{\mathrm{VI}})$ is as follows

$$\begin{aligned} &\text{minimize} \quad f(x) \\ &\text{subject to} \quad h_j(x) \le 0 \quad \text{for all } j = 1, \ldots, J, \\ &\qquad\qquad\quad Ax = b, \quad x \in X \triangleq \prod\nolimits_{i=1}^d X_i, \end{aligned} \tag{1.2}$$

where $x \in \mathbb{R}^n$, $A \in \mathbb{R}^{m \times n}$, $b \in \mathbb{R}^m$, $h_j : \mathbb{R}^n \to \mathbb{R}$ for all j, and n is possibly very large. In the following, we show a case where the problem (1.2) can be cast as $(\mathrm{P}^f_{\mathrm{VI}})$.

LEMMA 1.3. *Let the problem* (1.2) *be feasible and $h_j(x)$ be a continuously differentiable convex function for all j. Let the set $X_i \in \mathbb{R}^{n_i}$ be nonempty, closed, and convex for all i. Then, the problem* (1.2) *is equivalent to* $(\mathrm{P}^f_{\mathrm{VI}})$ *where $F : \mathbb{R}^n \to \mathbb{R}^n$ is defined as $F(x) \triangleq A^T(Ax - b) + \sum_{j=1}^J \max\{0, h_j(x)\}\nabla h_j(x)$.*

Proof. See Appendix A.1. ∎

Example 1.4 (optimization problems with complementarity constraints). Another class of problems that can be addressed in this work is as follows

$$\begin{aligned} &\text{minimize} \quad f(x) \\ &\text{subject to} \quad x^T F(x) = 0, \quad x \ge 0, \quad F(x) \ge 0, \end{aligned} \tag{1.3}$$

where $F : \mathbb{R}^n \to \mathbb{R}^n$ is a mapping. Then, problem (1.3) can be cast as $(\mathrm{P}^f_{\mathrm{VI}})$ where the set X is the nonnegative orthant, i.e., $X \triangleq \mathbb{R}^n_+$ (see Proposition 1.1.3 in [13]).

Example 1.5 (optimization problems with nonlinear equality constraints). Consider the following optimization problem

$$\begin{aligned} &\text{minimize} \quad f(x) \\ &\text{subject to} \quad F(x) = 0, \end{aligned} \tag{1.4}$$

where $F : \mathbb{R}^n \to \mathbb{R}^n$ is a mapping. Defining $X \triangleq \mathbb{R}^n$, SOL$(X, F)$ is equal to the feasible solution set of the problem (1.4). This implies that the problem (1.4) can be captured by the formulation $(\mathrm{P}^f_{\mathrm{VI}})$.

1.2. Existing methods and the research gap. We first begin by providing a brief overview of the solution methods for addressing a VI problem. Starting from the seminal work of Lemke and Howson [30] and Scarf [42], who developed the first solution methods for computing equilibria, in the past few decades there has been a surge of research on the development and analysis of the computational methods for solving VIs. Perhaps this interest lies in the strong interplay between the VIs and the formulation of optimization and equilibrium problems arising in many communication and networking problems [43]. Korpelevich's celebrated extragradient method [27] and its extensions [33, 20, 7, 17, 8, 53, 15, 9, 55] were developed, which require weaker assumptions than their gradient counterparts. In the past decade, there has been a trending interest in addressing VIs in the stochastic regimes. Among these, Jiang and Xu [18] developed the stochastic approximation methods for solving VIs with strongly monotone and smooth mappings. This work was later extended to the case with merely monotone mappings [28, 21, 16] and nonsmooth mappings [54].

Despite many advances in the theory and algorithms for VIs, solving the problem ($\mathrm{P}^f_{\mathrm{VI}}$) has remained challenging. To the best of our knowledge, the computational complexity of the existing solution methods for addressing ($\mathrm{P}^f_{\mathrm{VI}}$) is unknown. In addressing the standard constrained optimization problems, Lagrangian duality and relaxation rules have often proven to be very successful [6]. However, when it comes to solving ($\mathrm{P}^f_{\mathrm{VI}}$), the duality theory cannot be practically employed. This is primarily because unlike in the standard constrained optimization problems where the objective function provides a metric for distinguishing solutions, there is no immediate analogue in the VI problems. Inspired by the contributions of Andrey Tikhonov in 1980s on addressing illposed optimization problems, the existing methods for solving ($\mathrm{P}^f_{\mathrm{VI}}$) share in common a sequential regularization (SR) scheme presented by Algorithm 1.1. The SR scheme is a two-loop framework where at each iteration, given a fixed parameter η_t, a regularized VI denoted by $\mathrm{VI}(X, F+\eta_t\mathbf{I}_n)$ is required to be solved. In the special case where $f(x) := \frac{1}{2}\|x\|^2$, it can be shown when $\eta_t \to 0$, under the monotonicity of the mapping F and closedness and convexity of the set X, any limit point of the *Tikhonov trajectory* denoted by $\{x^*_{\eta_t}\}$, where $x^*_{\eta_t} \in \mathrm{SOL}(X, F+\eta_t\mathbf{I}_n)$, converges to the least ℓ_2-norm vector in $\mathrm{SOL}(X,F)$ (cf. Chapter 12 in [13]). The SR approach is associated with two main drawbacks: (i) It is a computationally inefficient scheme, as it requires solving a series of increasingly more difficult VI problems. (ii) The iteration complexity of the SR scheme in addressing the problem ($\mathrm{P}^f_{\mathrm{VI}}$) is unknown. Accordingly, the main goal in this work lies in the development of an efficient scheme equipped with computational complexity analysis for solving the problem ($\mathrm{P}^f_{\mathrm{VI}}$).

1.3. Summary of contributions. Our main contributions are as follows:
(i) *Development of a single timescale method equipped with convergence rate guarantees*: In addressing ($\mathrm{P}^f_{\mathrm{VI}}$), we develop an efficient first-order method called averaging randomized block iteratively regularized gradient (aRB-IRG). The proposed method

Algorithm 1.1. The existing SR scheme for solving problem ($\mathrm{P}^f_{\mathrm{VI}}$) when $f := \frac{1}{2}\|\cdot\|^2$

```
1: Input: Set X, mapping F, and an initial regularization parameter η_0 > 0.
2: for t = 0, 1, ... do
3:     Compute x*_{η_t} defined as x*_{η_t} ∈ SOL(X, F + η_t I_n).
4:     Update η_t to η_{t+1} such that η_{t+1} < η_t.
5: end for
```

is single timescale in the sense that, unlike the SR approach, it does not require solving a VI at each iteration. Instead, it only uses evaluations of the mapping F and the subgradient of the objective function f at each iteration. In the first part of the paper, we consider the case where the set X is bounded. We let f be a subdifferentiable merely convex function and F be a monotone mapping. In Theorem 3.3, we derive a suboptimality convergence rate in terms of the expected value of the objective function. We also derive a convergence rate for the infeasibility that is characterized by the expected value of a dual gap function. We also derive deterministic variants of the aforementioned convergence rates when we suppress the randomized block-coordinate scheme. In the second part of the paper, we consider the case where the set X is unbounded and f is smooth and strongly convex. Utilizing the properties of the Tikhonov trajectory, we establish the global convergence of the scheme in an almost sure and a mean sense. To the best of our knowledge, this work appears to be the first paper that provides the two rate statements for problems of the form ($\mathrm{P}^f_{\mathrm{VI}}$). In particular, the complexity analysis in this work contributes to the existing convergence theory in several previous papers including [50, 28, 21, 54, 24, 25, 56]. Moreover, in the special case where the VI constraint represents the optimal solution set of an optimization problem, ($\mathrm{P}^f_{\mathrm{VI}}$) captures a class of bilevel optimization problems. This class of problems has been studied in a number of recent papers in deterministic [47, 5, 41, 14], stochastic [3], and distributed [52] regimes. However, the complexity analysis in the aforementioned papers lacks a suboptimality rate, or lacks an infeasibility rate, or requires much stronger assumptions such as strong convexity and smoothness of f.

(ii) *Advancing the convergence rate properties of the randomized block-coordinate schemes*: Block-coordinate schemes, and specifically their randomized variants, have been widely studied in addressing the standard optimization problems (e.g., see [34, 38, 45, 12, 55]). However, in addressing VI problems, there are only a handful of recent papers, including [29, 55], that employ this technique and are equipped with rate guarantees. The aforementioned papers address standard VI problems that can be viewed a special case of the model ($\mathrm{P}^f_{\mathrm{VI}}$) where $f(x) := 0$. In this work, we extend the convergence and rate analysis of the randomized block-coordinate schemes to the much broader regime of optimization problems with CVI constraints.

Outline of the paper. The paper is organized as follows. The proposed algorithm is presented in section 2. The complexity analysis is provided in section 3. In section 4, we provide the convergence analysis when the set X is unbounded. The experimental results are in section 5, and the conclusions follow in section 6.

Notation and preliminary definitions. Throughout, a vector $x \in \mathbb{R}^n$ is assumed to be a column vector and x^T denotes the transpose of x. We use $x^{(i)} \in \mathbb{R}^{n_i}$ to denote the ith block-coordinate of vector x where $x = (x^{(1)}; \ldots ; x^{(d)})$ and $\sum_{i=1}^d n_i = n$. The Euclidean norm of a vector x is denoted by $\|x\|$, i.e., $\|x\| \triangleq \sqrt{x^T x}$. For a mapping $F : \mathbb{R}^n \to \mathbb{R}^n$, we denote the ith block-coordinate of F by $F_i : \mathbb{R}^n \to \mathbb{R}^{n_i}$, i.e., $F(x) = (F_1(x); \ldots ; F_d(x))$. A mapping $F : \mathbb{R}^n \to \mathbb{R}^n$ is said to be monotone on a convex set $X \subseteq \mathbb{R}^n$ if for any $x, y \in X$, we have $(F(x) - F(y))^T(x - y) \geq 0$. The mapping F is said to be μ–strongly monotone on a convex set $X \subseteq \mathbb{R}^n$ if $\mu > 0$ and for any $x, y \in X$, we have $(F(x) - F(y))^T(x - y) \geq \mu\|x - y\|^2$. Also, F is said to be Lipschitz with parameter $L > 0$ on the set X if for any $x, y \in X$, we have $\|F(x)-F(y)\| \leq L\|x-y\|$. A continuously differentiable function $f : \mathbb{R}^n \to \mathbb{R}$ is called μ–strongly convex on a convex set X if $f(x) \geq f(y) + \nabla f(y)^T(x - y) + \frac{\mu}{2}\|x - y\|^2$. Function f is μ–strongly convex if and only if ∇f is μ–strongly monotone on X. For a convex function f with the domain $\mathrm{dom}(f)$, the subgradient of f at $x \in \mathrm{dom}(f)$ is denoted by $\tilde{\nabla} f(x)$ and it satisfies $f(x)+\tilde{\nabla} f(x)^T(y-x) \leq f(y)$ for all $y \in \mathrm{dom}(f)$. The subdifferential set of f at x is the set of all subgradients of f at x and is denoted by $\partial f(x)$. The Euclidean projection of vector s onto a convex set S is denoted by $\mathcal{P}_S(s)$, where $\mathcal{P}_S(s) \triangleq \mathrm{argmin}_{y\in S} \|s-y\|$. We use $\mathbf{I}_n$ to denote the identity matrix of size $n \times n$. The probability of an event Z is denoted by $\mathrm{Prob}(Z)$ and the expectation of a random variable z is denoted by $\mathsf{E}[z]$. We use $\mathbb{R}^n_+$ and $\mathbb{R}^n_{++}$ to denote $\{x \in \mathbb{R}^n \mid x \geq 0\}$ and $\{x \in \mathbb{R}^n \mid x > 0\}$, respectively.

2. Outline of the algorithm. In this section, we state the main assumptions and present the proposed scheme for solving the optimization problem $(\mathrm{P}^f_{\mathrm{VI}})$.

Assumption 2.1. Consider the problem $(\mathrm{P}^f_{\mathrm{VI}})$ under the following conditions:
(a) The set X_i is nonempty, closed, and convex for all $i = 1, \ldots, d$.
(b) The function f is convex and has bounded subgradients over the set X.
(c) The mapping $F : \mathbb{R}^n \to \mathbb{R}^n$ is continuous, monotone, and bounded over the set X.
(d) The optimal solution set of problem $(\mathrm{P}^f_{\mathrm{VI}})$ in nonempty.

Assumption 2.1(b) implies that f is Lipschitz continuous over the set X. Under this assumption, we address a broad class of problems of the form $(\mathrm{P}^f_{\mathrm{VI}})$ where the objective function is possibly nondifferentiable and nonstrongly convex. In the following, we discuss the conditions under which Assumption 2.1(d) is satisfied.

Remark 2.2 (existence of an optimal solution). Suppose Assumptions 2.1(a), (b), and (c) hold. The existence of an optimal solution to the problem $(\mathrm{P}^f_{\mathrm{VI}})$ can be established under different conditions. We provide two instances as follows: (i) Suppose there exists a vector $\bar{x} \in X$ such that the set $\bar{X} \triangleq \{x \in X : \ F(x)^T(x-\bar{x}) \leq 0\}$ is bounded. Then, from Proposition 2.2.3 in [13], $\mathrm{SOL}(X, F)$ is nonempty and compact. Consequently, the Weierstrass theorem implies the existence of an optimal solution to the problem $(\mathrm{P}^f_{\mathrm{VI}})$. (ii) Suppose the set X is compact. Then, from Corollary 2.2.5 in [13], the set $\mathrm{SOL}(X, F)$ is nonempty and compact. Again, Assumption 2.1(d) is guaranteed by the Weierstrass theorem.

Throughout, we let $C_F > 0$ denote the bound on the Euclidean norm of the mapping F, i.e., $\|F(x)\| \leq C_F$ for all $x \in X$. Also, we let $C_f > 0$ denote the bound on the norm of the subgradients of f, i.e., $\|\tilde{\nabla} f(x)\| \leq C_f$ for all $\tilde{\nabla} f(x) \in \partial f(x)$ and $x \in X$. The outline of the proposed method is presented by Algorithm 2.1. At iteration k, a block-coordinate index i_k is selected randomly as follows.

Assumption 2.3 (block-coordinate selection rule). At each iteration $k \geq 0$, the random variable i_k is generated from an independent and identically distributed discrete probability distribution such that $\mathrm{Prob}\,(i_k = i) = \mathsf{p}_i$ where $\mathsf{p}_i > 0$ for $i \in \{1, \ldots, d\}$ and $\sum_{i=1}^d \mathsf{p}_i = 1$.

Then, the i_kth block-coordinate of the iterate x_k is updated using (2.1). Here, γ_k denotes the stepsize at iteration k and η_k denotes the regularization parameter at iteration k. We note that these sequences are updated iteratively. Here, we incorporate the information of the mapping F and the subgradient mapping $\tilde{\nabla} f$ by employing an iterative regularization scheme. At each iteration, a projection operation is performed onto a randomly selected set X_{i_k}. For this reason, the proposed algorithm finds some relevance with the prior study on randomized projection methods such as [32]. We will show that the convergence and rate analysis of the proposed method mainly rely on the choices of $\{\gamma_k\}$ and $\{\eta_k\}$. Accordingly, one key research objective

Algorithm 2.1. aRB-IRG

Input: A random initial point $x_0 \in X$, $\bar{x}_0 := x_0$, initial stepsize $\gamma_0 > 0$, initial regularization parameter $\eta_0 > 0$, a scalar $0 \leq r < 1$, and $S_0 := \gamma_0^r$.

for k = 0, 1, … **do**

  Generate a realization of random variable i_k according to Assumption 2.3.

  Evaluate $F_{i_k}(x_k)$ and $\tilde{\nabla}_{i_k} f(x_k)$ where $\tilde{\nabla} f(x_k) \in \partial f(x_k)$.

  Update x_k as

$$x_{k+1}^{(i)} := \begin{cases} \mathcal{P}_{X_{i_k}}\left(x_k^{(i_k)} - \gamma_k \left(F_{i_k}(x_k) + \eta_k \tilde{\nabla}_{i_k} f(x_k)\right)\right) & \text{if } i = i_k, \\ x_k^{(i)} & \text{if } i \neq i_k. \end{cases} \tag{2.1}$$

  Obtain γ_{k+1} and η_{k+1} (cf. Theorems 3.3 and 4.11 for the update rules).

  Update the averaged iterate $\bar{x}_k$ as

$$S_{k+1} := S_k + \gamma_{k+1}^r, \quad \bar{x}_{k+1} := \frac{S_k \bar{x}_k + \gamma_{k+1}^r x_{k+1}}{S_{k+1}}. \tag{2.2}$$

end for

in this section is to develop suitable update rules for the two sequences so that we can establish the convergence and derive rate statements. To obtain the rate results, we employ an averaging step using the equations given by (2.2), where the sequence $\{\bar{x}_k\}$ is obtained as a weighted average of $\{x_k\}$. The averaging weights are characterized by the stepsize γ_k and a scalar $r \in \mathbb{R}$. Note that in γ_k^r, the scalar r denotes the exponent. It will be shown that the rate results can be provided when $0 \leq r < 1$ (cf. Theorem 3.3).

Remark 2.4. Importantly, unlike Algorithm 1.1, Algorithm 2.1 is a single time-scale scheme that does not require solving any inner-level VI problem. In particular, the update rule given by step (2.1) mainly requires evaluations of random blocks of the mappings F and $\tilde{\nabla} f$. For this reason, step (2.1) is computationally more efficient than step 3 in Algorithm 1.1.

2.1. Preliminaries. In the following, we provide some definitions and preliminary results that will be used to analyze the convergence of Algorithm 2.1.

DEFINITION 2.5 (distance function). *For any $x, y \in \mathbb{R}^n$, function $\mathcal{D}(x,y)$ is defined as $\mathcal{D}(x,y) \triangleq \sum_{i=1}^d \mathsf{p}_i^{-1} \| x^{(i)} - y^{(i)} \|^2$, where p_i is given by Assumption* 2.3.

Remark 2.6. Under Assumption 2.3, we can relate the distance function $\mathcal{D}$ with the ℓ_2-norm as $\mathsf{p}_{min}\mathcal{D}(x,y) \leq \|x-y\|^2 \leq \mathsf{p}_{max}\mathcal{D}(x,y)$ for all $x, y \in \mathbb{R}^n$, where $\mathsf{p}_{min} \triangleq \min_{1 \leq i \leq d}\{\mathsf{p}_i\}$ and $\mathsf{p}_{max} \triangleq \max_{1 \leq i \leq d}\{\mathsf{p}_i\}$.

One of the main challenges in the convergence analysis of computational methods for solving VI problems lies in the lack of access to a standard metric to quantify the quality of the solution iterates. This is in contrast with solving the standard optimization problems where the objective function can serve as an immediate performance metric for the underlying algorithm. Addressing this challenge in the literature of VI problems has led to the study of so-called gap functions (cf. [13, 54]). Of these, in the analysis of this section, we use the dual gap function defined as follows.

Definition 2.7 (the dual gap function [31]). *Let a nonempty closed set $X \subseteq \mathbb{R}^n$ and a mapping $F : X \to \mathbb{R}^n$ be given. Then, for any $x \in X$, the dual gap function $\mathrm{GAP} : X \to \mathbb{R} \cup \{+\infty\}$ is defined as $\mathrm{GAP}(x) \triangleq \sup_{y \in X} F(y)^T(x-y)$.*

Remark 2.8. When $X \neq \emptyset$, Definition 2.7 implies that the dual gap function is nonnegative over X. It is also known that when F is continuous and monotone and the set X is closed and convex, $\mathrm{GAP}(x^*) = 0$ if and only if $x^* \in \mathrm{SOL}(X, F)$ (cf. [20]). Thus, we conclude that under Assumption 2.1, the dual gap function is well-defined.

Definition 2.9 (regularized mapping). *Given a vector $x \in X$, a subgradient $\tilde{\nabla} f(x) \in \partial f(x)$, and an integer $k \geq 0$, the regularized mapping $G_k : X \to \mathbb{R}$ is defined as $G_k(x) \triangleq F(x) + \eta_k \tilde{\nabla} f(x)$. The ith block-coordinate of G_k is denoted by $G_{k,i}$.*

Definition 2.10 (history of the method). *Throughout, we let the history of the algorithm to be denoted by $\mathcal{F}_k \triangleq \{x_0, i_0, i_1, \ldots, i_{k-1}\}$ for $k \geq 1$ with $\mathcal{F}_0 \triangleq \{x_0\}$.*

Next, we show that $\bar{x}_k$ generated by Algorithm 2.1 is a well-defined weighted average.

Lemma 2.11 (weighted averaging). *Let $\{\bar{x}_k\}$ be generated by Algorithm 2.1. Let us define the weights $\lambda_{k,N} \triangleq \frac{\gamma_k^r}{\sum_{j=0}^N \gamma_j^r}$ for $k \in \{0, \ldots, N\}$ and $N \geq 0$. Then, for any $N \geq 0$, we have $\bar{x}_N = \sum_{k=0}^N \lambda_{k,N} x_k$. Also, when X is a convex set, we have $\bar{x}_N \in X$.*

Proof. See Appendix A.2. ∎

In the following, we define two terms that characterize the error between the true maps with their randomized block variants.

Definition 2.12 (randomized block error terms). *Let $\mathbf{U}_i \in \mathbb{R}^{n \times n_i}$ for $i = 1, \ldots, d$ be the collection of matrices such that $\mathbf{I}_n = [\mathbf{U}_1, \ldots, \mathbf{U}_d] \in \mathbb{R}^{n \times n}$. Consider the following definitions for $k \geq 0$:*

$$\Delta_k \triangleq F(x_k) - \mathsf{p}_{i_k}^{-1} \mathbf{U}_{i_k} F_{i_k}(x_k), \qquad \delta_k \triangleq \tilde{\nabla} f(x_k) - \mathsf{p}_{i_k}^{-1} \mathbf{U}_{i_k} \tilde{\nabla}_{i_k} f(x_k). \tag{2.3}$$

Lemma 2.13 (properties of Δ_k and δ_k). *Consider Definition 2.12. We have:*
(a) $\mathsf{E}[\Delta_k \mid \mathcal{F}_k] = \mathsf{E}[\delta_k \mid \mathcal{F}_k] = 0$.
(b) $\mathsf{E}\left[\|\Delta_k\|^2 \mid \mathcal{F}_k\right] \leq \left(\mathsf{p}_{min}^{-1} - 1\right) C_F^2$ *and* $\mathsf{E}\left[\|\delta_k\|^2 \mid \mathcal{F}_k\right] \leq \left(\mathsf{p}_{min}^{-1} - 1\right) C_f^2$.

Proof. See Appendix A.3 ∎

We will use the next result in deriving the suboptimality and infeasibility rate results.

Lemma 2.14 (bounds on the harmonic series). *Let $0 \leq \alpha < 1$ be a given scalar. Then, for any integer $N \geq 2^{\frac{1}{1-\alpha}} - 1$, we have $\frac{(N+1)^{1-\alpha}}{2(1-\alpha)} \leq \sum_{k=0}^N \frac{1}{(k+1)^\alpha} \leq \frac{(N+1)^{1-\alpha}}{1-\alpha}$.*

Proof. See Appendix A.4 ∎

3. Convergence rate analysis. In the following result, we derive an inequality that will be later used to construct bounds on the objective function value and the dual gap function at the averaged sequence generated by Algorithm 2.1.

Lemma 3.1. *Consider the sequence $\{x_k\}$ in Algorithm 2.1. Suppose $\{\gamma_k\}$ and $\{\eta_k\}$ are strictly positive sequences. Let Assumptions 2.1 and 2.3 hold. Let the auxiliary sequence $\{u_k\} \subset X$ be defined as $u_{k+1} \triangleq \mathcal{P}_X(u_k - \gamma_k(\Delta_k + \eta_k \delta_k))$, where $u_0 \in X$ is an arbitrary vector. Then, for all $y \in X$ and all $k \geq 0$, we have*

$$\begin{aligned}\gamma_k^r F(y)^T(x_k-y)+\gamma_k^r\eta_k\tilde{\nabla}f(x_k)^T(x_k-y) &\le \frac{\gamma_k^{r-1}}{2}\left(\mathcal{D}\left(x_k,y\right)+\|u_k-y\|^2\right)\\ &-\frac{\gamma_k^{r-1}}{2}\left(\mathcal{D}\left(x_{k+1},y\right)+\|u_{k+1}-y\|^2\right)+\gamma_k^r(x_k-u_k)^T\left(\Delta_k+\eta_k\delta_k\right)\\ &+\gamma_k^{r+1}\left(\|\Delta_k\|^2+\eta_k^2\|\delta_k\|^2\right)+0.5\mathsf{p}_{i_k}^{-1}\gamma_k^{r+1}\left\|G_{k,i_k}\left(x_k\right)\right\|^2.\end{aligned}\tag{3.1}$$

Proof. Let $k\ge 1$ be fixed. From Definition 2.5 and (2.1), for any $y\in X$, we have

$$\mathcal{D}\left(x_{k+1},y\right)=\mathsf{p}_{i_k}^{-1}\left\|x_{k+1}^{(i_k)}-y^{(i_k)}\right\|^2+\sum_{i=1,\,i\neq i_k}^{d}\mathsf{p}_i^{-1}\left\|x_k^{(i)}-y^{(i)}\right\|^2.\tag{3.2}$$

Next, we find a bound on the term $\|x_{k+1}^{(i_k)}-y^{(i_k)}\|^2$. From the block structure of X, we have $y^{(i_k)}\in X_{i_k}$. Invoking the nonexpansiveness property of the projection mapping, the update rule (2.1), Definition 2.9, and the preceding relation, we obtain

$$\left\|x_{k+1}^{(i_k)}-y^{(i_k)}\right\|^2\le\left\|x_k^{(i_k)}-\gamma_k G_{k,i_k}\left(x_k\right)-y^{(i_k)}\right\|^2.$$

Combining the preceding relation with (3.2), we obtain

$$\begin{aligned}\mathcal{D}\left(x_{k+1},y\right)\le&\sum_{i=1,\,i\neq i_k}^{d}\mathsf{p}_i^{-1}\left\|x_k^{(i)}-y^{(i)}\right\|^2+\mathsf{p}_{i_k}^{-1}\left\|x_k^{(i_k)}-y^{(i_k)}\right\|^2\\ &-2\,\mathsf{p}_{i_k}^{-1}\gamma_k\left(x_k^{(i_k)}-y^{(i_k)}\right)^T G_{k,i_k}\left(x_k\right)+\mathsf{p}_{i_k}^{-1}\gamma_k^2\left\|G_{k,i_k}\left(x_k\right)\right\|^2.\end{aligned}$$

Invoking Definition 2.5 again, we obtain

$$\mathcal{D}\left(x_{k+1},y\right)\le\mathcal{D}\left(x_k,y\right)-2\,\mathsf{p}_{i_k}^{-1}\gamma_k\left(x_k^{(i_k)}-y^{(i_k)}\right)^T G_{k,i_k}\left(x_k\right)+\mathsf{p}_{i_k}^{-1}\gamma_k^2\left\|G_{k,i_k}\left(x_k\right)\right\|^2.\tag{3.3}$$

From Definitions 2.12 and 2.9, we can write

$$\begin{aligned}&\mathsf{p}_{i_k}^{-1}\left(x_k^{(i_k)}-y^{(i_k)}\right)^T G_{k,i_k}\left(x_k\right)=\mathsf{p}_{i_k}^{-1}(x_k-y)^T\left(\mathbf{U}_{i_k}G_{k,i_k}\left(x_k\right)\right)\\ &\quad=\mathsf{p}_{i_k}^{-1}(x_k-y)^T\left(\mathbf{U}_{i_k}F_{i_k}(x_k)+\eta_k\mathbf{U}_{i_k}\tilde{\nabla}_{i_k}f\left(x_k\right)\right)\\ &\quad=(x_k-y)^T\left(F(x_k)-\Delta_k+\eta_k\tilde{\nabla}f(x_k)-\eta_k\delta_k\right)=(x_k-y)^T\left(G_k(x_k)-\Delta_k-\eta_k\delta_k\right).\end{aligned}$$

Combining the preceding inequality and relation (3.3), we obtain

$$\mathcal{D}\left(x_{k+1},y\right)\le\mathcal{D}\left(x_k,y\right)-2\gamma_k(x_k-y)^T\left(G_k(x_k)-\Delta_k-\eta_k\delta_k\right)+\mathsf{p}_{i_k}^{-1}\gamma_k^2\left\|G_{k,i_k}\left(x_k\right)\right\|^2.\tag{3.4}$$

Consider the definition of the auxiliary sequence $\{u_k\}$ in Lemma 3.1. Invoking the nonexpansiveness property of the projection again, we can obtain

$$\begin{aligned}\|u_{k+1}-y\|^2&\le\|u_k-\gamma_k\left(\Delta_k+\eta_k\delta_k\right)-y\|^2\\ &=\|u_k-y\|^2-2\gamma_k(u_k-y)^T\left(\Delta_k+\eta_k\delta_k\right)+\gamma_k^2\|\Delta_k+\eta_k\delta_k\|^2.\end{aligned}$$

Thus, we have

$$\|u_{k+1}-y\|^2 \leq \|u_k-y\|^2 - 2\gamma_k(u_k-y)^T(\Delta_k+\eta_k\delta_k) + 2\gamma_k^2\|\Delta_k\|^2 + 2\gamma_k^2\eta_k^2\|\delta_k\|^2.$$

Adding the preceding inequality and the inequality (3.4) together, we obtain

$$\begin{aligned} 2\gamma_k(x_k-y)^T G_k(x_k) \leq & \left(\mathcal{D}(x_k,y)+\|u_k-y\|^2\right) - \left(\mathcal{D}(x_{k+1},y)+\|u_{k+1}-y\|^2\right) \\ & + 2\gamma_k(x_k-u_k)^T(\Delta_k+\eta_k\delta_k) + 2\gamma_k^2\left(\|\Delta_k\|^2+\eta_k^2\|\delta_k\|^2\right) \\ & + \mathsf{p}_{i_k}^{-1}\gamma_k^2 \left\| G_{k,i_k}(x_k)\right\|^2 . \end{aligned} \tag{3.5}$$

From the monotonicity property of the mapping F and Definition 2.9, we have

$$(x_k-y)^T G_k(x_k) \geq (x_k-y)^T F(y) + \eta_k \tilde{\nabla} f(x_k)^T (x_k-y).$$

This provides a lower bound on the left-hand side of (3.5). The inequality (3.1) is obtained by substituting this bound in (3.5) and multiplying both sides by $\frac{\gamma_k^{r-1}}{2}$, where $r-1$ denotes the exponent in γ_k^{r-1}. ∎

In the following, we develop upper bounds for suboptimality and infeasibility of the weighted average iterate generated by Algorithm 2.1. Both of these error bounds are characterized in terms of the stepsize and the regularization parameter.

Proposition 3.2 (error bounds for Algorithm 2.1). *Let the sequence $\{\bar{x}_k\}$ be generated by Algorithm 2.1, where $0 \leq r < 1$. Suppose $\{\gamma_k\}$ and $\{\eta_k\}$ are strictly positive and nonincreasing sequences. Let Assumptions 2.1 and 2.3 hold and assume that the set X is bounded, i.e., $\|x\| \leq M$ for all $x \in X$ and some $M > 0$.*

(a) *Let x^* be an optimal solution to problem ($\mathrm{P}^f_{\mathrm{VI}}$). Then, for all $N \geq 1$,*

$$\mathsf{E}[f(\bar{x}_N)] - f(x^*) \leq \frac{\frac{4M^2\gamma_N^{r-1}}{\eta_N} + \sum_{k=0}^N \eta_k^{-1}\gamma_k^{r+1}\left(C_F^2+\eta_k^2 C_f^2\right)}{\mathsf{p}_{min}\sum_{k=0}^N \gamma_k^r}. \tag{3.6}$$

(b) *Consider the dual gap function in Definition 2.7. Then, for all $N \geq 1$,*

$$\mathsf{E}[\mathrm{GAP}(\bar{x}_N)] \leq \frac{4M^2\gamma_N^{r-1} + \sum_{k=0}^N \gamma_k^r\left(2\mathsf{p}_{min}\eta_k C_f M + \gamma_k C_F^2 + \gamma_k\eta_k^2 C_f^2\right)}{\mathsf{p}_{min}\sum_{k=0}^N \gamma_k^r}. \tag{3.7}$$

Proof. We define the following terms for all $k \geq 0$, which appear in (3.1):

$$\begin{aligned} &\Theta_{k,1} \triangleq \gamma_k^r(x_k-u_k)^T(\Delta_k+\eta_k\delta_k), \quad \Theta_{k,2} \triangleq \gamma_k^{r+1}\left(\|\Delta_k\|^2+\eta_k^2\|\delta_k\|^2\right), \\ &\Theta_{k,3} \triangleq 0.5\mathsf{p}_{i_k}^{-1}\gamma_k^{r+1}\left\| G_{k,i_k}(x_k)\right\|^2 . \end{aligned} \tag{3.8}$$

Next, we estimate the expected values of these terms. Consider the notation of $\mathcal{F}_k$ given by Definition 2.10. Note that x_k is $\mathcal{F}_k$-measurable. Also, from the definition of u_k in Lemma 3.1, u_k is $\mathcal{F}_k$-measurable. Note, however, that $\Theta_{k,j}$ is $\mathcal{F}_{k+1}$-measurable for all $j \in \{1,2,3\}$. Taking these into account and using the total probability law, for any $k \geq 0$ and $j \in \{1,2,3\}$, we have $\mathsf{E}[\Theta_{k,j}] = \mathsf{E}_{\mathcal{F}_k}[\mathsf{E}_{i_k}[\Theta_{k,j} \mid \mathcal{F}_k]]$. From this relation and Lemma 2.13, we have for any $k \geq 0$

$$\mathsf{E}[\Theta_{k,1}] = 0, \qquad \mathsf{E}[\Theta_{k,2}] = \left(\mathsf{p}_{min}^{-1}-1\right)\gamma_k^{r+1}\left(C_F^2+\eta_k^2 C_f^2\right). \tag{3.9}$$

Also, using Definition 2.9 and the triangle inequality, we can write

$$\mathsf{E}_{i_k}\left[\Theta_{k,3} \mid \mathcal{F}_k\right] = \sum_{i=1}^{d} \mathsf{p}_i \left(0.5\mathsf{p}_i^{-1}\gamma_k^{r+1} \left\| G_{k,i}\left(x_k\right)\right\|^2\right)$$
$$\leq \gamma_k^{r+1} \sum_{i=1}^{d} \left(\|F_i(x_k)\|^2 + \eta_k^2\|\tilde{\nabla}_i f(x_k)\|^2\right) = \gamma_k^{r+1}\|F(x_k)\|^2 + \eta_k^2\|\tilde{\nabla} f(x_k)\|^2.$$

From the preceding inequality, we obtain

$$\mathsf{E}[\Theta_{k,3}] \leq \gamma_k^{r+1}\left(C_F^2 + \eta_k^2 C_f^2\right). \tag{3.10}$$

We are now ready to show the inequalities (3.6) and (3.7) as follows:

(a) Consider (3.1). From the definition of subgradients of the convex function f, we have that $f(x_k) - f(y) \leq \tilde{\nabla} f(x_k)^T(x_k - y)$. Thus, from (3.8) we obtain for any $y \in X$

$$\begin{aligned}\gamma_k^r F(y)^T(x_k - y) &+ \gamma_k^r\eta_k\left(f(x_k) - f(y)\right)\\ &\leq \frac{\gamma_k^{r-1}}{2}\left(\mathcal{D}\left(x_k, y\right) + \|u_k - y\|^2\right)\\ &\quad - \frac{\gamma_k^{r-1}}{2}\left(\mathcal{D}\left(x_{k+1}, y\right) + \|u_{k+1} - y\|^2\right) + \Theta_{k,1} + \Theta_{k,2} + \Theta_{k,3}.\end{aligned}$$

Let us substitute $y := x^*$, where x^* denotes an optimal solution to $(\mathrm{P}_{\mathrm{VI}}^f)$. Note that x^* must be a feasible solution to $(\mathrm{P}_{\mathrm{VI}}^f)$, i.e., $F(x^*)^T(x_k - x^*) \geq 0$. Thus, we obtain

$$\begin{aligned}\gamma_k^r\eta_k\left(f(x_k) - f(x^*)\right) &\leq \frac{\gamma_k^{r-1}}{2}\left(\mathcal{D}\left(x_k, x^*\right) + \|u_k - x^*\|^2\right)\\ &\quad - \frac{\gamma_k^{r-1}}{2}\left(\mathcal{D}\left(x_{k+1}, x^*\right) + \|u_{k+1} - x^*\|^2\right) + \Theta_{k,1} + \Theta_{k,2} + \Theta_{k,3}.\end{aligned} \tag{3.11}$$

Dividing both sides by η_k and adding and subtracting $\frac{\gamma_{k-1}^{r-1}}{2\eta_{k-1}}(\mathcal{D}(x_k, x^*) + \|u_k - x^*\|^2)$ in the right-hand side of (3.11), we obtain for $k \geq 1$

$$\begin{aligned}\gamma_k^r\left(f(x_k) - f(x^*)\right) &\leq \frac{\gamma_{k-1}^{r-1}}{2\eta_{k-1}}\left(\mathcal{D}\left(x_k, x^*\right) + \|u_k - x^*\|^2\right)\\ &\quad - \frac{\gamma_k^{r-1}}{2\eta_k}\left(\mathcal{D}\left(x_{k+1}, x^*\right) + \|u_{k+1} - x^*\|^2\right)\\ &\quad + \frac{1}{2}\left(\frac{\gamma_k^{r-1}}{\eta_k} - \frac{\gamma_{k-1}^{r-1}}{\eta_{k-1}}\right)\left(\mathcal{D}\left(x_k, x^*\right) + \|u_k - x^*\|^2\right)\\ &\quad + \eta_k^{-1}\left(\Theta_{k,1} + \Theta_{k,2} + \Theta_{k,3}\right).\end{aligned} \tag{3.12}$$

Since $r - 1 < 0$ and that $\{\gamma_k\}$ and $\{\eta_k\}$ are nonincreasing, we have $\frac{\gamma_k^{r-1}}{\eta_k} - \frac{\gamma_{k-1}^{r-1}}{\eta_{k-1}} \geq 0$. Also, from the boundedness of the set X, since x_k, x^*, and u_k belong to X, using Remark 2.6 and the triangle inequality, we have

(3.13)
$$\mathcal{D}\left(x_k, x^*\right) + \|u_k - x^*\|^2 \leq \mathsf{p}_{min}^{-1}\left\|x_k - x^*\right\|^2 + \|u_k - x^*\|^2 \leq 4M^2\left(\mathsf{p}_{min}^{-1} + 1\right) \leq \frac{8M^2}{\mathsf{p}_{min}}.$$

Summing over (3.12) from $k = 1$ to N and using (3.13), we obtain

$$\sum_{k=1}^{N} \gamma_k^r \left(f(x_k) - f(x^*)\right) \leq \frac{\gamma_0^{r-1}}{2\eta_0}\left(\mathcal{D}\left(x_1, x^*\right) + \|u_1 - x^*\|^2\right) + 4M^2 \mathsf{p}_{min}^{-1}\left(\frac{\gamma_N^{r-1}}{\eta_N} - \frac{\gamma_0^{r-1}}{\eta_0}\right) + \sum_{k=1}^{N} \eta_k^{-1}\left(\Theta_{k,1} + \Theta_{k,2} + \Theta_{k,3}\right),$$

where we drop the nonpositive term. From relation (3.11) when $k = 0$, we have

$$\gamma_0^r \left(f(x_0) - f(x^*)\right) \leq \frac{\gamma_0^{r-1}}{2\eta_0}\left(\mathcal{D}\left(x_0, x^*\right) + \|u_0 - x^*\|^2\right) - \frac{\gamma_0^{r-1}}{2\eta_0}\left(\mathcal{D}\left(x_1, x^*\right) + \|u_1 - x^*\|^2\right) + \eta_0^{-1}\left(\Theta_{0,1} + \Theta_{0,2} + \Theta_{0,3}\right).$$

Adding the last two inequalities, multiplying and dividing the left-hand side by $\sum_{k=0}^{N} \gamma_k^r$, and then invoking Lemma 2.11 and convexity of f, we obtain

$$\left(\sum_{k=0}^{N} \gamma_k^r\right)\left(f\left(\bar{x}_N\right) - f(x^*)\right) \leq \frac{\gamma_0^{r-1}}{2\eta_0}\left(\mathcal{D}\left(x_0, x^*\right) + \|u_0 - x^*\|^2\right) + 4M^2 \mathsf{p}_{min}^{-1}\left(\frac{\gamma_N^{r-1}}{\eta_N} - \frac{\gamma_0^{r-1}}{\eta_0}\right) + \sum_{k=0}^{N} \eta_k^{-1}\left(\Theta_{k,1} + \Theta_{k,2} + \Theta_{k,3}\right).$$

Taking the expectation on both sides and invoking (3.13), we obtain

$$\mathsf{E}[f\left(\bar{x}_N\right)] - f(x^*) \leq \frac{\frac{4M^2 \mathsf{p}_{min}^{-1}\gamma_N^{r-1}}{\eta_N} + \sum_{k=0}^{N} \eta_k^{-1}\mathsf{E}[\Theta_{k,1} + \Theta_{k,2} + \Theta_{k,3}]}{\sum_{k=0}^{N} \gamma_k^r}.$$

From the relations (3.9) and (3.10), we obtain

$$\mathsf{E}[f\left(\bar{x}_N\right)] - f(x^*) \leq \frac{\frac{4M^2 \mathsf{p}_{min}^{-1}\gamma_N^{r-1}}{\eta_N} + \sum_{k=0}^{N} \eta_k^{-1}\left(\mathsf{p}_{min}^{-1}\gamma_k^{r+1}\left(C_F^2 + \eta_k^2 C_f^2\right)\right)}{\sum_{k=0}^{N} \gamma_k^r},$$

which implies the inequality (3.6).

(b) From the Cauchy–Schwarz inequality, the definitions of C_f and M, and the triangle inequality, we have $\tilde{\nabla} f(x_k)^T(y - x_k) \leq \|\tilde{\nabla} f(x_k)\|\|x_k - y\| \leq 2C_f M$. Adding the preceding inequality with the relation (3.1), from (3.8) we obtain

$$\gamma_k^r F(y)^T(x_k - y) \leq \frac{\gamma_k^{r-1}}{2}\left(\mathcal{D}\left(x_k, y\right) + \|u_k - y\|^2\right) - \frac{\gamma_k^{r-1}}{2}\left(\mathcal{D}\left(x_{k+1}, y\right) + \|u_{k+1} - y\|^2\right) + 2\gamma_k^r \eta_k C_f M + \Theta_{k,1} + \Theta_{k,2} + \Theta_{k,3}. \tag{3.14}$$

Adding and subtracting the term $\frac{\gamma_{k-1}^{r-1}}{2}(\mathcal{D}(x_k, y) + \|u_k - y\|^2)$, we obtain

$$\gamma_k^r F(y)^T(x_k - y) \leq \frac{\gamma_{k-1}^{r-1}}{2}\left(\mathcal{D}\left(x_k, y\right) + \|u_k - y\|^2\right) - \frac{\gamma_k^{r-1}}{2}\left(\mathcal{D}\left(x_{k+1}, y\right) + \|u_{k+1} - y\|^2\right) + \frac{1}{2}\left(\gamma_k^{r-1} - \gamma_{k-1}^{r-1}\right)\left(\mathcal{D}\left(x_k, y\right) + \|u_k - y\|^2\right) + 2\gamma_k^r \eta_k C_f M + \Theta_{k,1} + \Theta_{k,2} + \Theta_{k,3}.$$

Substituting the bound given by (3.13) in the preceding relation, we obtain

$$\gamma_k^r F(y)^T(x_k - y) \leq \frac{\gamma_{k-1}^{r-1}}{2}\left(\mathcal{D}(x_k, y) + \|u_k - y\|^2\right) - \frac{\gamma_k^{r-1}}{2}\left(\mathcal{D}(x_{k+1}, y) + \|u_{k+1} - y\|^2\right) + 4M^2\mathsf{p}_{min}^{-1}\left(\gamma_k^{r-1} - \gamma_{k-1}^{r-1}\right) + 2\gamma_k^r\eta_k C_f M + \Theta_{k,1} + \Theta_{k,2} + \Theta_{k,3}.$$

Summing both sides from $k = 1$ to N, we obtain

$$\sum_{k=1}^{N}\gamma_k^r F(y)^T(x_k - y) \leq \frac{\gamma_0^{r-1}}{2}\left(\mathcal{D}(x_1, y) + \|u_1 - y\|^2\right) + 4M^2\mathsf{p}_{min}^{-1}\left(\gamma_N^{r-1} - \gamma_0^{r-1}\right) + \sum_{k=1}^{N}\left(2\gamma_k^r\eta_k C_f M + \Theta_{k,1} + \Theta_{k,2} + \Theta_{k,3}\right). \tag{3.15}$$

Writing the inequality (3.14) for $k = 0$, we have

$$\gamma_0^r F(y)^T(x_0 - y) \leq \frac{\gamma_0^{r-1}}{2}\left(\mathcal{D}(x_0, y) + \|u_0 - y\|^2\right) - \frac{\gamma_0^{r-1}}{2}\left(\mathcal{D}(x_1, y) + \|u_1 - y\|^2\right) + 2\gamma_0^r\eta_0 C_f M + \Theta_{0,1} + \Theta_{0,2} + \Theta_{0,3}. \tag{3.16}$$

Adding (3.15) and (3.16) together, we obtain

$$\sum_{k=0}^{N}\gamma_k^r F(y)^T(x_k - y) \leq \frac{\gamma_0^{r-1}}{2}\left(\mathcal{D}(x_0, y) + \|u_0 - y\|^2\right) + 4M^2\mathsf{p}_{min}^{-1}\left(\gamma_N^{r-1} - \gamma_0^{r-1}\right) + \sum_{k=0}^{N}\left(2\gamma_k^r\eta_k C_f M + \Theta_{k,1} + \Theta_{k,2} + \Theta_{k,3}\right). \tag{3.17}$$

Recalling $\bar{x}_N = \sum_{k=0}^{N}\lambda_{k,N}x_k$ in Lemma 2.11, applying the bound given by (3.13), and using the triangle inequality, we obtain

$$\left(\sum_{k=0}^{N}\gamma_k^r\right)F(y)^T(\bar{x}_N - y) \leq 4M^2\mathsf{p}_{min}^{-1}\gamma_N^{r-1} + \sum_{k=0}^{N}\left(2\gamma_k^r\eta_k C_f M + \Theta_{k,1} + \Theta_{k,2} + \Theta_{k,3}\right).$$

Taking the supremum with respect to y over the set X from the left-hand side, invoking Definition 2.7, and then dividing both sides by $\sum_{k=0}^{N}\gamma_k^r$, we obtain

$$\mathrm{GAP}(\bar{x}_N) \leq \frac{4M^2\mathsf{p}_{min}^{-1}\gamma_N^{r-1} + \sum_{k=0}^{N}\left(2\gamma_k^r\eta_k C_f M + \Theta_{k,1} + \Theta_{k,2} + \Theta_{k,3}\right)}{\sum_{k=0}^{N}\gamma_k^r}.$$

Taking the expectation on both sides, using the relations (3.9) and (3.10), and rearranging the terms, we obtain the inequality (3.7). ◻

We are now ready to present the convergence rate results of the proposed method.

Theorem 3.3 (convergence rate statements for Algorithm 2.1). *Consider Algorithm* 2.1*. Let Assumptions* 2.1 *and* 2.3 *hold and assume that the set X is bounded such that $\|x\| \leq M$ for all $x \in X$ and some $M > 0$. Suppose for all $k \geq 0$, $\gamma_k := \frac{\gamma_0}{\sqrt{k+1}}$ and $\eta_k := \frac{\eta_0}{(k+1)^b}$, where $\gamma_0 > 0$, $\eta_0 > 0$, and $0 < b < 0.5$. Then, for any $0 \leq r < 1$, the following results hold:*

(i) *Let x^* be an optimal solution to the problem* $(\mathrm{P}^f_{\mathrm{VI}})$. *Then, for all* $N \geq 2^{\frac{2}{1-r}} - 1$,

(3.18)
$$\mathsf{E}[f(\bar{x}_N)] - f(x^*) \leq \frac{2-r}{\mathsf{p}_{min}\eta_0}\left(\frac{4M^2}{\gamma_0} + \frac{\gamma_0\left(C_F^2 + \eta_0^2 C_f^2\right)}{0.5 - 0.5r + b}\right)\frac{1}{(N+1)^{0.5-b}}.$$

(ii) *Consider the dual gap function in Definition* 2.7. *Then, for all* $N \geq 2^{\frac{2}{1-r}} - 1$,

(3.19)
$$\mathsf{E}[\mathrm{GAP}(\bar{x}_N)] \leq \frac{2-r}{\mathsf{p}_{min}}\left(\frac{4M^2}{\gamma_0} + \frac{\gamma_0\left(C_F^2 + \eta_0^2 C_f^2\right)}{0.5 - 0.5r} + \frac{2\mathsf{p}_{min} C_f M \eta_0}{1 - 0.5r - b}\right)\frac{1}{(N+1)^b}.$$

Proof. Let us define the following terms:

$$\Lambda_{N,1} \triangleq \mathsf{p}_{min}\sum_{k=0}^{N}\gamma_k^r, \quad \Lambda_{N,2} \triangleq \frac{4M^2\gamma_N^{r-1}}{\eta_N}, \quad \Lambda_{N,3} \triangleq \left(C_F^2 + \eta_0^2 C_f^2\right)\sum_{k=0}^{N}\eta_k^{-1}\gamma_k^{r+1},$$

$$\Lambda_{N,4} \triangleq 4M^2\gamma_N^{r-1}, \quad \Lambda_{N,5} \triangleq \left(C_F^2 + \eta_0^2 C_f^2\right)\sum_{k=0}^{N}\gamma_k^{r+1}, \quad \Lambda_{N,6} \triangleq 2\mathsf{p}_{min} C_f M \sum_{k=0}^{N}\eta_k\gamma_k^r.$$

Note that from (3.6) and (3.7), we have

$$\mathsf{E}[f(\bar{x}_N)] - f(x^*) \leq \frac{\Lambda_{N,2} + \Lambda_{N,3}}{\Lambda_{N,1}}, \quad \mathsf{E}[\mathrm{GAP}(\bar{x}_N)] \leq \frac{\Lambda_{N,4} + \Lambda_{N,5} + \Lambda_{N,6}}{\Lambda_{N,1}}. \tag{3.20}$$

Next, we apply Lemma 2.14 to estimate the terms $\Lambda_{N,i}$. Substituting γ_k and η_k by their update rules, we obtain

$$\Lambda_{N,1} = \mathsf{p}_{min}\sum_{k=0}^{N}\frac{\gamma_0^r}{(k+1)^{0.5r}} \geq \frac{\mathsf{p}_{min}\gamma_0^r (N+1)^{1-0.5r}}{2(1-0.5r)},$$

$$\Lambda_{N,2} = \frac{4M^2(N+1)^{0.5(1-r)+b}}{\eta_0\gamma_0^{1-r}}, \quad \Lambda_{N,4} = \frac{4M^2(N+1)^{0.5(1-r)}}{\gamma_0^{1-r}},$$

$$\Lambda_{N,3} = \sum_{k=0}^{N}\frac{\left(C_F^2 + \eta_0^2 C_f^2\right)\gamma_0^{r+1}}{\eta_0(k+1)^{0.5(r+1)-b}} \leq \frac{\gamma_0^{r+1}\left(C_F^2 + \eta_0^2 C_f^2\right)(N+1)^{1-0.5(r+1)+b}}{\eta_0(1 - 0.5(r+1) + b)},$$

$$\Lambda_{N,5} = \left(C_F^2 + \eta_0^2 C_f^2\right)\sum_{k=0}^{N}\frac{\gamma_0^{r+1}}{(k+1)^{0.5(r+1)}} \leq \frac{\left(C_F^2 + \eta_0^2 C_f^2\right)\gamma_0^{r+1}(N+1)^{1-0.5(r+1)}}{1 - 0.5(r+1)},$$

$$\Lambda_{N,6} = 2\mathsf{p}_{min} C_f M \eta_0 \gamma_0^r \sum_{k=0}^{N}\frac{1}{(k+1)^{0.5r+b}} \leq \frac{2\mathsf{p}_{min} C_f M \eta_0 \gamma_0^r (N+1)^{1-0.5r-b}}{1 - 0.5r - b}.$$

For these inequalities to hold, we need to ensure that the conditions of Lemma 2.14 are met. Accordingly, we must have $0 \leq 0.5r < 1$, $0 \leq 0.5(r+1) - b < 1$, $0 \leq 0.5r + b < 1$, and $0 \leq 0.5(r+1) < 1$. These relations hold because $0 \leq r < 1$ and $0 < b < 0.5$. Another set of conditions when applying Lemma 2.14 includes $N \geq \max\{2^{1/(1-0.5r)}, 2^{1/(1-0.5(r+1)+b)}, 2^{1/(1-0.5r-b)}, 2^{1/(1-0.5(r+1))}\} - 1$. This relation is indeed satisfied as a consequence of $N \geq 2^{\frac{2}{1-r}} - 1$, $0 < b < 0.5$, and $0 \leq r < 1$. We conclude that all the necessary conditions for applying Lemma 2.14 and obtaining the aforementioned bounds for the terms $\Lambda_{N,i}$ are satisfied. To show that the inequalities (3.18) and (3.19) hold, it suffices to substitute the preceding bounds on the terms $\Lambda_{N,i}$ into the two inequalities given by (3.20). The details are as follows.

$$\mathsf{E}[f(\bar{x}_N)] - f(x^*) \leq \frac{\Lambda_{N,2} + \Lambda_{N,3}}{\Lambda_{N,1}} = \frac{2-r}{\mathsf{p}_{min}\gamma_0^r (N+1)^{1-0.5r}} \left(\frac{4M^2(N+1)^{0.5-0.5r+b}}{\eta_0 \gamma_0^{1-r}} + \left(\frac{\gamma_0^{r+1}}{\eta_0}\right) \frac{\left(C_F^2 + \eta_0^2 C_f^2\right)(N+1)^{0.5-0.5r+b}}{0.5-0.5r+b} \right).$$

The inequality (3.18) is obtained by rearranging the terms in the preceding relation.

$$\mathsf{E}[\mathrm{GAP}(\bar{x}_N)] \leq \frac{\Lambda_{N,4} + \Lambda_{N,5} + \Lambda_{N,6}}{\Lambda_{N,1}} \leq \frac{2-r}{\mathsf{p}_{min}\gamma_0^r (N+1)^{1-0.5r}} \left(\frac{4M^2(N+1)^{0.5-0.5r}}{\gamma_0^{1-r}} + \frac{\left(C_F^2 + \eta_0^2 C_f^2\right)\gamma_0^{r+1}(N+1)^{0.5-0.5r}}{0.5-0.5r} + \frac{2\mathsf{p}_{min} C_f M \eta_0 \gamma_0^r (N+1)^{1-0.5r-b}}{1-0.5r-b} \right).$$

Then, (3.19) can be obtained by rearranging the terms in the preceding inequality. ☐

Remark 3.4 (iteration complexity of Algorithm 2.1). As an immediate result from Theorem 3.3, choosing $\gamma_k := \frac{\gamma_0}{\sqrt{k+1}}$ and $\eta_k := \frac{\eta_0}{\sqrt[4]{k+1}}$, we obtain

$$\mathsf{E}[f(\bar{x}_N) - f(x^*)] = \mathsf{E}[\mathrm{GAP}(\bar{x}_N)] = \mathcal{O}\left(\frac{1}{\sqrt[4]{N}}\right).$$

This implies that Algorithm 2.1 achieves an iteration complexity of $\mathcal{O}\left(\epsilon^{-4}\right)$ in solving ($\mathrm{P}^f_{\mathrm{VI}}$), where $\epsilon > 0$ denotes the expected tolerance in both of the suboptimality and infeasibility metrics.

The rate statements derived in Theorem 3.3 are in a mean sense. In the following, we consider a deterministic variant of Algorithm 2.1 where we suppress the randomized block-coordinate scheme. The outline of this deterministic method is presented by Algorithm 3.1. In Corollary 3.5, we show that nonasymptotic deterministic rate statements can be derived for Algorithm 3.1.

Algorithm 3.1. a-IRG

1: **Input:** An arbitrary initial point $x_0 \in X$, $\bar{x}_0 := x_0$, initial stepsize $\gamma_0 > 0$, initial regularization parameter $\eta_0 > 0$, a scalar $0 \leq r < 1$, and $S_0 := \gamma_0^r$.
2: **for** k = 0, 1, … **do**
3: Evaluate $F(x_k)$ and $\tilde{\nabla} f(x_k)$ where $\tilde{\nabla} f(x_k) \in \partial f(x_k)$.
4: For all $i \in \{1, \ldots, d\}$, do the following update

$$x_{k+1}^{(i)} := \mathcal{P}_{X_i}\left(x_k^{(i)} - \gamma_k \left(F_i(x_k) + \eta_k \tilde{\nabla}_i f(x_k)\right)\right). \tag{3.21}$$

5: Obtain γ_{k+1} and η_{k+1} (cf. Corollary 3.5 for the update rules).
6: Update the averaged iterate $\bar{x}_k$ as

$$S_{k+1} := S_k + \gamma_{k+1}^r, \quad \bar{x}_{k+1} := \frac{S_k \bar{x}_k + \gamma_{k+1}^r x_{k+1}}{S_{k+1}}. \tag{3.22}$$

7: **end for**

COROLLARY 3.5 (convergence rate statements for Algorithm 3.1). *Consider Algorithm 3.1. Let Assumption 2.1 hold and assume that the set X is bounded such that $\|x\| \leq M$ for all $x \in X$ and some $M > 0$. Suppose for $k \geq 0$, $\gamma_k := \frac{\gamma_0}{\sqrt{k+1}}$ and $\eta_k := \frac{\eta_0}{(k+1)^b}$, where $\gamma_0 > 0$, $\eta_0 > 0$, and $0 < b < 0.5$. Then, for any $0 \leq r < 1$, the following results hold:*

(i) *Let x^* be an optimal solution to the problem $(\mathrm{P}^f_{\mathrm{VI}})$. Then, for all $N \geq 2^{\frac{2}{1-r}} - 1$,*

$$f(\bar{x}_N) - f(x^*) \leq \frac{2-r}{\eta_0}\left(\frac{4M^2}{\gamma_0} + \frac{\gamma_0\left(C_F^2 + \eta_0^2 C_f^2\right)}{0.5 - 0.5r + b}\right)\frac{1}{(N+1)^{0.5-b}}. \tag{3.23}$$

(ii) *Consider the dual gap function in Definition 2.7. Then, for all $N \geq 2^{\frac{2}{1-r}} - 1$,*

$$\mathrm{GAP}(\bar{x}_N) \leq (2-r)\left(\frac{4M^2}{\gamma_0} + \frac{\gamma_0\left(C_F^2 + \eta_0^2 C_f^2\right)}{0.5 - 0.5r} + \frac{2C_f M \eta_0}{1 - 0.5r - b}\right)\frac{1}{(N+1)^b}. \tag{3.24}$$

Proof. See Appendix A.6. ∎

4. Addressing the case where X is unbounded. The convergence and rate statements provided by Theorem 3.3 require the set X to be bounded. We, however, note that in some applications, e.g., in the models presented in Examples 1.4 and 1.5, this assumption may not hold. Accordingly, in this section, our aim is to analyze the convergence of Algorithm 2.1 when X is unbounded. To this end, we consider the following main assumption.

Assumption 4.1. Consider problem $(\mathrm{P}^f_{\mathrm{VI}})$ under the following conditions:

(a) The set X_i is nonempty, closed, and convex for all $i = 1, \ldots, d$.
(b) The function f is continuously differentiable and μ_f–strongly convex over X.
(c) The mapping $F : \mathbb{R}^n \to \mathbb{R}^n$ is continuous and monotone over X.
(d) The solution set $\mathrm{SOL}(X, F)$ is nonempty.

Remark 4.2 (existence and uniqueness of the optimal solution). Under Assumption 4.1, the constraint set of $(\mathrm{P}^f_{\mathrm{VI}})$, i.e., $\mathrm{SOL}(X, F)$, is nonempty, closed, and convex. The convexity of this set is implied by Theorem 2.3.5 in [13] and its closedness property is obtained by the continuity of the mapping F and closedness of the set X. Because in the problem $(\mathrm{P}^f_{\mathrm{VI}})$, the objective function f is strongly convex and the constraint set is nonempty, closed, and convex, we conclude from Proposition 1.1.2 in [6] that the problem $(\mathrm{P}^f_{\mathrm{VI}})$ has a unique optimal solution. Throughout this section, we let x^* denote this unique optimal solution.

4.1. Preliminaries. In this part, we provide some preliminary results that will be used in the convergence analysis. We begin by defining a generalized variant of the Tikhonov trajectory that is associated with the problem of interest in this paper.

DEFINITION 4.3 (Tikhonov trajectory). *Consider the problem $(\mathrm{P}^f_{\mathrm{VI}})$ under Assumption 4.1. Let $\{\eta_k\}$ be a sequence of strictly positive scalars for all $k \geq 0$, and let $x^*_{\eta_k} \in X$ denote the unique solution to the regularized VI problem given by $\mathrm{VI}(X, F + \eta_k \nabla f)$. Then, the sequence $\left\{x^*_{\eta_k}\right\}$ is defined as the Tikhonov trajectory associated with the problem $(\mathrm{P}^f_{\mathrm{VI}})$.*

Remark 4.4. The uniqueness of the solution of $\mathrm{VI}(X, F+\eta_k\nabla f)$ in Definition 4.3 is due to the strong monotonicity of the mapping $F+\eta_k\nabla f$ and closedness and convexity of the set X (see Theorem 2.3.3 in [13]). Definition 4.3 generalizes the notion of Tikhonov trajectory provided in [13] in the following way: in [13], $x^*_{\eta_k}$ is defined as the solution to the regularized problem $\mathrm{VI}(X, F+\eta_k\mathbf{I}_n)$. This is indeed the special case where we choose $f(x) := \frac{1}{2}\|x\|^2$ in Definition 4.3.

To analyze the convergence, we utilize the properties of the Tikhonov trajectory. The following result ascertains the asymptotic convergence of this trajectory to the optimal solution of the problem $(\mathrm{P}^f_{\mathrm{VI}})$. It also provides an upper bound on the error between any two successive vectors of the trajectory.

LEMMA 4.5. *Consider Definition* 4.3 *and let Assumption* 4.1 *hold. Let* $\{\eta_k\}$ *be a sequence such that* $\lim_{k\to\infty}\eta_k = 0$ *and* $\eta_k > 0$ *for all* $k \geq 0$*. Then, we have:*

(a) *The Tikhonov trajectory* $\{x^*_{\eta_k}\}$ *converges to a unique limit point, that is,* x^**.*

(b) *There exists* $\bar{C}_f > 0$ *such that* $\|x^*_{\eta_k} - x^*_{\eta_{k-1}}\| \leq \frac{\bar{C}_f}{\mu_f}|1-\frac{\eta_{k-1}}{\eta_k}|$ *for all* $k \geq 1$*.*

Proof. See Appendix A.5. ∎

The following lemmas will be employed to establish the asymptotic convergence result.

LEMMA 4.6 (Theorem 6, page 75 in [26]). *Let* $\{u_t\} \subset \mathbb{R}^n$ *denote a sequence of vectors where* $\lim_{t\to\infty} u_t = \hat{u}$*. Also, let* $\{\alpha_k\}$ *denote a sequence of strictly positive scalars such that* $\sum_{k=0}^{\infty}\alpha_k = \infty$*. Suppose* $v_k \in \mathbb{R}^n$ *is defined by* $v_k \triangleq \frac{\sum_{t=0}^{k}\alpha_t u_t}{\sum_{t=0}^{k}\alpha_t}$ *for all* $k \geq 0$*. Then,* $\lim_{k\to\infty} v_k = \hat{u}$*.*

LEMMA 4.7 (Lemma 10, page 49 in [37]). *Let* $\{v_k\}$ *be a sequence of nonnegative random variables, where* $\mathsf{E}[v_0] < \infty$*, and let* $\{\alpha_k\}$ *and* $\{\beta_k\}$ *be deterministic scalar sequences such that* $\mathsf{E}[v_{k+1}|v_0,\ldots,v_k] \leq (1-\alpha_k)v_k + \beta_k$ *for all* $k \geq 0$*,* $0 \leq \alpha_k \leq 1$*,* $\beta_k \geq 0$*,* $\sum_{k=0}^{\infty}\alpha_k = \infty$*,* $\sum_{k=0}^{\infty}\beta_k < \infty$*, and* $\lim_{k\to\infty}\frac{\beta_k}{\alpha_k} = 0$*. Then,* $v_k \to 0$ *almost surely and* $\lim_{k\to\infty}\mathsf{E}[v_k] = 0$*.*

4.2. Convergence analysis. As a key step toward performing the convergence analysis for Algorithm 2.1 when the set X is unbounded, next we derive a recursive inequality for the distance between the generated sequence $\{x_k\}$ by the algorithm and the Tikhonov trajectory $\{x^*_{\eta_k}\}$. To this end, we first make the following assumption.

Assumption 4.8. Consider the problem $(\mathrm{P}^f_{\mathrm{VI}})$ under the following assumptions:

(a) There exist nonnegative scalars L_F and B_F such that for all $x, y \in X$,

$$\|F(x)-F(y)\|^2 \leq L_F^2\|x-y\|^2 + B_F.$$

(b) The gradient mapping ∇f is Lipschitz with parameter $L_f > 0$.

Remark 4.9. By allowing L_F or B_F to be zero, Assumption 4.8 provides a unifying structure for considering both smooth and nonsmooth cases. In particular, when $L_F = 0$, part (a) refers to a bounded but possibly non-Lipschitzian mapping F. Also, when $B_F = 0$, part (a) refers to a Lipschitzian but possibly unbounded mapping F.

The following recursive relation will play a key role in establishing the convergence.

LEMMA 4.10 (a recursive error bound for Algorithm 2.1). *Consider the sequence* $\{x_k\}$ *in Algorithm* 2.1*. Let Assumptions* 4.1, 2.3, *and* 4.8 *hold. Suppose* $\{\gamma_k\}$

and $\{\eta_k\}$ *are nonincreasing and strictly positive where* $\lim_{k\to\infty}\eta_k = 0$ *and* $\frac{\gamma_k}{\eta_k} \le \frac{\mu_f \mathsf{p}_{min}}{2\mathsf{p}_{max}(L_F^2+\eta_0^2 L_f^2)}$ *for all* $k \ge 0$*. Then, for all* $k \ge 1$,

$$\begin{aligned}\mathsf{E}\left[\mathcal{D}\left(x_{k+1}, x^*_{\eta_k}\right)|\mathcal{F}_k\right] &\le \frac{\mathsf{p}_{max}}{\mathsf{p}_{min}}\left(1-\frac{\mathsf{p}_{min}\mu_f\gamma_k\eta_k}{2}\right)\mathcal{D}\left(x_k, x^*_{\eta_{k-1}}\right)\\ &\quad + \frac{\bar{C}_f^2\left(\mu_f\gamma_0\eta_0 + 2/\mathsf{p}_{min}\right)}{\mu_f^3\,\mathsf{p}_{min}\gamma_k\eta_k}\left(\frac{\eta_{k-1}}{\eta_k}-1\right)^2 + 2\gamma_k^2 B_F. \qquad (4.1)\end{aligned}$$

Proof. From Definition 2.5, we have

$$\mathcal{D}\left(x_{k+1}, x^*_{\eta_k}\right) = \mathsf{p}_{i_k}^{-1}\left\|x_{k+1}^{(i_k)} - x^{*^{(i_k)}}_{\eta_k}\right\|^2 + \sum_{i=1,\, i\ne i_k}^{d}\mathsf{p}_i^{-1}\left\|x_k^{(i)} - x^{*^{(i)}}_{\eta_k}\right\|^2. \qquad (4.2)$$

Next, we find a bound on the term $\| x_{k+1}^{(i_k)} - x^{*^{(i_k)}}_{\eta_k}\|^2$. From the properties of the natural map (cf. Proposition 1.5.8 in [13]), Definition 2.9, and that $x^*_{\eta_k}\in X$, we have $x^*_{\eta_k} = \mathcal{P}_X(x^*_{\eta_k} - \gamma_k G_k(x^*_{\eta_k}))$. From Assumption 4.1(a) and that $x^*_{\eta_k}\in \mathrm{SOL}(X, G_k)\subseteq X$, we have $x^{*^{(i_k)}}_{\eta_k}\in X_{i_k}$. Invoking the nonexpansiveness property of the projection mapping, (2.1), and the preceding relation, we obtain

$$\left\|x_{k+1}^{(i_k)} - x^{*^{(i_k)}}_{\eta_k}\right\|^2 \le \left\|x_k^{(i_k)} - \gamma_k G_{k,i_k}\left(x_k\right) - x^{*^{(i_k)}}_{\eta_k} + \gamma_k G_{k,i_k}\left(x^*_{\eta_k}\right)\right\|^2.$$

Combining the preceding relation with (4.2), we obtain

$$\begin{aligned}\mathcal{D}\left(x_{k+1}, x^*_{\eta_k}\right) \le &\sum_{i=1,\, i\ne i_k}^{d}\mathsf{p}_i^{-1}\left\|x_k^{(i)} - x^{*^{(i)}}_{\eta_k}\right\|^2 + \mathsf{p}_{i_k}^{-1}\left\|x_k^{(i_k)} - x^{*^{(i_k)}}_{\eta_k}\right\|^2\\ &- 2\,\mathsf{p}_{i_k}^{-1}\gamma_k\left(x_k^{(i_k)} - x^{*^{(i_k)}}_{\eta_k}\right)^T\left(G_{k,i_k}\left(x_k\right) - G_{k,i_k}\left(x^*_{\eta_k}\right)\right)\\ &+ \mathsf{p}_{i_k}^{-1}\gamma_k^2\left\|G_{k,i_k}\left(x_k\right) - G_{k,i_k}\left(x^*_{\eta_k}\right)\right\|^2.\end{aligned}$$

Invoking Definition 2.5, from the preceding relation we obtain

$$\begin{aligned}\mathcal{D}\left(x_{k+1}, x^*_{\eta_k}\right) \le \mathcal{D}\left(x_k, x^*_{\eta_k}\right) &- 2\,\mathsf{p}_{i_k}^{-1}\gamma_k\left(x_k^{(i_k)} - x^{*^{(i_k)}}_{\eta_k}\right)^T\left(G_{k,i_k}\left(x_k\right) - G_{k,i_k}\left(x^*_{\eta_k}\right)\right)\\ &+ \mathsf{p}_{i_k}^{-1}\gamma_k^2\left\|G_{k,i_k}\left(x_k\right) - G_{k,i_k}\left(x^*_{\eta_k}\right)\right\|^2.\end{aligned}$$

Taking the conditional expectation from both sides of the preceding relation and noting that $\mathcal{D}(x_k, x^*_{\eta_k})$ is $\mathcal{F}_k$-measurable, we obtain the inequality

$$\begin{aligned}\mathsf{E}\left[\mathcal{D}\left(x_{k+1}, x^*_{\eta_k}\right)|\mathcal{F}_k\right] \le \mathcal{D}\left(x_k, x^*_{\eta_k}\right) &+ \gamma_k^2\mathsf{E}\left[\mathsf{p}_{i_k}^{-1}\left\|G_{k,i_k}\left(x_k\right) - G_{k,i_k}\left(x^*_{\eta_k}\right)\right\|^2\right]\\ &- 2\gamma_k\mathsf{E}\left[\mathsf{p}_{i_k}^{-1}\left(x_k^{(i_k)} - x^{*^{(i_k)}}_{\eta_k}\right)^T\left(G_{k,i_k}\left(x_k\right) - G_{k,i_k}\left(x^*_{\eta_k}\right)\right)\right]. \qquad (4.3)\end{aligned}$$

Next, we estimate the second and third expectations in the preceding relation

$$\begin{aligned}&\mathsf{E}\left[\mathsf{p}_{i_k}^{-1}\left(x_k^{(i_k)} - x^{*^{(i_k)}}_{\eta_k}\right)^T\left(G_{k,i_k}\left(x_k\right) - G_{k,i_k}\left(x^*_{\eta_k}\right)\right)\right]\\ &\quad= \sum_{i=1}^{d}\mathsf{p}_i\mathsf{p}_i^{-1}\left(x_k^{(i)} - x^{*^{(i)}}_{\eta_k}\right)^T\left(G_{k,i}\left(x_k\right) - G_{k,i}\left(x^*_{\eta_k}\right)\right)\\ &\quad= \left(x_k - x^*_{\eta_k}\right)^T\left(G_k(x_k) - G_k\left(x^*_{\eta_k}\right)\right). \qquad (4.4)\end{aligned}$$

We can also write

$$\begin{aligned}&\mathsf{E}\Big[\mathsf{p}_{i_k}^{-1}\left\|G_{k,i_k}(x_k)-G_{k,i_k}\left(x^*_{\eta_k}\right)\right\|^2\Big]\\ &\quad=\sum_{i=1}^d \mathsf{p}_i\mathsf{p}_i^{-1}\left\|G_{k,i}(x_k)-G_{k,i}\left(x^*_{\eta_k}\right)\right\|^2=\left\|G_k(x_k)-G_k\left(x^*_{\eta_k}\right)\right\|^2.\end{aligned}\tag{4.5}$$

From Assumption 4.8, taking into account that G_k is $(\eta_k\mu_f)$–strongly monotone, and combining (4.3), (4.4), and (4.5) we obtain

$$\begin{aligned}\mathsf{E}\big[\mathcal{D}\left(x_{k+1},x^*_{\eta_k}\right)|\mathcal{F}_k\big]\le\mathcal{D}\left(x_k,x^*_{\eta_k}\right)-2\mu_f\gamma_k\eta_k\left\|x_k-x^*_{\eta_k}\right\|^2\\ +2\gamma_k^2\left(\left(L_F^2+\eta_k^2L_f^2\right)\left\|x_k-x^*_{\eta_k}\right\|^2+B_F\right).\end{aligned}$$

From Remark 2.6 and that $\{\eta_k\}$ is a nonincreasing sequence, we obtain

$$\begin{aligned}\mathsf{E}\big[\mathcal{D}\left(x_{k+1},x^*_{\eta_k}\right)|\mathcal{F}_k\big]\le\left(1-2\mu_f\gamma_k\eta_k\mathsf{p}_{min}+2\gamma_k^2\mathsf{p}_{max}\left(L_F^2+{\eta_0}^2L_f^2\right)\right)\mathcal{D}\left(x_k,x^*_{\eta_k}\right)\\ +2\gamma_k^2B_F.\end{aligned}$$

From the assumption $\gamma_k\le\frac{\mu_f\eta_k\mathsf{p}_{min}}{2\mathsf{p}_{max}(L_F^2+\eta_0^2L_f^2)}$ and the preceding inequality, we obtain

$$\mathsf{E}\big[\mathcal{D}\left(x_{k+1},x^*_{\eta_k}\right)|\mathcal{F}_k\big]\le\left(1-\mu_f\gamma_k\eta_k\mathsf{p}_{min}\right)\mathcal{D}\left(x_k,x^*_{\eta_k}\right)+2\gamma_k^2B_F.\tag{4.6}$$

The preceding relation is not yet fully recursive as the term $x^*_{\eta_k}$ on the right-hand side must change to $x^*_{\eta_{k-1}}$. Next, we find an upper bound for $\mathcal{D}(x_k,x^*_{\eta_k})$ in terms of $\mathcal{D}(x_k,x^*_{\eta_{k-1}})$. Note that we have $\|u+v\|^2\le(1+\theta)\|u\|^2+(1+\frac{1}{\theta})\|v\|^2$ for any vectors $u,v\in\mathbb{R}^n$ and $\theta>0$. Utilizing this inequality, by setting $u:=x_k-x^*_{\eta_{k-1}}$, $v:=x^*_{\eta_{k-1}}-x^*_{\eta_k}$, and $\theta:=\frac{\mathsf{p}_{min}\mu_f\gamma_k\eta_k}{2}$ we obtain

$$\begin{aligned}\left\|x_k-x^*_{\eta_k}\right\|^2\le\left(1+\frac{\mathsf{p}_{min}\mu_f\gamma_k\eta_k}{2}\right)\left\|x_k-x^*_{\eta_{k-1}}\right\|^2\\ +\left(1+\frac{2}{\mathsf{p}_{min}\mu_f\gamma_k\eta_k}\right)\left\|x^*_{\eta_{k-1}}-x^*_{\eta_k}\right\|^2.\end{aligned}$$

Together with Lemma 4.5(b) and Remark 2.6, we have

$$\begin{aligned}\mathsf{p}_{min}\,\mathcal{D}\left(x_k,x^*_{\eta_k}\right)\le\left(1+\frac{\mathsf{p}_{min}\mu_f\gamma_k\eta_k}{2}\right)\mathsf{p}_{max}\,\mathcal{D}\left(x_k,x^*_{\eta_{k-1}}\right)\\ +\left(1+\frac{2}{\mathsf{p}_{min}\mu_f\gamma_k\eta_k}\right)\frac{\bar{C}_f^2}{\mu_f^2}\left(1-\frac{\eta_{k-1}}{\eta_k}\right)^2.\end{aligned}$$

Dividing both sides by p_{min} and substituting this in (4.6), we obtain

$$\begin{aligned}\mathsf{E}\big[\mathcal{D}\left(x_{k+1},x^*_{\eta_k}\right)|\mathcal{F}_k\big]\le\frac{\mathsf{p}_{max}}{\mathsf{p}_{min}}\left(1-\gamma_k\eta_k\mu_f\mathsf{p}_{min}\right)\left(1+\frac{\mathsf{p}_{min}\mu_f\gamma_k\eta_k}{2}\right)\mathcal{D}\left(x_k,x^*_{\eta_{k-1}}\right)\\ +\frac{\bar{C}_f^2}{\mu_f^2\,\mathsf{p}_{min}}\left(1+\frac{2}{\mathsf{p}_{min}\mu_f\gamma_k\eta_k}\right)\left(1-\frac{\eta_{k-1}}{\eta_k}\right)^2+2\gamma_k^2B_F.\end{aligned}$$

(4.1) is obtained by noting that $(1-\gamma_k\eta_k\mu_f\mathsf{p}_{min})(1+\frac{\mathsf{p}_{min}\mu_f\gamma_k\eta_k}{2})\le 1-\frac{\mathsf{p}_{min}\mu_f\gamma_k\eta_k}{2}$. ∎

In the following result, we provide a class of update rules for the stepsize and the regularization sequences such that Algorithm 2.1 attains both an almost sure convergence and a convergence in the mean sense.

THEOREM 4.11 (convergence of Algorithm 2.1 when X is unbounded). *Consider* $(\mathrm{P}^f_{\mathrm{VI}})$. *Let the sequence* $\{\bar{x}_k\}$ *be generated by Algorithm* 2.1. *Let Assumptions* 4.1, 2.3, *and* 4.8 *hold. Suppose the random block-coordinate* i_k *in Assumption* 2.3 *is drawn from a uniform distribution for all* $k \geq 0$. *Let the stepsize* $\{\gamma_k\}$ *and the regularization parameter* $\{\eta_k\}$ *be given by* $\gamma_k := \gamma_0(k+1)^{-a}$ *and* $\eta_k := \eta_0(k+1)^{-b}$, *respectively, where* $\gamma_0 > 0$, $\eta_0 > 0$, $0 < b < 0.5 < a$, *and* $a + b < 1$. *Then, the following results hold for all* $0 \leq r < 1$:

(i) *The sequence* $\{\bar{x}_k\}$ *converges almost surely to the unique optimal solution of* $(\mathrm{P}^f_{\mathrm{VI}})$.

(ii) *We have that* $\lim_{k\to\infty} \mathsf{E}[\|\bar{x}_k - x^*\|] = 0$.

Proof. The proof is done in two main steps. In the first step, we show that the nonaveraged sequence $\{x_k\}$ converges to x^* in an almost sure sense and that $\lim_{k\to\infty} \mathsf{E}[\|x_k - x^*\|] = 0$. In the second step, we show that these results hold for the weighted average sequence $\{\bar{x}_k\}$ as well.

Step 1. The proof of this step is done by applying Lemma 4.7 to the recursive inequality (4.1) with $\mathsf{p}_i := \frac{1}{d}$ for all $i \in \{1, \ldots, d\}$. The details are as follows. First, we note that from the update rules of γ_k and η_k and that $a > b$, we have $\lim_{k\to\infty} \frac{\gamma_k}{\eta_k} = 0$. Thus, there exists an integer $k_0 \geq 1$ such that for all $k \geq k_0$, we have $\frac{\gamma_k}{\eta_k} \leq \frac{\mu_f \mathsf{p}_{min}}{2\mathsf{p}_{max}(L_F^2 + \eta_0^2 L_f^2)}$. This implies that the conditions of Lemma 4.10 are satisfied and the inequality (4.1) holds for all $k \geq k_0$. To apply Lemma 4.7, we define the following terms for all $k \geq 1$:

$$v_k \triangleq \mathcal{D}\left(x_k, x^*_{\eta_{k-1}}\right), \quad \alpha_k \triangleq \frac{\mu_f \gamma_k \eta_k}{2d},$$
$$\beta_k \triangleq \left(\frac{d\bar{C}_f^2\left(\mu_f \eta_0 \gamma_0 + 2d\right)}{\mu_f^3 \gamma_k \eta_k}\right)\left(\frac{\eta_{k-1}}{\eta_k} - 1\right)^2 + 2\gamma_k^2 B_F.$$

Since $\gamma_k \eta_k \to 0$, there exists an integer $k_1 \geq k_0$ such that for any $k \geq k_1$ we have $0 \leq \alpha_k \leq 1$. From the assumption that $a+b < 1$, we have that $\sum_{k=k_1}^{\infty} \alpha_k = \infty$. Next, we show that $\sum_{k=k_1}^{\infty} \beta_k < \infty$. From the update rules of γ_k and η_k and invoking the Taylor series expansion, for $k \geq 2$ we can write

$$\begin{aligned}\frac{\eta_{k-1}}{\eta_k} - 1 &= \left(1 + \frac{1}{k}\right)^b - 1 = \left(1 + \frac{b}{k} + \frac{b(b-1)}{2!}\frac{1}{k^2} + \frac{b(b-1)(b-2)}{3!}\frac{1}{k^3} + \cdots\right) - 1 \\ &= \frac{b}{k}\left(1 - \frac{(1-b)}{2!k} + \frac{(1-b)(2-b)}{3!k^2} - \frac{(1-b)(2-b)(3-b)}{4!k^3} + \cdots\right) \leq \frac{b}{k}\sum_{i=0}^{\infty}\frac{1}{k^{2i}},\end{aligned}$$

where the inequality is obtained using $b < 1$ and neglecting the negative terms. This implies that $\frac{\eta_{k-1}}{\eta_k} - 1 \leq \frac{b}{k(1-k^{-2})}$ and thus $(\frac{\eta_{k-1}}{\eta_k} - 1)^2 \leq (\frac{4b}{3k})^2 \leq \frac{2b^2}{k^2}$ for all $k \geq 2$. Using the preceding relation, invoking the definition of β_k and the update formulas of γ_k and η_k, we have that $\beta_k = \mathcal{O}(k^{-(2-a-b)}) + \mathcal{O}(k^{-2a})$. From the assumptions on a and b, we obtain that $\sum_{k=k_1}^{\infty} \beta_k < \infty$. Also, from the assumption $a > b$, we get $\lim_{k\to\infty} \beta_k/\alpha_k = 0$. Therefore, all conditions of Lemma 4.7 are satisfied. As such, we have that $\mathcal{D}(x_k, x^*_{\eta_{k-1}}) \to 0$ almost surely and also $\lim_{k\to\infty} \mathsf{E}\left[\mathcal{D}(x_k, x^*_{\eta_{k-1}})\right] = 0$. From Remark 2.6 and that i_k is drawn uniformly, we obtain

$$\begin{aligned}\|x_k - x^*\|^2 &\leq 2\left\|x_k - x^*_{\eta_{k-1}}\right\|^2 + 2\left\|x^*_{\eta_{k-1}} - x^*\right\|^2 \\ &= \frac{2}{d}\mathcal{D}\left(x_k, x^*_{\eta_{k-1}}\right) + 2\left\|x^*_{\eta_{k-1}} - x^*\right\|^2, \qquad (4.7)\end{aligned}$$

where the first inequality is obtained from the triangle inequality. Taking the limit from both sides of the preceding relation when $k \to \infty$ and invoking Lemma 4.5(a), we obtain $\lim_{k\to\infty} \|x_k - x^*\|^2 \leq \frac{2}{d} \lim_{k\to\infty} \mathcal{D}(x_k, x^*_{\eta_{k-1}})$. From the almost sure convergence of $\mathcal{D}(x_k, x^*_{\eta_{k-1}})$ to zero, we conclude that $\{x_k\}$ converges to x^* almost surely. To show the convergence in mean, let us take the expectation from both sides of (4.7). Noting that the Tikhonov trajectory is deterministic, we obtain that $\mathsf{E}\left[\|x_k - x^*\|^2\right] \leq \frac{2}{d}\mathsf{E}\left[\mathcal{D}(x_k, x^*_{\eta_{k-1}})\right] + 2\|x^*_{\eta_{k-1}} - x^*\|^2$. Now, taking the limit from both sides of the preceding relation when $k \to \infty$, invoking Lemma 4.5(a), and recalling $\lim_{k\to\infty} \mathsf{E}\left[\mathcal{D}(x_k, x^*_{\eta_{k-1}})\right] = 0$, we conclude that $\lim_{k\to\infty} \mathsf{E}\left[\|x_k - x^*\|^2\right] = 0$. Invoking Jensen's inequality, we can conclude that $\lim_{k\to\infty} \mathsf{E}\left[\|x_k - x^*\|\right] = 0$.

Step 2. Invoking Lemma 2.11 and using the triangle inequality, we have

$$\|\bar{x}_k - x^*\| = \left\|\sum_{t=0}^{k} \lambda_{t,k} x_t - x^*\right\| = \left\|\sum_{t=0}^{k} \lambda_{t,k}\left(x_t - x^*\right)\right\| \leq \sum_{t=0}^{k} \lambda_{t,k}\left\|x_t - x^*\right\|, \tag{4.8}$$

where $\lambda_{t,k} \triangleq \gamma_t^r / \sum_{j=0}^k \gamma_j^r$. In view of Lemma 4.6, let us define $u_t \triangleq \|x_t - x^*\|$, $v_k \triangleq \sum_{t=0}^k \lambda_{t,k}\|x_t - x^*\|$, and $\alpha_t \triangleq \gamma_t^r$. Note that since $ar \leq 1$, we have $\sum_{t=0}^\infty \alpha_t = \gamma_0^r \sum_{t=0}^\infty (t+1)^{-ar} = \infty$. Also, from Step 1 we have that $\hat{u} \triangleq \lim_{t\to\infty} u_t = 0$ in an almost sure sense. Thus, from Lemma 4.6, we conclude that $\{v_k\}$ converges to zero almost surely. Thus, (4.8) implies that $\{\bar{x}_k\}$ converges to x^* almost surely. Next, we apply Lemma 4.6 again, but in a slightly different fashion to show that $\lim_{k\to\infty} \mathsf{E}[\|\bar{x}_k - x^*\|] = 0$. From (4.8), we have

$$\mathsf{E}[\|\bar{x}_k - x^*\|] \leq \sum_{t=0}^{k} \lambda_{t,k} \mathsf{E}[\|x_t - x^*\|]\,. \tag{4.9}$$

In view of Lemma 4.6, let us define $u_t \triangleq \mathsf{E}[\|x_t - x^*\|]$, $v_k \triangleq \sum_{t=0}^k \lambda_{t,k}\mathsf{E}[\|x_t - x^*\|]$, and $\alpha_t \triangleq \gamma_t^r$. First, note that from Step 1, we have $\hat{u} \triangleq \lim_{t\to\infty} u_t = 0$. In view of Lemma 4.6, $\lim_{k\to\infty} v_k = 0$. Thus, from (4.9), we conclude that $\lim_{k\to\infty} \mathsf{E}[\|\bar{x}_k - x^*\|] = 0$. Hence, the proof is completed. ∎

5. Experimental results. In this section, we revisit the problem of finding the best Nash equilibrium formulated as in (1.1). We consider a case where the Nash game is characterized as a Cournot competition over a network. The Cournot game is one of the most extensively studied economic models for competition among multiple firms, including imperfectly competitive power markets as well as rate control over communication networks [19, 21, 13]. Consider a collection of d firms who compete to sell a commodity over a network with J nodes. The decision of each firm $i \in \{1, \ldots, d\}$ includes variables y_{ij} and s_{ij}, denoting the generation and sales of the firm i at the node j, respectively. Considering the definitions $y_i \triangleq (y_{i1}; \ldots; y_{iJ})$ and $s_i \triangleq (s_{i1}; \ldots; s_{iJ})$, we can compactly denote the decision variable of the ith firm as $x^{(i)} \triangleq (y_i; s_i) \in \mathbb{R}^{2J}$. The goal of the ith firm lies in minimizing the net cost function $g_i(x^{(i)}, x^{(-i)})$ over the network defined as follows

$$g_i\left(x^{(i)}; x^{(-i)}\right) \triangleq \sum_{j=1}^{J} c_{ij}(y_{ij}) - \sum_{j=1}^{J} s_{ij} p_j\left(\bar{s}_j\right),$$

where $c_{ij} : \mathbb{R} \to \mathbb{R}$ denotes the production cost function of the firm i at the node j, $\bar{s}_j \triangleq \sum_{i=1}^d s_{ij}$ denotes the aggregate sales from all the firms at the node j, and

$p_j : \mathbb{R} \to \mathbb{R}$ denotes the price function with respect to the aggregate sales $\bar{s}_j$ at the node j. We assume that the cost functions are linear and the price functions are given as $p_j(\bar{s}_j) \triangleq \alpha_j - \beta_j(\bar{s}_j)^\sigma$ where $\sigma \geq 1$ and α_j and β_j are positive scalars. Throughout, we assume that the transportation costs are negligible. We let the generation be capacitated as $y_{ij} \leq \mathcal{B}_{ij}$, where $\mathcal{B}_{ij}$ is a positive scalar for $i \in \{1, \ldots, d\}$ and $j \in \{1, \ldots, J\}$. Last, for any firm i, the total sales must match with the total generation. Consequently, the strategy set of the firm i is given as follows:

$$X_i \triangleq \left\{ (y_i; s_i) \mid \sum_{j=1}^{J} y_{ij} = \sum_{j=1}^{J} s_{ij}, \quad y_{ij}, s_{ij} \geq 0, \quad y_{ij} \leq \mathcal{B}_{ij}, \quad \text{for all } j = 1, \ldots, J \right\}.$$

Following the model (1.1), we employ the Marshallian aggregate surplus function defined as $f(x) \triangleq \sum_{i=1}^{d} g_i\left(x^{(i)}; x^{(-i)}\right)$. We note that the convexity of the function f is implied by $\sigma \geq 1$ and the monotonicity of mapping F is guaranteed either when $\sigma = 1$ or when $1 < \sigma \leq 3$ and $d \leq \frac{3\sigma - 1}{\sigma - 1}$ (cf. section 4 in [21]).

The set-up. In the experiment, we consider a Cournot game among four firms over three nodes. We let the slopes of the linear cost functions take values between 10 and 50. We assume that $\alpha_j := 50$ and $\beta_j := 0.05$ for all j, $\mathcal{B}_{ij} := 120$ for all i and j, and $\sigma := 1.01$. To report the performance of Algorithm 2.1 in terms of the suboptimality, we plot a sample average approximation of $\mathsf{E}[f(\bar{x}_N)]$ using the sample size of 25. With regard to the infeasibility, we compute a sample average approximation of $\mathsf{E}[\mathrm{GAP}(\bar{x}_N)]$ using the same sample size. Following Remark 3.4, we use $\gamma_k := \frac{\gamma_0}{\sqrt{k+1}}$ and $\eta_k := \frac{\eta_0}{\sqrt[4]{k+1}}$. To select the block-coordinates in Algorithm 2.1, we use a discrete uniform distribution.

Results and insights. Figure 1 shows the experimental results. Here, in the top three figures, we compare the performance of Algorithm 2.1 with that of Algorithm 1.1 in terms of infeasibility measured by the sample averaged gap function. Importantly, the proposed algorithm performs significantly better than the SR scheme. This claim is supported by considering the different values of the parameter r and the initial conditions of the proposed scheme in terms of the initial stepsize γ_0 and the initial

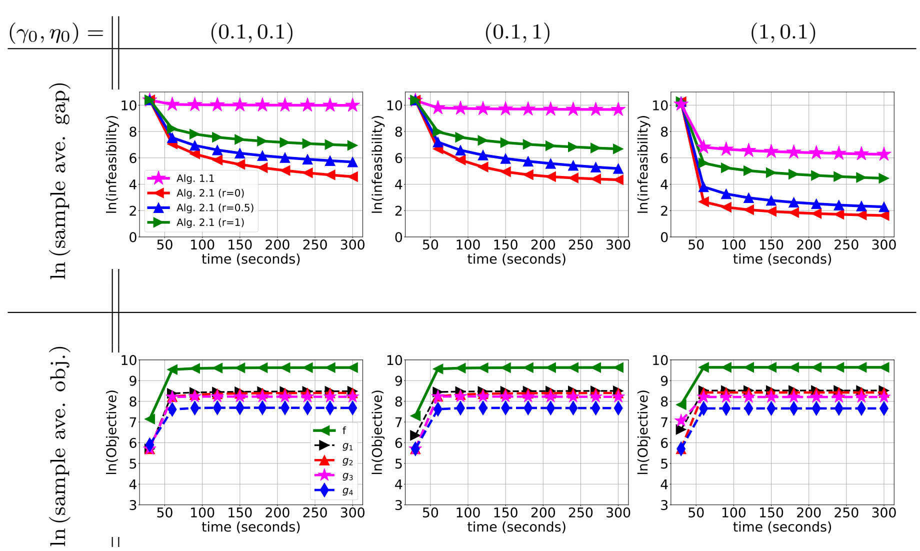

FIG. 1. *Algorithm* 2.1 *in terms of infeasibility and the objective function value.*

regularization parameter η_0. The three figures in the bottom row of Figure 1 demonstrate the performance of Algorithm 2.1 in terms of reaching a stability in the objective values. This includes the Marshallian objective function f as well as the individual objective functions g_i. Note that all the objective values in Figure 1 appear to reach a desired level of stability after around 60 seconds. This interesting observation could be linked to the impact of the averaging scheme (2.2). Generally, it is expected that the trajectories of the objective function values in Figure 1 will be noisy due to the randomness in the block-coordinate selection rule. However, the weighted averaging scheme employed in Algorithm 2.1 appears to induce much robustness with respect to this uncertainty, resulting in an accelerated convergence.

6. Conclusions. Motivated by the applications arising from noncooperative multiagent networks, we consider a class of optimization problems with Cartesian variational inequality (CVI) constraints. The computational complexity of the solution methods for addressing this class of problems appears to be unknown. We develop a single timescale algorithm equipped with nonasymptotic suboptimality and infeasibility convergence rates. Moreover, in the case where the set associated with the CVI is unbounded, we establish the global convergence of the sequence generated by the proposed algorithm. We apply the method in computing the best Nash equilibrium in a networked Cournot competition. Our experimental results show that the proposed method outperforms the classical sequential regularized schemes.

Appendix A. Additional proofs.

A.1. Proof of Lemma 1.3. Let us define the function $\phi : \mathbb{R}^n \to \mathbb{R}$ as $\phi(x) \triangleq \frac{1}{2}\|Ax-b\|^2 + \frac{1}{2}\sum_{j=1}^{J}(\max\{0, h_j(x)\})^2$. We first note that ϕ is a differentiable function such that $\nabla\phi(x) = F(x)$ where F is given by Lemma 1.3 (e.g., see page 380 in [6]). Next, we also note that ϕ is convex. To see this, note that from the convexity of $h_j(x)$, the function $h_j^+(x) \triangleq \max\{0, h_j(x)\}$ is convex. Then, the function $(h_j^+(x))^2$ can be viewed as a composition of $s(u) \triangleq u^2$ for $u \in \mathbb{R}$ and the convex function h_j^+. Since h_j^+ is nonnegative on its domain and $s(u)$ is nondecreasing on $[0, +\infty)$, we have that $(h_j^+(x))^2$ is a convex function. As such, ϕ is a convex function as well. Consequently, from the first-order optimality conditions for convex programs, we have $\mathrm{SOL}(X, F) = \mathrm{argmin}_{x\in X}\,\phi(x)$. To show the desired equivalence between problems $(\mathrm{P}_{\mathrm{VI}}^f)$ and (1.2), it suffices to show that $\mathcal{X} = \mathrm{argmin}_{x\in X}\,\phi(x)$ where $\mathcal{X}$ denotes the feasible set of problem (1.2). To show this statement, first we let $\bar{x} \in \mathcal{X}$. Then, from the definition of $\phi(x)$, we have $\phi(\bar{x}) = 0$. This implies that $\bar{x} \in \mathrm{argmin}_{x\in X}\,\phi(x)$. Thus, we have $\mathcal{X} \subseteq \mathrm{argmin}_{x\in X}\,\phi(x)$. Second, let $\tilde{x} \in \mathrm{argmin}_{x\in X}\,\phi(x)$. The feasibility assumption of the set $\mathcal{X}$ implies that there exists an $x_0 \in X$ such that $Ax_0 = b$ and $h_j(x_0) \leq 0$ for all j. This implies that $\phi(x_0) = 0$. From the nonnegativity of ϕ and that $\tilde{x} \in \mathrm{argmin}_{x\in X}\,\phi(x)$, we must have $\phi(\tilde{x}) = 0$ and $\tilde{x} \in X$. Therefore, we obtain $A\tilde{x} = b$, $h_j(\tilde{x}) \leq 0$ for all j, and $\tilde{x} \in X$. Thus, we have $\mathrm{argmin}_{x\in X}\,\phi(x) \subseteq \mathcal{X}$. Hence, we conclude that $\mathcal{X} = \mathrm{argmin}_{x\in X}\,\phi(x) = \mathrm{SOL}(X, F)$ and the proof is completed.

A.2. Proof of Lemma 2.11. We use induction to show $\bar{x}_N = \sum_{k=0}^{N}\lambda_{k,N}x_k$ for any $N \geq 0$. For $N = 0$, the relation holds due to the initialization $\bar{x}_0 := x_0$ in Algorithm 2.1 and that $\lambda_{0,0} = 1$. Next, let the relation hold for some $N \geq 0$. From the hypothesis, (2.2), and that $S_N = \sum_{k=0}^{N}\gamma_k^r$ for all $N \geq 0$, we can write

$$\bar{x}_{N+1} = \frac{S_N\bar{x}_N + \gamma_{N+1}^r x_{N+1}}{S_{N+1}} = \frac{\sum_{k=0}^{N+1}\gamma_k^r x_k}{\sum_{k=0}^{N+1}\gamma_k^r} = \sum_{k=0}^{N+1}\lambda_{k,N+1}x_k,$$

implying that the induction hypothesis holds for $N+1$. Thus, we conclude that the desired averaging formula holds for all $N \geq 0$. To complete the proof, note that since $\sum_{k=0}^{N} \lambda_{k,N} = 1$, under the convexity of the set X, we have $\bar{x}_N \in X$.

A.3. Proof of Lemma 2.13. (a) From Definition 2.12, we can write

$$\mathsf{E}[\Delta_k \mid \mathcal{F}_k] = F(x_k) - \sum_{i=1}^{d} \mathsf{p}_i \mathsf{p}_i^{-1} \mathbf{U}_i F_i(x_k) = F(x_k) - \sum_{i=1}^{d} \mathbf{U}_i F_i(x_k) = 0.$$

The relation $\mathsf{E}[\delta_k \mid \mathcal{F}_k] = 0$ can be shown in a similar fashion.
(b) We can write

$$\begin{aligned}
\mathsf{E}\left[\|\Delta_k\|^2 \mid \mathcal{F}_k\right] &= \sum_{i=1}^{d} \mathsf{p}_i \left\|F(x_k) - \mathsf{p}_i^{-1} \mathbf{U}_i F_i(x_k)\right\|^2 \\
&= \sum_{i=1}^{d} \mathsf{p}_i \left(\|F(x_k)\|^2 + \mathsf{p}_i^{-2} \|\mathbf{U}_i F_i(x_k)\|^2 - 2\mathsf{p}_i^{-1} F(x_k)^T \mathbf{U}_i F_i(x_k)\right) \\
&= \|F(x_k)\|^2 + \sum_{i=1}^{d} \mathsf{p}_i^{-1} \|\mathbf{U}_i F_i(x_k)\|^2 - 2\sum_{i=1}^{d} \|F_i(x_k)\|^2 \leq \left(\mathsf{p}_{min}^{-1} - 1\right) C_F^2.
\end{aligned}$$

The relation $\mathsf{E}\left[\|\delta_k\|^2 \mid \mathcal{F}_k\right] \leq \left(\mathsf{p}_{min}^{-1} - 1\right) C_f^2$ can be shown using a similar approach.

A.4. Proof of Lemma 2.14. Given $0 \leq \alpha < 1$, let us define the function $\phi : \mathbb{R}_{++} \to \mathbb{R}$ as $\phi(x) \triangleq x^{-\alpha}$ for all $x > 0$. Since $\alpha \geq 0$, the function ϕ is nonincreasing. We can write

$$\sum_{k=0}^{N} \frac{1}{(k+1)^\alpha} = 1 + \sum_{k=2}^{N+1} \frac{1}{k^\alpha} \leq 1 + \int_1^{N+1} \frac{dx}{x^\alpha} = 1 + \frac{(N+1)^{1-\alpha} - 1}{1-\alpha} \leq \frac{(N+1)^{1-\alpha}}{1-\alpha},$$

implying the desired upper bound. To show that the lower bound holds, we can write

$$\sum_{k=0}^{N} \frac{1}{(k+1)^\alpha} = \sum_{k=1}^{N+1} \frac{1}{k^\alpha} \geq \int_1^{N+2} \frac{dx}{x^\alpha} \geq \int_1^{N+1} \frac{dx}{x^\alpha} \geq \frac{(N+1)^{1-\alpha} - 0.5(N+1)^{1-\alpha}}{1-\alpha},$$

where the last inequality is obtained using the assumption that $N \geq 2^{\frac{1}{1-\alpha}} - 1$. Therefore, the desired lower bound holds as well. This completes the proof.

A.5. Proof of Lemma 4.5. (a) From the definition of x^* and $x^*_{\eta_k}$ (cf. Definition 4.3), we have that

$$F(x^*)^T (x - x^*) \geq 0 \qquad \text{for all } x \in X, \tag{A.1}$$

$$\left(F\left(x^*_{\eta_k}\right) + \eta_k \nabla f\left(x^*_{\eta_k}\right)\right)^T \left(y - x^*_{\eta_k}\right) \geq 0 \qquad \text{for all } y \in X. \tag{A.2}$$

For $x := x^*_{\eta_k}$ and $y := x^*$, adding the resulting two relations together, we obtain

$$\eta_k \nabla f\left(x^*_{\eta_k}\right)^T \left(x^* - x^*_{\eta_k}\right) \geq \left(F\left(x^*\right) - F\left(x^*_{\eta_k}\right)\right)^T \left(x^* - x^*_{\eta_k}\right).$$

From the monotonicity of the mapping F and the preceding relation, we obtain that $\nabla f(x^*_{\eta_k})^T(x^* - x^*_{\eta_k}) \geq 0$. Also, from the strong convexity of f, we have

$$f(x^*) \geq f\left(x^*_{\eta_k}\right) + \nabla f\left(x^*_{\eta_k}\right)^T \left(x^* - x^*_{\eta_k}\right) + \frac{\mu_f}{2} \left\|x^* - x^*_{\eta_k}\right\|^2.$$

From the preceding relations, we obtain

$$f(x^*) \geq f\left(x^*_{\eta_k}\right) + \frac{\mu_f}{2}\left\|x^* - x^*_{\eta_k}\right\|^2 \qquad \text{for all } k \geq 0. \tag{A.3}$$

Thus, $f(x^*) \geq f(x^*_{\eta_k})$ for all $k \geq 0$. Recall that from Remark 4.2, under Assumption 4.1, $x^* \in X$ and $x^*_{\eta_k} \in X$ both exist and are unique. Therefore, $f(x^*_{\eta_k})$ is bounded above for all $k \geq 0$. From this statement and invoking the coercive property of f (implied by the strong convexity of f), we can conclude that $\{x^*_{\eta_k}\}$ is a bounded sequence. Therefore, it must have at least one limit point. Let $\{x^*_{\eta_k}\}_{k\in\mathcal{K}}$ be an arbitrary subsequence such that $\lim_{k\to\infty,\ k\in\mathcal{K}} x^*_{\eta_k} = \hat{x}$, where $\lim_{k\to\infty,\ k\in\mathcal{K}}$ denotes the subsequential limit when $k \in \mathcal{K}$ and k goes to infinity. We show that $\hat{x} \in \mathrm{SOL}(X, F)$. Taking the limit from both sides of (A.2) with respect to the aforementioned subsequence and using the continuity of F and ∇f, we obtain that for all $y \in X$, $(F(\hat{x}) + \lim_{k\to\infty,\ k\in\mathcal{K}} \eta_k \nabla f(\hat{x}))^T(y - \hat{x}) \geq 0$. Note that the mapping $\nabla f(\hat{x})$ is bounded. This is because $\hat{x} \in X$ (due to the closedness of X) and ∇f is continuous on the set X. Therefore, from the preceding inequality and $\lim_{k\to\infty} \eta_k = 0$, we obtain $F(\hat{x})^T(y - \hat{x}) \geq 0$ for all $y \in X$, implying that $\hat{x} \in \mathrm{SOL}(X, F)$ and so $\hat{x}$ is a feasible solution to $(\mathrm{P}^f_{\mathrm{VI}})$. Next, we show that $\hat{x}$ is the optimal solution to $(\mathrm{P}^f_{\mathrm{VI}})$. From (A.3), continuity of f, and neglecting the term $\frac{\mu_f}{2}\|x^* - x^*_{\eta_k}\|^2$, we obtain $f(x^*) \geq f(\lim_{k\to\infty,\ k\in\mathcal{K}} x^*_{\eta_k}) = f(\hat{x})$. Hence, from the uniqueness of x^*, all the limit points of $\{x^*_{\eta_k}\}$ fall in the singleton $\{x^*\}$ and the proof is completed.

(b) If $x^*_{\eta_k} = x^*_{\eta_{k-1}}$, the desired relation holds. Suppose for $k \geq 1$, we have $x^*_{\eta_k} \neq x^*_{\eta_{k-1}}$. From $x^*_{\eta_{k-1}} \in \mathrm{SOL}\left(X, F + \eta_{k-1}\nabla f\right)$ and $x^*_{\eta_k} \in \mathrm{SOL}\left(X, F + \eta_k \nabla f\right)$, we have that

$$\left(F\left(x^*_{\eta_{k-1}}\right) + \eta_{k-1}\nabla f\left(x^*_{\eta_{k-1}}\right)\right)^T\left(x - x^*_{\eta_{k-1}}\right) \geq 0 \qquad \text{for all } x \in X,$$
$$\left(F\left(x^*_{\eta_k}\right) + \eta_k \nabla f\left(x^*_{\eta_k}\right)\right)^T\left(y - x^*_{\eta_k}\right) \geq 0 \qquad \text{for all } y \in X.$$

Adding the resulting two relations together, for $x := x^*_{\eta_k}$ and $y := x^*_{\eta_{k-1}}$ we have

$$\left(-F\left(x^*_{\eta_k}\right) - \eta_k \nabla f\left(x^*_{\eta_k}\right) + F(x^*_{\eta_{k-1}}) + \eta_{k-1}\nabla f\left(x^*_{\eta_{k-1}}\right)\right)^T\left(x^*_{\eta_k} - x^*_{\eta_{k-1}}\right) \geq 0.$$

The monotonicity of F implies that $(F(x^*_{\eta_k}) - F(x^*_{\eta_{k-1}}))^T(x^*_{\eta_k} - x^*_{\eta_{k-1}}) \geq 0$. Adding this relation to the preceding inequality, we have

$$\left(\eta_k \nabla f\left(x^*_{\eta_k}\right) - \eta_{k-1}\nabla f\left(x^*_{\eta_{k-1}}\right)\right)^T\left(x^*_{\eta_{k-1}} - x^*_{\eta_k}\right) \geq 0.$$

Adding and subtracting the term $\eta_k \nabla f(x^*_{\eta_{k-1}})^T(x^*_{\eta_{k-1}} - x^*_{\eta_k})$, we obtain

$$\begin{aligned} &(\eta_k - \eta_{k-1})\nabla f\left(x^*_{\eta_{k-1}}\right)^T\left(x^*_{\eta_{k-1}} - x^*_{\eta_k}\right) \\ &\quad \geq \eta_k\left(\nabla f\left(x^*_{\eta_{k-1}}\right) - \nabla f\left(x^*_{\eta_k}\right)\right)^T\left(x^*_{\eta_{k-1}} - x^*_{\eta_k}\right). \end{aligned} \tag{A.4}$$

From the strong convexity of function f, we have

$$\left(\nabla f\left(x^*_{\eta_{k-1}}\right) - \nabla f\left(x^*_{\eta_k}\right)\right)^T\left(x^*_{\eta_{k-1}} - x^*_{\eta_k}\right) \geq \mu_f\left\|x^*_{\eta_k} - x^*_{\eta_{k-1}}\right\|^2. \tag{A.5}$$

From (A.4) and (A.5), and using the Cauchy–Schwarz inequality, we obtain

$$|\eta_k - \eta_{k-1}|\left\|\nabla f\left(x^*_{\eta_{k-1}}\right)\right\|\left\|x^*_{\eta_{k-1}} - x^*_{\eta_k}\right\| \geq \eta_k \mu_f\left\|x^*_{\eta_k} - x^*_{\eta_{k-1}}\right\|^2.$$

Since $x^*_{\eta_k} \neq x^*_{\eta_{k-1}}$, dividing the both sides by $\eta_k \|x^*_{\eta_k} - x^*_{\eta_{k-1}}\|$, we obtain

$$\left|1 - \frac{\eta_{k-1}}{\eta_k}\right| \left\|\nabla f\left(x^*_{\eta_{k-1}}\right)\right\| \geq \mu_f \left\|x^*_{\eta_k} - x^*_{\eta_{k-1}}\right\|. \tag{A.6}$$

From part (a), the trajectory $\{x^*_{\eta_k}\}$ is bounded. Also, for any $k \geq 0$, $x^*_{\eta_k} \in X$ by the definition. Since X is closed, there exists a compact set $S \subset X$ such that $\{x^*_{\eta_k}\} \subset S$. This statement and the continuity of ∇f imply that there exists $\bar{C}_f > 0$ such that $\|\nabla f(x^*_{\eta_{k-1}})\| \leq \bar{C}_f$ for all $k \geq 1$. Thus, from (A.6), we obtain the desired inequality.

A.6. Proof of Corollary 3.5. Let us rewrite ($\mathrm{P}^f_{\mathrm{VI}}$) as the equivalent problem:

$$\begin{aligned} &\text{minimize} \quad && f(x) \\ &\text{subject to} \quad && x \in \mathrm{SOL}(Y, F), \end{aligned} \tag{A.7}$$

where $Y \triangleq \prod_{i=1}^{d'} Y_i$ and $d' \triangleq 1$ and $Y_1 \triangleq X$. Note that this setting immediately implies that $Y = X$. Now, let us consider Algorithm 2.1 for solving (A.7) where we assume that $x_0 \in X$ is an arbitrary fixed vector. Since $d' = 1$, Assumption 2.3 holds with $\mathrm{Prob}\,(i_k = 1) = 1$ for all $k \geq 0$. This setting implies that Algorithm 2.1 reduces to a deterministic scheme where step 5 in Algorithm 2.1 is equivalent to the following update rule:

$$x_{k+1} := \mathcal{P}_X\Big(x_k - \gamma_k\left(F\left(x_k\right) + \eta_k \tilde{\nabla} f\left(x_k\right)\right)\Big), \tag{A.8}$$

where we used $Y = Y_1 = X$. Next, we note that from the properties of the Euclidean projection mapping, for any $z \in X$ where $X \triangleq \prod_{i=1}^{d} X_i$, we have that $\mathcal{P}_X(z) = \prod_{i=1}^{d} \mathcal{P}_{X_i}\left(z^{(i)}\right)$. In view of this property, (A.8) compactly represents the d updates given by (3.21). Therefore, Algorithm 3.1 is equivalent to Algorithm 2.1 and thus all the results in Theorem 3.3 will hold with $\mathsf{p}_{min} = 1$. Note that in both (3.18) and (3.19), the expectation is eliminated. This completes the proof.

REFERENCES

[1] T. Alpcan and T. Başar, *A game-theoretic framework for congestion control in general topology networks*, in Proceedings of the 41st IEEE Conference on Decision and Control, 2002, pp. 1218–1224.

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

[3] M. Amini and F. Yousefian, *An Iterative Regularized Mirror Descent Method for Ill-Posed Nondifferentiable Stochastic Optimization*, https://arxiv.org/abs/1901.09506, 2019.

[4] 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 J. Comput., 38 (2008), pp. 1602—-1623.

[5] A. Beck and S. Sabach, *A first order method for finding minimal norm-like solutions of convex optimization problems*, Math. Program., 147 (2014), pp. 25–46.

[6] D. P. Bertsekas, *Nonlinear Programming*, 3rd ed., Athena Scientific, Bellmont, MA, 2016.

[7] Y. Censor, A. Gibali, and S. Reich, *The subgradient extragradient method for solving variational inequalities in Hilbert space*, J. Optim. Theory Appl., 148 (2011), pp. 318–335.

[8] Y. Censor, A. Gibali, and S. Reich, *Extensions of Korpelevich's extragradient method for the variational inequality problem in Euclidean space*, Optimization, 61 (2012), pp. 1119–1132.

[9] Y. Chen, G. Lan, and Y. Ouyang, *Accelerated schemes for a class of variational inequalities*, Math. Program., 165 (2017), pp. 113–149.

[10] J. R. Correa, A. S. Schulz, and N. E. Stier-Moses, *Selfish routing in capacitated networks*, Math. Oper. Res., 29 (2004), pp. 961–976.

[11] C. D. Dang and G. Lan, *On the convergence properties of non-Euclidean extragradient methods for variational inequalities with generalized monotone operators*, Comput. Optim. Appl., 60 (2015), pp. 277–310.
[12] C. D. Dang and G. Lan, *Stochastic block mirror descent methods for nonsmooth and stochastic optimization*, SIAM J. Optim., 25 (2015), pp. 856–881.
[13] F. Facchinei and J.-S. Pang, *Finite-Dimensional Variational Inequalities and Complementarity Problems*, Springer Ser. Oper. Res., Springer-Verlag, New York, 2003.
[14] G. Garrigos, L. Rosasco, and S. Villa, *Iterative regularization via dual diagonal descent*, J. Math. Imaging Vision, 60 (2018), pp. 189–215.
[15] A. N. Iusem, A. Jofré, R. I. Oliveira, and P. Thompson, *Extragradient method with variance reduction for stochastic variational inequalities*, SIAM J. Optim., 27 (2016), pp. 686–724.
[16] A. N. Iusem, A. Jofré, and P. Thompson, *Incremental constraint projection methods for monotone stochastic variational inequalities*, Math. Oper. Res., 44 (2019), pp. 236–263.
[17] A. N. Iusem and M. Nasri, *Korpelevich's method for variational inequality problems in Banach spaces*, J. Global Optim., 50 (2011), pp. 59–76.
[18] H. Jiang and H. Xu, *Stochastic approximation approaches to the stochastic variational inequality problem*, IEEE Trans. Automat. Control, 53 (2008), pp. 1462–1475.
[19] R. Johari, *Efficiency Loss in Market Mechanisms for Resource Allocation*, Ph.D. thesis, MIT, 2004.
[20] A. Juditsky, A. Nemirovski, and C. Tauvel, *Solving variational inequalities with stochastic mirror-prox algorithm*, Stoch. Syst., 1 (2011), pp. 17–58.
[21] A. Kannan and U. V. Shanbhag, *Distributed computation of equilibria in monotone Nash games via iterative regularization techniques*, SIAM J. Optim., 22 (2012), pp. 1177–1205.
[22] A. Kannan, U. V. Shanbhag, and H. M. Kim, *Strategic behavior in power markets under uncertainty*, Energy Systems, 2 (2011), pp. 115–141.
[23] A. Kannan, U. V. Shanbhag, and H. M. Kim, *Addressing supply-side risk in uncertain power markets: Stochastic Nash models, scalable algorithms and error analysis*, Optim. Methods Softw., 28 (2013), pp. 1095–1138.
[24] H. D. Kaushik and F. Yousefian, *A randomized block coordinate iterative regularized subgradient method for high-dimensional ill-posed convex optimization*, in Proceedings of the American Control Conference, IEEE, 2019, pp. 3420–3425, https://doi.org/10.23919/ACC.2019.8815256.
[25] H. D. Kaushik and F. Yousefian, *An incremental gradient method for large-scale distributed nonlinearly constrained optimization*, in Proceedings of the American Control Conference, 2021, pp. 953–958, https://doi.org/10.23919/ACC50511.2021.9483035.
[26] K. Knopp, *Theory and Applications of Infinite Series*, Blackie & Son, Glasgow, 1951.
[27] G. M. Korpelevich, *An extragradient method for finding saddle points and for other problems*, Eknomika Matematicheskie Metody, 12 (1976), pp. 747—-756.
[28] J. Koshal, A. Nedić, and U. V. Shanbhag, *Regularized iterative stochastic approximation methods for stochastic variational inequality problems*, IEEE Trans. Automat. Control, 58 (2013), pp. 594–609.
[29] J. Lei, U. V. Shanbhag, J.-S. Pang, and S. Sen, *On synchronous, asynchronous, and randomized best-response schemes for stochastic Nash games*, Math. Oper. Res., 45 (2020), pp. 157–190.
[30] C. E. Lemke and J. T. Howson, Jr., *Equilibrium points of bimatrix games*, J. SIAM, 12 (1964), pp. 413–423.
[31] P. Marcotte and D. Zhu, *Weak sharp solutions of variational inequalities*, SIAM J. Optim., 9 (1998), pp. 179–189.
[32] A. Nedić, *Random algorithms for convex minimization problems*, Math. Program., 129 (2011), pp. 225–253.
[33] A. Nemirovski, *Prox-method with rate of convergence $\mathcal{O}(1/t)$ for variational inequalities with Lipschitz continuous monotone operators and smooth convex-concave saddle point problems*, SIAM J. Optim., 15 (2004), pp. 229–251.
[34] Yu. Nesterov, *Efficiency of coordinate descent methods on huge-scale optimization problems*, SIAM J. Optim., 22 (2012), pp. 341–362.
[35] N. Nisan, T. Roughgarden, E. Tardos, and V. V. Vazirani, *Algorithmic Game Theory*, Cambridge University Press, New York, 2007.
[36] M. J. Osborne and A. Rubinstein, *A Course in Game Theory*, MIT Press, Cambridge, MA, 1994.
[37] B. T. Polyak, *Introduction to Optimization*, Optimization Software, New York, 1987.
[38] P. Ricktárik and M. Takáč, *Iteration complexity of randomized block-coordinate descent methods for minimizing a composite function*, Math. Program., 144 (2014), pp. 1–38.

[39] R. T. Rockafellar and R. J.-B. Wets, *Variational Analysis*, Springer-Verlag, Berlin, 1998.

[40] T. Roughgarden, *Stackelberg scheduling strategies*, SIAM J. Comput., 33 (2004), pp. 332–350.

[41] S. Sabach and S. Shtern, *A first order method for solving convex bilevel optimization problems*, SIAM J. Optim., 27 (2017), pp. 640–660.

[42] H. Scarf, *The approximation of fixed points of a continuous mapping*, SIAM J. Appl. Math., 15 (1967), pp. 1328–1343.

[43] G. Scutari, D. P. Palomar, F. Facchinei, and J.-S. Pang, *Convex optimization, game theory, and variational inequality theory*, IEEE Signal Processing Magazine, 27 (2010), pp. 35–49.

[44] G. Scutari, D. P. Palomar, F. Facchinei, and J.-S. Pang, *Monotone games for cognitive radio systems*, in Distributed Decision Making and Control, Springer, New York, pp. 83–112.

[45] S. Shalev-Shwartz and T. Zhang, *Stochastic dual coordinate ascent methods for regularized loss minimization*, J. Mach. Learn. Res., 14 (2013), pp. 567–599.

[46] U. V. Shanbhag, G. Infanger, and P. W. Glynn, *A complementarity framework for forward contracting under uncertainty*, Oper. Res., 59 (2011), pp. 810–834.

[47] M. V. Solodov, *An explicit descent method for bilevel convex optimization*, J. Convex Anal., 14 (2007), pp. 227–237.

[48] J. Wang, G. Scutari, and D. P. Palomar, *Robust MIMO cognitive radio via game theory*, IEEE Trans. Signal Process., 59 (2011), pp. 1183–1201.

[49] M. Wang and D. P. Bertsekas, *Incremental constraint projection methods for variational inequalities*, Math. Program., 150 (2015), pp. 321–363.

[50] H.-K. Xu, *Viscosity approximation methods for nonexpansive mappings*, J. Math. Anal. Appl., 298 (2004), pp. 279–291.

[51] H. Yin, U. V. Shanbhag, and P. G. Mehta, *Nash equilibrium problems with scaled congestion costs and shared constraints*, IEEE Trans. Automat. Control, 56 (2011), pp. 1702–1708.

[52] F. Yousefian, *Bilevel distributed optimization in directed networks*, in Proceedings of the American Control Conference, 2021, pp. 2230–2235, https://doi.org/10.23919/ACC50511.2021.9483429.

[53] F. Yousefian, A. Nedić, and U. V. Shanbhag, *Optimal robust smoothing extragradient algorithms for stochastic variational inequality problems*, in Proceedings of the 53rd IEEE Conference on Decision and Control, 2014, pp. 5831–5836.

[54] F. Yousefian, A. Nedić, and U. V. Shanbhag, *On smoothing, regularization, and averaging in stochastic approximation methods for stochastic variational inequality problems*, Math. Program., 165 (2017), pp. 391–431, https://doi.org/10.1007/s10107-017-1175-y.

[55] F. Yousefian, A. Nedić, and U. V. Shanbhag, *On stochastic mirror-prox algorithms for stochastic Cartesian variational inequalities: Randomized block coordinate and optimal averaging schemes*, Set-Valued Var. Anal., 26 (2018), pp. 789–819, https://doi.org/10.1007/s11228-018-0472-9.

[56] F. Yousefian, A. Nedić, and U. V. Shanbhag, *On stochastic and deterministic quasi-Newton methods for nonstrongly convex optimization: Asymptotic convergence and rate analysis*, SIAM J. Optim., 30 (2020), pp. 1144–1172, https://doi.org/10.1137/17M1152474.