An Incremental Gradient Method for Optimization Problems With Variational Inequality Constraints

Harshal D. Kaushik, Sepideh Samadi, and Farzad Yousefian, Member, IEEE

Abstract—We consider minimizing a sum of agent-specific non-differentiable merely convex functions over the solution set of a variational inequality (VI) problem in that each agent is associated with a local monotone mapping. This problem finds an application in computation of the best equilibrium in nonlinear complementarity problems arising in transportation networks. We develop an iteratively regularized incremental gradient method where at each iteration, agents communicate over a directed cycle graph to update their solution iterates using their local information about the objective and the mapping. The proposed method is single-timescale in the sense that it does not involve any excessive hard-to-project computation per iteration. We derive nonasymptotic agent-wise convergence rates for the suboptimality of the global objective function and infeasibility of the VI constraints measured by a suitably defined dual gap function. The proposed method appears to be the first fully iterative scheme equipped with iteration complexity that can address distributed optimization problems with VI constraints over cycle graphs.

Index Terms—Computational complexity, distributed algorithms, mathematical programming, optimization methods.

I. INTRODUCTION

Consider a system with $m$ agents where the $i$th agent is associated with a component function $f_i : \mathbb{R}^n \to \mathbb{R}$ and a mapping $F_i : \mathbb{R}^n \to \mathbb{R}^n$. We consider a new distributed optimization framework of the form

$$\begin{align*}
\text{minimize} & \quad \sum_{i=1}^{m} f_i(x) \\
\text{subject to} & \quad x \in \text{SOL}(X, \sum_{i=1}^{m} F_i) \quad (P)
\end{align*}$$

where $X \subseteq \mathbb{R}^n$ is a set and SOL$(X, \sum_{i=1}^{m} F_i)$ denotes the solution set of the variational inequality $V(X, \sum_{i=1}^{m} F_i)$ defined as follows. A vector $x \in X$ solves $V(X, \sum_{i=1}^{m} F_i)$ if $(y - x)^T \sum_{i=1}^{m} F_i(x) \geq 0$ for all $y \in X$. Problem $(P)$ represents a distributed optimization framework in the sense that the information about $f_i$ and $F_i$ is locally known to the $i$th agent, while set $X$ is globally known. Model $(P)$ captures the canonical formulation of distributed optimization

$$\begin{align*}
\text{minimize} & \quad \sum_{i=1}^{m} f_i(x) \\
\text{subject to} & \quad x \in X
\end{align*}$$

that has been extensively studied. Indeed, by choosing $F_i(x) := 0$, $(P)$ is equivalent to model $(1)$. Motivating example: Nonlinear complementarity problems (NCP) have been employed to formulate diverse applications in engineering and economics. The celebrated Wardrop’s principle of equilibrium in traffic networks and also, the Walras’s law of competitive equilibrium in economics are among important examples that can be represented using NCP [1]. Formally, NCP is defined as follows. Given a mapping $F : \mathbb{R}_+^n \to \mathbb{R}^n$, $x \in \mathbb{R}^n$ solves NCP $F$ if $0 \leq x \perp F(x)$, where $\perp$ denotes the perpendicularity operator between two vectors. It is known that NCP $(F)$ can be cast as VI$(\mathbb{R}_+^n, F)$ (see [2, Proposition 1.1.3]). In many applications where $F$ is merely monotone, NCP $(F)$ may admit multiple equilibria. In such cases, one may consider finding the best equilibrium with respect to a global metric $f : \mathbb{R}^n \to \mathbb{R}$. For example, in traffic networks, the total travel time of the network users can be considered as the objective $f$. In fact, the problem of computing the best equilibrium of an NCP is important to be addressed particularly in the design of transportation networks where there is a need to estimate the efficiency of the equilibrium [3, 4]. In this regime, the goal is to minimize $f(x)$, where $x \in \text{SOL}(F)$. Consider a stochastic NCP associated with the mapping $\mathbb{E}[F(\bullet, \xi(\omega))]$, where $\xi : \Omega \to \mathbb{R}^d$ is a random variable associated with the probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and $F : \mathbb{R}^n_+ \times \Omega \to \mathbb{R}^n$ is a stochastic real-valued mapping. Let $S_i$ denote a local index set of independent and identically distributed samples from the random variable $\xi$. Employing a sample average approximation scheme, one can consider a distributed NCP given by $x \geq 0$, $\sum_{i=1}^{n} \sum_{\ell \in S_i} F(x, \xi_{\ell}) \geq 0$, $x^T \sum_{\ell \in S_i} F(x, \xi_{\ell}) = 0$. Let $f : \mathbb{R}^n_+ \times \Omega \to \mathbb{R}^n$ denote a stochastic objective function that measures the performance of a given equilibrium at a realization of $\xi$. Then, the problem of distributed computation of the best equilibrium of the preceding NCP is formulated as

$$\begin{align*}
\text{minimize} & \quad \sum_{i=1}^{m} \sum_{\ell \in S_i} f(x, \xi_{\ell}) \\
\text{subject to} & \quad x \in SOL \left( \mathbb{R}^n_+ \cup \sum_{i=1}^{m} \sum_{\ell \in S_i} F(\bullet, \xi_{\ell}) \right)
\end{align*}$$

Model $(P)$ captures problem $(2)$ by defining $X \triangleq \mathbb{R}^n_+$, $f_i(x) \triangleq \sum_{\ell \in S_i} f(x, \xi_{\ell})$, and $F_i(x) \triangleq \sum_{\ell \in S_i} F(x, \xi_{\ell})$. Scope and literature review: In addressing the proposed formulation $(P)$, our focus in this work lies in the development of an incremental gradient (IG) method. IG methods are among popular avenues for addressing the classical model $(1)$ [5, 6, 7, 8]. In these schemes, utilizing the additive structure of the problem, the algorithm cycles through the data blocks and updates the local estimates of the optimal solution in a sequential manner [9, 10]. In addressing the constrained problems

Manuscript received 11 December 2022; accepted 18 February 2023.

Color versions of one or more figures in this article are available at https://doi.org/10.1109/TAC.2023.3251851.

Digital Object Identifier 10.1109/TAC.2023.3251851
with easy-to-project constraint sets, the projected incremental gradient (P-IG) method and its subgradient variant were developed [11]. The P-IG scheme is described as follows. Given an initial point \(x_{0, 1} \in X\), where \(X \subseteq \mathbb{R}^{n}\) denotes the constraint set, for each \(k \geq 0\), consider the update rules given by

\[
x_{k+1, i} := \Pi_X (x_{k,i} - \gamma_k \nabla f_i (x_{k,i})), \quad \text{for } i \in [m]
\]

\[
x_{k+1, 1} := x_{k, m+1}
\]

where \(x_{k,i} \in \mathbb{R}^{n}\) denotes agent \(i\)'s local copy of the decision variables at iteration \(k\), \(\Pi\) denotes the Euclidean projection operator defined as \(\Pi_X (x) \triangleq \text{argmin}_{x \in X} \|x - z\|_2\), \(\gamma_k > 0\) denotes the stepsize, and \([m] \triangleq \{1, \ldots, m\}\). Recently, under strong convexity and twice continuous differentiability, and also, boundedness of the generated iterates, the standard IG method was proved to converge with the rate \(O(1/k)\) in the unconstrained case [8]. Accelerated variants of IG schemes with provable convergence speeds were also developed, including the incremental aggregated gradient method [7], [12], SAG [13], and SAGA [14]. Most of the past research efforts on the design and analysis of IG methods for constrained optimization problems have focused on addressing easy-to-project sets or sets with linear functional inequalities. This has been done through employing duality theory, projection, or penalty methods (see [15], [16], [17], [18], [19]).

Research gap: Despite the extensive work in the area of constrained optimization, no provably convergent single-timescale method exists in the literature that can be employed to solve distributed optimization problems with VI constraints. This is mainly because, unlike in standard constrained optimization, the Lagrangian duality theory does not appear to lend itself to be directly employed for addressing VI constraints. The classical scheme for addressing optimization problems with VI constraints is a sequential regularization (SR) framework [2, Ch. 12], where the regularized problem \(V^i(x, F + \eta_k \nabla f)\) is solved at every iteration \(k\) of the scheme, while \(\eta_k\) is updated in the outer iteration and reduced to zero. A key drawback of the SR scheme is that its iteration complexity is unknown and the scheme often convergences very slowly in practice. Moreover, the asymptotic convergence of this scheme is only established when \(f\) is strongly convex and smooth.

Contributions:

1) Complexity guarantees for addressing model (P): We develop an IG method equipped with agent-specific iteration complexity guarantees for solving distributed optimization problems with VI constraints of the form (P). To this end, employing a regularization-based relaxation technique, we propose a projected averaging iteratively regularized incremental gradient method (pair-IG) presented by Algorithm 1. In Theorem 1, under merely convexity of the global objective function and merely monotonicity of the global mapping, we derive new nonasymptotic suboptimality and infeasibility convergence rates for each agent’s generated iterates. This implies a total iteration complexity of \(O((C_f + C_p)^k \epsilon^{-4})\) for obtaining an \(\epsilon\)-approximate solution, where \(C_f\) and \(C_p\) denote the bounds on the global objective function’s subgradients and the global mapping over a compact convex set \(X\), respectively. Iterative regularization (IR) has been recently employed as a constraint-relaxation technique in a class of bilevel optimization problems [20], [21] and also in regimes where the duality theory may not be directly applied [22], [23]. Of these, in our recent work [23], we employed the IR technique to derive a provably convergent method for solving problem (P) in a centralized framework, where the information of the objective function is globally known by the agents. Unlike in [23], here we assume that the agents have access only to local information about both the objective function and the mapping.

2) Distributed averaging scheme: In pair-IG, we employ a distributed averaging scheme where agents can choose their initial averaged iterate arbitrarily and independent from each other. This relaxation in the proposed IG method appears to be novel, even for the classical IG schemes in addressing (1).

Notation: Throughout, a vector \(x \in \mathbb{R}^{n}\) is assumed to be a column vector and \(x^T\) denotes its transpose. We use \(\|x\|\) to denote the \(\ell_2\)-norm. For a convex function \(f: \mathbb{R}^{n} \rightarrow \mathbb{R}\), \(x \in \text{dom}(f)\), and any \(y \in \text{dom}(f)\), a vector \(\nabla f(x) \in \mathbb{R}^{n}\) is called a subgradient of \(f\) at \(x\) if \(f(x) + \nabla f(x)^T (y - x) \leq f(y)\) for all \(y \in \text{dom}(f)\). We let \(\partial f(x)\) denote the subdifferential set of function \(f\) at \(x\). The Euclidean projection of vector \(x\) onto set \(X\) is denoted by \(\Pi_X (x)\). We let \(\mathbb{R}^{n}_+\) denote the set \(\{x \in \mathbb{R}^n \mid x \geq 0\}\) and \(\{x \in \mathbb{R}^n \mid x > 0\}\), respectively. Given a set \(S \subseteq \mathbb{R}^{n}\), we let \(\text{int}(S)\) denote the interior of \(S\).

II. ALGORITHM OUTLINE

In this section, we present the main assumptions on problem (P), the outline of the proposed algorithm, and a few preliminary results that will be applied later in the rate analysis. Throughout, we let \(f(x) \triangleq \sum_{i=1}^{m} f_i(x) + F(x)\) denote the global objective and the global mapping, respectively.

Assumption 1 (Properties of problem (P)):

1) Function \(f_i : \mathbb{R}^{n} \rightarrow \mathbb{R}\) is real-valued and merely convex (possibly nondifferentiable) on its domain for all \(i \in [m]\).

2) Mapping \(F_i : \mathbb{R}^{n} \rightarrow \mathbb{R}\) is real-valued, continuous, and merely monotone on its domain for all \(i \in [m]\).

3) Set \(X \subseteq \text{dom}(f) \cap \text{dom}(F)\) is nonempty, convex, and compact.

Remark 1: Assumption 1 immediately implies the following. From [2, Th. 2.3.5 and Corollary 2.2.5], SOL(X, F) is convex, and compact. For all \(i\), the nonemptyness of the subdifferential \(\partial f_i(x)\) for any \(x \in \text{dom}(f_i)\) is implied from [24, Th. 3.14]. Also, [24, Th. 3.16] implies that \(f_i\) has bounded subgradients over the compact set \(X\). Further, mapping \(F_i\) is bounded over the set \(X\).

In view of compactness of the set \(X\) and continuity of \(f\), throughout this article, we let positive scalars \(M_X < \infty\) and \(M_F < \infty\) be defined as \(M_X \triangleq \sup_{x \in X} \|x\|\) and \(M_F \triangleq \sup_{x \in X} f(x)\), respectively. We also let \(f, F : \mathbb{R} \rightarrow \mathbb{R}\) denote the optimal objective value of problem (P). In view of Remark 1, throughout we let scalars \(C_F > 0\) and \(C_f > 0\) be defined such that for all \(i \in [m]\) and for all \(x \in X\), we have \(\|F_i(x)\| \leq \frac{C_F}{\sqrt{m}}\) and \(\|\nabla f_i(x)\| \leq \frac{C_f}{\sqrt{m}}\) for all \(\nabla f_i(x) \in \partial f_i(x)\). In the following, we comment on the Lipschitz continuity of the local and global objective functions.

Remark 2: Under Assumption 1 and from [24, Th. 3.61], function \(f_i\) is Lipschitz continuous with parameter \(\frac{C_f}{\sqrt{m}}\) over the set \(X\), i.e., for all \(i \in [m]\), we have \(|f_i(x) - f_i(y)| \leq \frac{C_f}{\sqrt{m}} \|x - y\|\) for all \(x, y \in X\).

We also have \(\|\nabla f_i(x)\| \leq C_f\) for all \(x \in X\) and all \(\nabla f_i(x) \in \partial f_i(x)\). This implies that \(\|f(x) - f(y)\| \leq C_f \|x - y\|\) for all \(x, y \in X\).

We now present an overview of the proposed method given by Algorithm 2. We use vector \(x_{k, 1}\) to denote the local copy of the global decision vector maintained by agent \(i\) at iteration \(k\). At each iteration, agents update their iterates in a cyclic manner. Each agent \(i \in [m]\) uses only its local information including the subgradient of function \(f_i\) and mapping \(F_i\) and evaluates the regularized mapping \(F_i + \eta_k \nabla f_i\) at \(x_{k, i}\). Here, \(\eta_k\) denotes the stepsize and the regularization parameter at iteration \(k\), respectively. Importantly, through employing an IR technique, we let both of these parameters be updated iteratively at suitable prescribed rates (cf., Theorem 1). Each agent computes and returns a weighted averaging iterate denoted by \(\bar{x}_{k, i}\), where the weights are characterized in terms of the stepsize \(\gamma_k\) and an arbitrary scalar \(r \in [0, 1]\). Notably, this averaging technique is carried out in a distributed fashion in the sense that agents do not require to start from the same initialized averaging iterate.
Algorithm 1: Projected Averaging Iteratively Regularized Incremental subGradient (pair-IG).

**input:** Agent $i$ arbitrarily chooses an initial vector $x_{0,i} \in X$. Agent $i$ arbitrarily chooses $x_{0,i} \in X$, for all $i \in [m]$. Let $S_0 := \gamma_0$ with an arbitrary $0 \leq r < 1$.

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

Update $S_{k+1} := S_k + \gamma_{k+1}$

**for** $i = 1, \ldots, m$ **do**

$$x_{k,i+1} := \mathcal{P}_X \left( x_{k,i} - \gamma_k \left( F_i(x_{k,i}) + \eta_k \nabla f_i(x_{k,i}) \right) \right)$$

$$\bar{x}_{k+1,i} := \left( \frac{S_k}{\gamma} \right) \bar{x}_{k,i} + \left( \frac{\gamma_{k+1}}{\gamma} \right) x_{k,i+1}$$

**end for**

Set $x_{k+1,i} := x_{k,m+1}$

**end for**

**return:** $x_{N,i}$ for all $i \in [m]$

Next, we show that for any $i \in [m]$, $\bar{x}_{N,i}$ is indeed a well-defined weighted average of $\bar{x}_{k,i}$, and the iterates $x_{k-1,i+1}$ for $k \in [N]$.

**Lemma 1:** Consider the sequence $\{x_{k,i}\}$ generated by agent $i \in [m]$ in Algorithm 1. For $k \in \{0, \ldots, N\}$, let us define the weights $\lambda_{k,i} \triangleq \frac{\sum_{j=0}^{k} \gamma_j}{\sum_{j=0}^{N} \gamma_j}$. Then, for all $i \in [m]$, we have

$$\bar{x}_{N,i} = \lambda_{0,N} \bar{x}_{0,i} + \sum_{k=1}^{N} \lambda_{k,N} x_{k-1,i+1}.$$ 

Further, for a convex set $X$, we have $\bar{x}_{N,i} \in X$.

**Proof:** We use induction on $N \geq 0$ to show the equation. For $N = 0$, from $\lambda_{0,0} = 1$, we have $\bar{x}_{0,i} = \lambda_{0,0} \bar{x}_{0,i}$. Now, assume that the equation holds for some $N \geq 0$. This implies

$$\bar{x}_{N,i} = \lambda_{0,N} \bar{x}_{0,i} + \sum_{k=1}^{N} \lambda_{k,N} x_{k-1,i+1}$$

$$= (\gamma_0 \bar{x}_{0,i} + \sum_{k=1}^{N} \gamma_k x_{k-1,i+1}) / \sum_{j=0}^{N} \gamma_j.$$ 

(5)

Using (5), we now show that the hypothesis statement holds for any $N + 1$. From (4), we have $\bar{x}_{N+1,i} = \left( \frac{S_N}{\gamma} \right) \bar{x}_{N,i} + \left( \frac{\gamma_{N+1}}{\gamma} \right) x_{N,i+1}$. Note that from (4) in Algorithm 1, we have $S_k = \sum_{j=0}^{k} \gamma_j$ for all $k \geq 0$. From this and using (5), we obtain

$$\bar{x}_{N+1,i} = \left( \frac{\sum_{k=0}^{N} \gamma_k}{\sum_{j=0}^{N} \gamma_j} \right) \bar{x}_{N,i} + \left( \frac{\gamma_{N+1}}{\sum_{j=0}^{N} \gamma_j} \right) x_{N,i+1}$$

$$= \gamma_0 \bar{x}_{0,i} + \sum_{k=1}^{N+1} \gamma_k x_{k-1,i+1} / \sum_{j=0}^{N+1} \gamma_j.$$ 

From the definition of $\lambda_{k,i}$, we conclude that the hypothesis holds for $N + 1$ and thus, the result holds for all $N \geq 0$. To show the second part, note that from the initialization in Algorithm 1 and the projection in (3), we have $\bar{x}_{0,i}, x_{k-1,i+1} \in X$ for all $i$ and $k \geq 1$. From the first part, $\bar{x}_{N,i}$ is a convex combination of $\bar{x}_{0,i}, x_{0,i+1}, \ldots, x_{N-1,i+1}$. Therefore, $\bar{x}_{N,i} \in X$ using the convexity of the set $X$.

For the ease of presentation throughout the analysis, we define a sequence $\{x_k\}$ as follows.

**Definition 1:** Consider Algorithm 1. Let $\{x_k\}$ be defined as $x_k \triangleq x_{k-1,m+1} = x_{k,1}$, for all $k \geq 1$, with $x_0 \triangleq x_{0,1}$.

In the following result, we characterize the distance between the local variable of any arbitrary agent with that of the first and the last agent at any given iteration.

**Lemma 2:** Consider Algorithm 1. Let Assumption 1 hold. Then, the following inequalities hold for all $i \in [m]$ and $k \geq 0$:

a) $\|x_k - x_{k,i}\| \leq (\frac{1}{\gamma}) \eta_k (C_P + \eta_k C_J).

b) $\|x_{k+1,i} - x_{k,i}\| \leq (\frac{1}{\gamma}) \eta_k (C_P + \eta_k C_J).$

**Proof:**

(a) Let $k \geq 0$ be an arbitrary integer. We use induction on $i$ to show this result. From Definition 1, for $i = 1$ and $k \geq 0$, we have $\|x_k - x_{k,i}\| = 0$, implying that the result holds for $i = 1$. Now suppose the hypothesis statement holds for some $i \in [m]$. We have

$$\|x_k - x_{k,i+1}\|$$

$$\leq \|x_k - x_{k,i}\| + \|x_{k,i} - x_{k,i+1}\|$$

$$\leq (\frac{1}{\gamma}) \eta_k (C_P + \eta_k C_J).$$

From (3), the hypothesis statement, and the nonexpansivity property of the projection, we obtain

$$\|x_k - x_{k,i+1}\|$$

$$\leq \|x_k - x_{k,i}\| + \|x_{k,i} - x_{k,i+1}\|$$

$$\leq (\frac{1}{\gamma}) \eta_k (C_P + \eta_k C_J).$$

This completes the proof of part (b).

We note that the generated agent-wise iterates $\bar{x}_{k,i}$ in Algorithm 1, as the scheme proceeds, may not be solutions to VI $(X,F)$ and so, they may not necessarily be feasible to problem (P). To quantify the infeasibility of these iterates, we employ a dual gap function (cf., [2, Ch. 1]) defined as follows.

**Definition 2:** Consider a closed convex set $X \subseteq \mathbb{R}^n$ and a continuous mapping $F : X \rightarrow \mathbb{R}^n$. The dual gap function at $x \in X$ is defined as $\text{GAP}(x) \triangleq \sup_{y \in X} F(y)^T (x - y)$.

Note that under Assumption 1, we have $\text{GAP}(x) \geq 0$. Further, it is known that when the mapping $F$ is continuous and monotone, $x \in \text{SOL}(X,F)$ if and only if $\text{GAP}(x) = 0$ (cf., [25]). The following result will be utilized in the rate analysis.

**Lemma 3 ([23, Lemma 2.14]):** Let $\beta \in [0, 1)$, $\Gamma \geq 1$, and $K$ be an integer. Then, for all $K \geq (1 - \sqrt{2 - 1})^{-1}$, we have

$$\text{GAP}(x) \leq (K+1)^{1-\beta} \leq \sum_{k=0}^{K} (k+1)^{-\beta}.$$ 

**III. RATE AND COMPLEXITY ANALYSIS**

In this section, we present the convergence and rate analysis of the proposed method under Assumption 1. After obtaining a preliminary inequality in Lemma 4 in terms of the sequence generated by the last agent, in Lemma 5, we derive inequalities that relate the global objective and the dual gap function at the iterate of other agents with those of the last agent. Utilizing these results, in Proposition 1, we obtain agent-specific bounds on the objective function value and the...
Algorithm $C - \left(\frac{\eta_k}{m} \right) \leq \left\| x_k - y \right\|^2 - \gamma_k^2 \left\| x_{k+1} - y \right\|^2 + \left[ C_F + \eta_k C_f \right]^2 \right)

\text{Proof:} \ Let y \in X be an arbitrary vector and $k \geq 0$ be fixed. From the update rule (3), for $i \in [m]$, we have

$$
\left\| x_{k,i+1} - y \right\|^2 = \left\| p_x \left( x_{k,i} - \gamma_k \left( F_i(x_{k,i}) + \eta_k \nabla f_i(x_{k,i}) \right) \right) - p_x(y) \right\|^2.
$$

Employing the nonexpansivity of the projection, we have

$$
\left\| x_{k,i+1} - y \right\|^2 \leq \left\| x_{k,i} - \gamma_k \left( F_i(x_{k,i}) + \eta_k \nabla f_i(x_{k,i}) \right) - y \right\|^2
$$

$$
= \left\| x_{k,i} - y \right\|^2 + \gamma_k^2 \left\| F_i(x_{k,i}) + \eta_k \nabla f_i(x_{k,i}) \right\|^2 - 2 \gamma_k \left( F_i(x_{k,i}) + \eta_k \nabla f_i(x_{k,i}) \right)^T \left( x_{k,i} - y \right).
$$

From the triangle inequality and recalling the bounds on $\nabla f_i(x)$ and $F_i(x)$, we obtain

$$
\left\| x_{k,i+1} - y \right\|^2 \leq \left\| x_{k,i} - y \right\|^2 + \gamma_k^2 \left( \frac{C_F + \eta_k C_f}{m} \right)^2
$$

$$
+ 2 \gamma_k \left( F_i(x_{k,i}) + \eta_k \nabla f_i(x_{k,i}) \right)^T \left( y - x_{k,i} \right).
$$

(7)

The last term in the preceding relation is bounded as follows:

$$
2 \gamma_k \left( F_i(x_{k,i}) + \eta_k \nabla f_i(x_{k,i}) \right)^T \left( y - x_{k,i} \right)
$$

$$
= 2 \gamma_k F_i(y)^T \left( y - x_{k,i} \right) + 2 \gamma_k \eta_k \nabla f_i(x_{k,i})^T \left( y - x_{k,i} \right)
$$

$$
\leq 2 \gamma_k \left( F_i(y)^T \left( y - x_{k,i} \right) + 2 \gamma_k \eta_k \left( f_i(y) - f_i(x_{k,i}) \right) \right)
$$

where the last inequality is implied by the monotonicity of $F_i$ and convexity of $f_i$. Combining with (7), we have

$$
\left\| x_{k,i+1} - y \right\|^2 \leq \left\| x_{k,i} - y \right\|^2 + \gamma_k^2 \left( \frac{C_F + \eta_k C_f}{m} \right)^2
$$

$$
+ 2 \gamma_k \left( F_i(y)^T \left( y - x_{k,i} \right) + 2 \gamma_k \eta_k \left( f_i(y) - f_i(x_{k,i}) \right) \right).
$$

Adding and subtracting $2 \gamma_k F_i(y)^T x_k + 2 \gamma_k \eta_k f_i(x_k)$, we get

$$
\left\| x_{k,i+1} - y \right\|^2 \leq \left\| x_{k,i} - y \right\|^2 + \gamma_k^2 \left( \frac{C_F + \eta_k C_f}{m} \right)^2
$$

$$
+ 2 \gamma_k \left( F_i(y)^T \left( x_k - x_{k,i} \right) + 2 \gamma_k \eta_k \left( f_i(y) - f_i(x_{k,i}) \right) \right)
$$

Using the Cauchy–Schwarz inequality and Remark 2, we obtain

$$
\left\| x_{k,i+1} - y \right\|^2 \leq \left\| x_{k,i} - y \right\|^2 + \gamma_k^2 \left( \frac{C_F + \eta_k C_f}{m} \right)^2
$$

$$
+ 2 \gamma_k \left( F_i(y)^T \left( x_k - x_{k,i} \right) + 2 \gamma_k \eta_k \left( f_i(y) - f_i(x_{k,i}) \right) \right)
$$

$$
+ 2 \gamma_k \left( \frac{C_F}{m} \left\| x_k - x_{k,i} \right\| + \frac{\eta_k C_f}{m} \left\| x_k - x_{k,i} \right\| \right).
$$

Summing over $i \in [m]$ and considering Definition 1, we have

$$
\left\| x_{k+1} - y \right\|^2 \leq \left\| x_k - y \right\|^2 + \frac{\gamma_k^2 (C_F + \eta_k C_f)^2}{m}
$$

$$
+ 2 \gamma_k \left( F(y)^T \left( y - x_k \right) + 2 \gamma_k \eta_k \left( f(y) - f(x_k) \right) \right)
$$

$$
+ \frac{\gamma_k^2 (C_F + \eta_k C_f)}{m} \sum_{i=1}^m \left\| x_k - x_{k,i} \right\|.
$$

Invoking Lemma 2 and then, multiplying both sides by $\gamma_k^{-1}$, we can obtain the desired result.

In the next result, we provide inequalities that relate the objective function and the dual gap function at the generated averaged iterate of the last agent with that of any other agent, respectively. This result will be utilized in Proposition 1.

Lemma 5: Consider problem (P) and the sequences $\{x_{N,i}\}$ generated in Algorithm 1 for $i \in [m]$ for some $N \geq 1$. Let Assumption 1 hold and let $\{\gamma_k\}$ and $\{\eta_k\}$ be strictly positive and nonincreasing sequences. Then, for any $i \in [m]$, we have

$$
f(x_{N,i}) - f(x_{N,m}) \leq C_F \lambda_{0,N} \left\| x_{0,i} - x_{0,m} \right\|
$$

$$
+ \frac{(m-1)C_F}{m} \sum_{k=0}^N \lambda_{k,N} \gamma_k
$$

$$
\text{GAP} (x_{N,i}) - \text{GAP}(x_{N,m}) \leq C_F \lambda_{0,N} \left\| x_{0,i} - x_{0,m} \right\|
$$

$$
+ \frac{(m-1)C_F}{m} \sum_{k=0}^N \lambda_{k,N} \gamma_k
$$

where $\lambda_{k,N} \triangleq \sum_{j=0}^k \gamma_j$ for $k \in \{0, \ldots, N\}$.

Proof: Note that the results are trivial when $m = 1$. Throughout, we assume that $m \geq 2$. From the Lipschitz continuity of function $f$ from Remark 2 and invoking Lemma 1, we can write the following for all $i \in [m]$

$$
f(x_{N,i}) - f(x_{N,m}) \leq C_f \lambda_{0,N} \left\| x_{0,i} - x_{0,m} \right\|
$$

$$
+ C_f \sum_{k=1}^N \lambda_{k,N} \left\| x_{k-1,i} - x_{k-1,m+1} \right\|
$$

(9)

Next, using Lemma 2(b) for any $k \geq 1$ and $i \in [m]$, we have

$$
\left\| x_{k-1,i} - x_{k-1,m+1} \right\| \leq \frac{(m-1)\gamma_k \left( C_F + \eta_k C_f \right)}{m}
$$

(10)

From (9), (10), and the nonincreasing sequence $\{\eta_k\}$, we have

$$
f(x_{N,i}) - f(x_{N,m}) \leq \frac{(m-1)C_F}{m} \sum_{k=1}^N \lambda_{k,N} \gamma_k
$$

$$
+ C_f \lambda_{0,N} \left\| x_{0,i} - x_{0,m} \right\|.
$$

Since $\{\gamma_k\}$ is nonincreasing and $0 \leq r < 1$, we obtain

$$
\sum_{k=1}^N \lambda_{k,N} \gamma_k \leq \frac{1}{\gamma_0} \sum_{j=0}^{r+1} \gamma_j \sum_{k=1}^N \gamma_k
$$

$$
\leq \frac{1}{\gamma_0} \sum_{j=0}^{r+1} \gamma_j \sum_{k=0}^N \lambda_{k,N} \gamma_k
$$

From the last two relations, we obtain (8a). Next, we show (8b). From Definition 2, we have

$$
\text{GAP}(x_{N,i}) = \sup_{y \in X} F(y)^T (x_{N,i} - y)
$$

$$
\leq \sup_{y \in X} F(y)^T (x_{N,i} - x_{N,m}) + \sup_{y \in X} F(y)^T (x_{N,m} - y)
$$

$$
\leq C_F \left\| x_{N,i} - x_{N,m} \right\| + \text{GAP}(x_{N,m})
$$

Authorized licensed use limited to the terms of the applicable license agreement with IEEE. Restrictions apply.
Rearranging the terms, we obtain \( \text{GAP}(\bar{x}_{N,i}) - \text{GAP}(\bar{x}_{N,m}) \leq C_F \|\bar{x}_{N,i} - \bar{x}_{N,m}\| \). The rest of the proof can be done in a similar fashion to the proof of (8a).

Next, we construct agent-wise error bounds in terms of the objective function value and the dual gap function at the averaged iterates generated in Algorithm 1.

**Proposition 1 (Agent-wise error bounds):** Consider problem (P) and the averaged sequence \( \{\bar{x}_{k,i}\} \) generated by agent \( i \) in Algorithm 1 for \( i \in [m] \). Let Assumption 1 hold and \( \{\gamma_k\} \) and \( \{\eta_k\} \) be nonincreasing and strictly positive sequences. Then, we have for \( i \in [m], N \geq 1, \) and \( r \in [0, 1] \):

\[
(a) f(\bar{x}_{N,i}) - f^* \leq \left( \sum_{k=0}^{N} \gamma_k \right)^{-1} \left( 2M_F^2 \gamma_k^{-1} + \frac{(C_F + \eta_k C_F)}{2} \right) \]

\[
+b_k \gamma_k^0 \left( f(x_{0,0}) - f(x_{0,1}) + C_F \gamma_k^0 \|x_{0,i} - x_{0,0}\| \right)
\]

\[
(b) \text{GAP}(\bar{x}_{N,i}) \leq \left( \sum_{k=0}^{N} \gamma_k \right)^{-1} \left( 2M_F^2 \gamma_k^{-1} + 2M_F \sum_{k=0}^{N} \gamma_k \eta_k \right) + \left( \sum_{k=0}^{N} \gamma_k \eta_k \right) \]

\[
+ \left( \frac{(C_F + N \gamma_k C_F)}{2} \right) \sum_{k=0}^{N} \gamma_k \eta_k \gamma_k^0 \|x_{0,i} - x_{0,0}\|.
\]

**Proof:** (a) Let \( x^* \in X \) denote an arbitrary optimal solution to problem (P). From feasibility of \( x^* \), we have \( F(x^*)^T(x_k - x^*) \geq 0 \). Substituting \( y \) by \( x^* \) in relation (6) and using the preceding relation, we have

\[
2 \gamma_k \eta_k \left( f(x_k) - f^* \right) \leq \gamma_k^{-1} \left( \|x_k - x^*\|^2 - \|x_{k+1} - x^*\|^2 \right)
\]

Adding and subtracting the term \( \gamma_k^{-1} \left( \gamma_k^{-1} - \gamma_{k+1} \right) \|x_{k+1} - x^*\|^2 \), we have

\[
\gamma_k \left( f(x_k) - f^* \right) \leq \gamma_k^{-1} \left( \|x_k - x^*\|^2 - \|x_{k+1} - x^*\|^2 \right)
\]

Adding (11) for \( k = 0 \) and then, adding and subtracting \( f(x_{0,0}), \) we have

\[
\gamma_0^* \left( f(x_{0,0}) - f^* + f(x_0) - f(x_{0,1}) \right) \leq \gamma_0^{-1} \left( \|x_{0,1} - x_{0,0}\|^2 \right)
\]

Adding the preceding equation with (14), we obtain

\[
\gamma_0^* \left( f(x_{0,0}) - f^* \right) + \sum_{k=0}^{N} \gamma_k^* \left( f(x_k) - f^* \right) \leq \gamma_0^{-1} \left( \|x_{0,1} - x_{0,0}\|^2 \right)
\]

Next, dividing both sides by \( \sum_{k=0}^{N} \gamma_k \), we have

\[
\gamma_0^* \left( f(x_{0,0}) + \sum_{k=0}^{N} \gamma_k f(x_k) - f(x_{0,1}) \right) \leq \gamma_0^{-1} \left( \sum_{k=0}^{N} \gamma_k \right)^{-1} \left( 2M_F^2 \gamma_k^{-1} \right)
\]

Adding (8a) with the preceding inequality, we obtain the desired result.

(b) From (6), for an arbitrary \( y \in X \), we have

\[
2 \gamma_k^0 \eta_k \left( f(y) - f(x_k) \right) + \gamma_k^0 \left( C_F + \eta_k C_F \right)^2.
\]

From the triangle inequality and definition of \( M_{F} \), we have \( |f(y) - f(x_k)| \leq 2M_{F} \). We obtain

\[
2 \gamma_k^0 \eta_k \left( f(y) - f(x_k) \right) + \gamma_k^0 \left( C_F + \eta_k C_F \right)^2.
\]

Adding and subtracting \( \gamma_k^{-1} \|x_k - y\|^2 \), we have

\[
2 \gamma_k^0 \eta_k \left( f(y) - f(x_k) \right) + \gamma_k^0 \left( C_F + \eta_k C_F \right)^2.
\]
\begin{equation}
+ 4\gamma_k \eta_k M_f + \left(\gamma_k - \gamma_k^{-1}\right) \|x_k - y\|^2 + \gamma_k^{+1} \left(C_F + \eta_0 C_f\right)^2.
\end{equation}

(16)

Using the nonincreasing property of \(\{\gamma_k\}\) and recalling \(0 \leq r < 1\), we have \(\gamma_k^{+1} \leq \gamma_k^{-1} \leq 0\). Thus, we can write, term 2 \(\leq \left(\gamma_k^{+1} - \gamma_k^{-1}\right)4M_k^2\). Taking summation over \(k \in [N]\) in (16) and dropping a nonpositive term, we obtain

\begin{equation}
2 \sum_{k=1}^{N} \gamma_k F(y)^T (x_k - y) \leq \gamma_0^{-1} \|x_1 - y\|^2 + 4M_f \sum_{k=1}^{N} \gamma_k \eta_k
+ 4M_k^2 \left(\gamma_k^{+1} - \gamma_k^{-1}\right) + \left(C_F + \eta_0 C_f\right)^2 \sum_{k=1}^{N} \gamma_k^{+1}.
\end{equation}

(17)

Writing (15) for \(k = 0\) and adding and subtracting \(2\gamma_0^2 F(y)^T x_{0,m}\), we have

\begin{align*}
2\gamma_0^2 F(y)^T (x_{0,m} - y + x_0 - x_{0,m}) &\leq 4\gamma_0 \eta_0 M_f \\
+ \gamma_0^{-1} \left(\|x_0 - y\|^2 - \|x_1 - y\|^2\right) + \gamma_0^{+1} \left(C_F + \eta_0 C_f\right)^2.
\end{align*}

Adding the preceding relation with (17), we have

\begin{align*}
2\gamma_0^2 F(y)^T (x_{0,m} - y) + 2 \sum_{k=1}^{N} \gamma_k F(y)^T (x_k - y)
\leq 4M_k^2 \left(\gamma_k^{+1} - \gamma_k^{-1}\right) + \left(C_F + \eta_0 C_f\right)^2 \sum_{k=1}^{N} \gamma_k^{+1} \\
+ \gamma_0^{-1} \|x_0 - y\|^2 + 4M_f \sum_{k=0}^{N} \gamma_k \eta_k + 2\gamma_0^2 F(y)^T (x_{0,m} - x_0).
\end{align*}

Using the Cauchy–Schwarz inequality, we have that term 3 \(\leq 2\gamma_0 C_F \|x_{0,m} - x_0.1\|\). We also have \(\|x_0 - y\|^2 \leq 4M_k^2\). Dividing the both sides of the preceding inequality by \(2 \sum_{k=0}^{N} \gamma_k\) and invoking Lemma 1, we have

\begin{equation}
F(y)^T (x_{0,m} - y) \leq \left(\sum_{k=0}^{N} \gamma_k\right)^{-1} \left(\frac{2M_k^2}{\gamma_k} \sum_{k=0}^{N} \gamma_k \eta_k + \left(C_F + \eta_0 C_f\right)^2 \sum_{k=0}^{N} \gamma_k^{+1} + \gamma_0 C_F \|x_{0,m} - x_0.1\|\right).
\end{equation}

(18)

(b) GAP \(\hat{\mathcal{E}}_N,i\) \(\leq \left(\sum_{k=0}^{N} \gamma_k\right)^{-1} \left(\frac{2M_k^2}{\gamma_k} \sum_{k=0}^{N} \gamma_k \eta_k + \left(C_F + \eta_0 C_f\right)^2 \sum_{k=0}^{N} \gamma_k^{+1} + \gamma_0 C_F \|x_{0,m} - x_0.1\|\right) + \gamma_0 C_F \|x_{0,m} - x_0.1\| + \gamma_0 C_F \|\|x_{0,i} - \bar{x}_{0,m}\| + \left(\gamma_0 C_F + \gamma_0 C_f\right)^2 \|\|x_{0,i} - \bar{x}_{0,m}\|\|.
\end{equation}

(19)

Proof: (a) Consider the inequality in Proposition 1(a). Substituting \(\gamma_k\) and \(\eta_k\) by their update rules, we obtain

\begin{align*}
f(\tilde{x}_{N,i}) - f^* &\leq \left(\sum_{k=0}^{N} \gamma_k\right)^{-1} \left(\frac{2M_k^2}{\gamma_k} \sum_{k=0}^{N} \gamma_k \eta_k + \left(C_F + \eta_0 C_f\right)^2 \sum_{k=0}^{N} \gamma_k^{+1} + \gamma_0 C_F \|x_{0,m} - x_0.1\|\right) + \gamma_0 C_F \|x_{0,m} - x_0.1\| + \left(\gamma_0 C_F + \gamma_0 C_f\right)^2 \|\|x_{0,i} - \bar{x}_{0,m}\|\|.
\end{align*}

In the next step, to apply Lemma 3, we need to ensure that the conditions in that result are met. From \(0 \leq r < 1\) and \(0 < b < 0.5\), we have \(0 \leq 0.5(1+r) - b < 1\), \(0 \leq 0.5r + b < 1\), and \(0 \leq 0.5(1+r) < 1\). Further, from \(\sum_{k=0}^{N} \gamma_k\), \(\sum_{k=0}^{N} \gamma_k^{+1}\), and \(\sum_{k=0}^{N} \gamma_k^{+1} + \gamma_0 C_F \|x_{0,m} - x_0.1\|\),

Therefore, all the necessary conditions of Lemma 3 are met

\begin{align*}
f(\tilde{x}_{N,i}) - f^* &\leq \left(\sum_{k=0}^{N} \gamma_k\right)^{-1} \left(\frac{2M_k^2}{\gamma_k} \sum_{k=0}^{N} \gamma_k \eta_k + \left(C_F + \eta_0 C_f\right)^2 \sum_{k=0}^{N} \gamma_k^{+1} + \gamma_0 C_F \|x_{0,m} - x_0.1\|\right) + \gamma_0 C_F \|x_{0,m} - x_0.1\| + \left(\gamma_0 C_F + \gamma_0 C_f\right)^2 \|\|x_{0,i} - \bar{x}_{0,m}\|\|.
\end{align*}

From the preceding relation, we obtain

\begin{align*}
f(\tilde{x}_{N,i}) - f^* &\leq \left(\gamma_0 C_F + \gamma_0 C_f\right)^2 \|\|x_{0,i} - \bar{x}_{0,m}\|\|.
\end{align*}

Factoring out \(1/(N + 1)^{0.5-b}\), we obtain

\begin{align*}
f(\tilde{x}_{N,i}) - f^* &\leq \left(\gamma_0 C_F + \gamma_0 C_f\right)^2 \|\|x_{0,i} - \bar{x}_{0,m}\|\|.
\end{align*}

Note that from \(b > 0\) and \(r < 1\), we have \(0.5 - 0.5r + b > 0\). Hence, (18) holds.

(b) Consider the inequality in Proposition 1(b). We have

\begin{align*}
\text{GAP}(x_{N,i}) &\leq \left(\sum_{k=0}^{N} \gamma_k\right)^{-1} \left(\frac{2M_k^2}{\gamma_k} \sum_{k=0}^{N} \gamma_k \eta_k + \left(C_F + \eta_0 C_f\right)^2 \sum_{k=0}^{N} \gamma_k^{+1} + \gamma_0 C_F \|x_{0,m} - x_0.1\| + \gamma_0 C_F \|\|x_{0,i} - \bar{x}_{0,m}\|\| + \left(\gamma_0 C_F + \gamma_0 C_f\right)^2 \|\|x_{0,i} - \bar{x}_{0,m}\|\|.
\end{align*}

(20)
Substituting \( \{ \gamma_k \} \) and \( \{ \eta_k \} \) by their update rules, we obtain
\[
\text{GAP}(\bar{x}_{N,i}) \leq \left( \frac{N}{(k+1)^{\frac{1}{2}}} \right)^{-1} \left( 2M^2 \gamma_0 \right)^{0.5(1-r)} + \sum_{k=0}^{N} \frac{2M^2 \gamma_0 \gamma_0^{r+1}}{(k+1)^{0.5(1-r)}}
\]
Utilizing the bounds in Lemma 3, we obtain
\[
\text{GAP}(\bar{x}_{N,i}) \leq \left( \frac{\gamma_0(N+1)^{1-0.5r}}{2(1-0.5r-b)(N+1)^{r}} \right)^{-1} \left( 2M^2 \gamma_0 \right)^{0.5(1-r)} + \sum_{k=0}^{N} \frac{2M^2 \gamma_0 \gamma_0^{r+1}}{(k+1)^{0.5(1-r)}}
\]
Rearranging the terms, we obtain
\[
\text{GAP}(\bar{x}_{N,i}) \leq \left( \frac{\gamma_0(N+1)^{1-0.5r}}{2(1-0.5r-b)(N+1)^{r}} \right)^{-1} \left( 2M^2 \gamma_0 \right)^{0.5(1-r)} + \sum_{k=0}^{N} \frac{2M^2 \gamma_0 \gamma_0^{r+1}}{(k+1)^{0.5(1-r)}}
\]
Note that from \( b < 0.5 \) and \( 0 \leq r < 1 \), we have \( 1 - 0.5r \geq b \). Hence, (19) holds.

**Remark 3 (Iteration complexity of Algorithm 1):** Consider the rate presented by relations (18) and (19). Let us choose \( r = 0 \) and suppose \( \gamma_k := \frac{(C^T \bar{Q} + \gamma_0)^{r}}{(N+1)^{r}} \) and \( \eta_k := \frac{1}{k+1} \) for \( k \geq 0 \).

Let \( \epsilon > 0 \) be an arbitrary small scalar such that \( f(\bar{x}_{N,i}) \to f^* + \text{GAP}(\bar{x}_{N,i}) < \epsilon \) for all \( i \in [m] \). Then, we obtain the iteration complexity of \( N = O((C^T \bar{Q} + \gamma_0)^{\epsilon^{-1}}) \) for each agent. Interestingly, this iteration complexity matches the complexity of the proposed method in the earlier work [23] for addressing formulation (P) in a centralized regime where the information of the objective function \( f \) is globally known. This indicates that there is no sacrifice in the iteration complexity in addressing the distributed formulation (P). Another observation to make is that the iteration complexity of the proposed distributed method is independent of the number of agents \( m \).

**IV. NUMERICAL RESULTS**

**A traffic equilibrium problem:** For an illustrative example, we consider the transportation network in [26]. We first describe the network and present the NCP formulation. Then, we implement Algorithm 1 to solve model (2) and compute the best equilibrium.

Consider a transportation network with the set of nodes \( \{ n_1, n_2 \} \) and the set of directed arcs \( \{ a_1, a_2, a_3, a_4, a_5 \} \), as shown in Fig. 1. Note that \( a_1 \) and \( a_2 \) construct a two-way road. The same holds for \( a_3 \) and \( a_5 \). We let \( d \triangleq [d_1, d_2]^T \) denote the expected travel demand vector, where \( d_1 \) and \( d_2 \) correspond to the demand from \( n_1 \) to \( n_2 \), and from \( n_2 \) to \( n_1 \), respectively. Let vector \( h \triangleq [h_1, \ldots, h_5]^T \) denote the traffic flow on the arcs. The travel cost on arc \( a_i \) is assumed to be \( [Ch + \tilde{q}]_i \).

where the cost matrix \( C \in \mathbb{R}^{5 \times 5} \) and vector \( q \in \mathbb{R}^5 \) be given by

\[
C := \begin{bmatrix}
0.92 & 0 & 0 & 5 & 0 \\
0 & 5.92 & 0 & 0 & 5 \\
0 & 0 & 10.92 & 0 & 0 \\
2 & 0 & 0 & 10.92 & 0 \\
0 & 1 & 0 & 0 & 15.92
\end{bmatrix}, \quad q := \begin{bmatrix}
1000 \\
950 \\
3000 \\
1000 \\
1300
\end{bmatrix}
\]

We note that matrix \( C \) is positive semidefinite. Intuitively speaking, the structure of \( C \) implies that the cost of each arc in a two-way road depends on the flows on the both directions. Let \( u \triangleq [u_1, u_2]^T \) denote the (unknown) vector of minimum travel costs between the origin-destination (OD) pairs, i.e., \( u_1 \) denotes the minimum travel cost from \( n_1 \) to \( n_2 \), and \( u_2 \) denotes the minimum travel cost from \( n_2 \) to \( n_1 \). Mathematically, the Wardrop’s principle can be characterized as

\[
0 \leq Ch + q - B^T u \leq h, \quad 0 \leq Bh - d \leq u \geq 0
\]

where \( B \in \mathbb{R}^{2 \times 5} \) denotes the (OD pair, arc)-incidence matrix given as

\[
B := \begin{bmatrix}
1 & 1 & 1 & 0 & 0 \\
0 & 0 & 0 & 1 & 1
\end{bmatrix}
\]

We assume that demand vector \( d \) and cost vector \( q \) are subject to uncertainties and define decision vector \( x \in \mathbb{R}^7 \), random variable \( \xi \in \mathbb{R}^7 \), and stochastic mapping \( F(\cdot, \xi) : \mathbb{R}^7 \to \mathbb{R}^7 \) as

\[
x \triangleq \begin{bmatrix}
\bar{h} \\
\bar{u}
\end{bmatrix}, \quad \xi \triangleq \begin{bmatrix}
\bar{d} \\
\bar{q}
\end{bmatrix}, \quad F(x, \xi) \triangleq \begin{bmatrix}
C - B^T & 0 \\
B & 0
\end{bmatrix} \begin{bmatrix}
\bar{h} \\
\bar{q}
\end{bmatrix} + \begin{bmatrix}
\bar{d} \\
\bar{q}
\end{bmatrix}
\]

Then, from Section I, the Wardrop (21) can be characterized as \( \text{VI}(\tilde{N}_u, E[F(\cdot, \xi)]) \). Notably, due to positive semidefinite property of \( C \), mapping \( E[F(\cdot, \xi)] \) is merely monotone. Consequently, the aforementioned VI may have multiple equilibria. Among them, we seek to find the best equilibrium with respect to a welfare function \( f \) defined as the expected total travel time over the network by all users, i.e., \( f(x) \triangleq \mathbb{E}[((Ch + \tilde{q})^T I_5)] \) where \( I_5 \) is the 1-vector of size 5.

**Setup:** For this experiment, we assume that \( d_1 \sim \mathcal{N}(210, 10), d_2 \sim \mathcal{N}(120, 10) \). Also for \( i = 1, \ldots, 5 \), we let \( \tilde{q}_i \) be normally distributed with the mean equal to \( q_i \) and the standard deviation of 300, where vector \( q \) is given by (2). Following formulation (2), we generate 1000 samples for each parameter and distribute the data equally among ten agents. We let \( \gamma_k := \frac{1}{(k+1)^{0.5}} \) and \( \eta_k := \frac{1}{(k+1)^{0.5}} \) and consider different values for the initial stepsize \( \gamma_0 \) and the initial regularization parameter \( \eta_0 \). The results are as shown in Fig. 2. We use standard averaging by assuming that \( r = 0 \). Notably, for quantifying the infeasibility, we consider metric \( \phi(x) \triangleq \max \{0, x - \min \{0, -x\}\}^2 + \max \{0, -F(x)\}^2 + |x^T F(x)| \), where \( F(x) = \sum_{i=1}^{m} F_i(x) \) and \( F_i(x) = \sum_{i=1}^{m} F(x, \xi_i) \). Note that \( \phi(x) = 0 \) if and only if \( 0 \leq x \leq F(x) \geq 0 \). We choose this metric over the dual gap function employed earlier in the analysis because in this particular example, the dual gap function becomes infinity at some of the evaluations of the generated iterates. This is due to the unboundedness of set \( X := \mathbb{R}^5_+ \). Unlike the dual gap function, \( \phi(x) \) stays bounded and is more suitable to plot.
Insights: In Fig. 2, we observe that in all four different settings the infeasibility metric decreases as the algorithm proceeds. This indeed implies that the generated iterates by the agents tend to satisfy the NCP constraints with an increasing accuracy. In terms of the suboptimality metric, we observe that each agent’s objective value becomes more and more stable over time. Intuitively, this implies that the agents asymptotically reach to an equilibrium. We should note that although function $f$ is minimized, it is minimized only over the set of equilibria. The fact that the objective values in Fig. 2 are not necessarily decreasing is mainly because of the impact of feasibility violation of the iterates with respect to the NCP constraints throughout the implementations.

ACKNOWLEDGMENT

An extended version of this work is available at https://arxiv.org/abs/2105.14205.

REFERENCES

[1] M. C. Ferris and J.-S. Pang, “Engineering and economic applications of complementarity problems,” SIAM Rev., vol. 39, no. 4, pp. 669–713, 1997.

[2] F. Facchinei and J.-S. Pang, Finite-Dimensional Variational Inequalities and Complementarity Problems (Springer Series in Operations Research). New York, NY, USA: Springer-Verlag, 2003.

[3] N. Nisan, T. Roughgarden, E. Tardos, and V. Vazirani, Algorithmic Game Theory. New York, NY, USA: Cambridge Univ. Press, 2007.

[4] R. Johari, “Efficiency loss in market mechanisms for resource allocation,” Ph.D. dissertation, Dept. Elect. Eng. Comput. Sci., Massachusetts Inst. Technol., Cambridge, MA, USA, 2004.

[5] A. Nedić, “Subgradient methods for convex minimization,” Ph.D. dissertation, Dept. Elect. Eng. Comput. Sci., Massachusetts Inst. Technol., Cambridge, MA, USA, 2002.

[6] M. Wang and D. P. Bertsekas, “Incremental constraint projection methods for variational inequalities,” Math. Program., vol. 150, no. 2, pp. 321–363, 2015.

[7] M. Gürbüzbalaban, A. Ozdaglar, and P. A. Parrilo, “On the convergence rate of incremental aggregated gradient algorithms,” SIAM J. Optim., vol. 27, no. 2, pp. 1035–1048, 2017.

[8] M. Gürbüzbalaban, A. Ozdaglar, and P. A. Parrilo, “Convergence rate of incremental gradient and incremental Newton methods,” SIAM J. Optim., vol. 29, no. 4, pp. 2542–2565, 2019.

[9] D. P. Bertsekas, Nonlinear Programming, 3rd ed. Bellmont, MA, USA: Athena Sci., 2016.

[10] D. P. Bertsekas, “Incremental gradient, subgradient, and proximal methods for convex optimization: A survey,” Optim. Mach. Learn., vol. 2010, no. 1-38, p. 3, 2011.

[11] A. Nedić and D. P. Bertsekas, “Incremental subgradient methods for non-differentiable optimization,” SIAM J. Optim., vol. 12, no. 1, pp. 109–138, 2001.

[12] D. Blatt, A. O. Hero, and H. Gauchman, “A convergent incremental gradient method with a constant step size,” SIAM J. Optim., vol. 18, no. 1, pp. 29–51, 2007.

[13] N. L. Roux, M. Schmidt, and F. Bach, “A stochastic gradient method with an exponential convergence rate for finite training sets,” in Proc. Adv. Neural Inf. Process. Syst., 2012, pp. 2663–2671.

[14] A. Defazio, F. Bach, and S. Lacoste-Julien, “SAGA: A fast incremental gradient method with support for non-strongly convex composite objectives,” in Proc. Adv. Neural Inf. Process. Syst., 2014, pp. 1646–1654.

[15] T.-H. Chang, A. Nedić, and A. Scaglione, “Distributed constrained optimization by consensus-based primal-dual perturbation method,” IEEE Trans. Autom. Control, vol. 59, no. 6, pp. 1524–1538, Jun. 2014.

[16] N. S. Aybat and E. Y. Hamedani, “A primal-dual method for conic constrained distributed optimization problems,” in Proc. Adv. Neural Inf. Process. Syst., 2016, pp. 5049–5057.

[17] G. Scutari and Y. Sun, “Distributed nonconvex constrained optimization over time-varying digraphs,” Math. Program., vol. 176, pp. 497–544, 2019.

[18] A. Nedić and T. Tatarenko, “Convergence rate of a penalty method for strongly convex problems with linear constraints,” in Proc. IEEE 59th Conf. Decis. Control, 2020, pp. 372–377.

[19] D. P. Bertsekas, “Incremental aggregated proximal and augmented Lagrangian algorithms,” Lab. Inf. Decision Syst., MIT, Tech. Rep. LIDS-P-2176, 2015.

[20] F. Yousefian, “Bilevel distributed optimization in directed networks,” in Proc. IEEE Amer. Control Conf., 2021, pp. 2230–2235.

[21] M. Amini and F. Yousefian, “An iterative regularized mirror descent method for ill-posed nondifferentiable stochastic optimization,” 2019, arXiv:1901.09506.

[22] F. Yousefian, A. Nedić, and U. V. Shanbhag, “On smoothing, regularization, and averaging in stochastic approximation methods for stochastic variational inequality problems,” Math. Program., vol. 165, no. 1, pp. 391–431, 2017.

[23] H. D. Kaushik and F. Yousefian, “A method with convergence rates for optimization problems with variational inequality constraints,” SIAM J. Optim., vol. 31, no. 3, pp. 2171–2198, 2021.