Control Theory and Technology (2024) 22:431–441
https://doi.org/10.1007/s11768-024-00223-9

RESEARCH ARTICLE

Distributed algorithm for solving variational inequalities over time-varying unbalanced digraphs

Yichen Zhang[1] · Yutao Tang[1] · Zhipeng Tu[2] · Yiguang Hong[3,4]

Received: 24 September 2023 / Revised: 12 December 2023 / Accepted: 13 December 2023 / Published online: 31 May 2024
© The Author(s), under exclusive licence to South China University of Technology and Academy of Mathematics and Systems Science, Chinese Academy of Sciences 2024

Abstract
In this paper, we study a distributed model to cooperatively compute variational inequalities over time-varying directed graphs. Here, each agent has access to a part of the full mapping and holds a local view of the global set constraint. By virtue of an auxiliary vector to compensate the graph imbalance, we propose a consensus-based distributed projection algorithm relying on local computation and communication at each agent. We show the convergence of this algorithm over uniformly jointly strongly connected unbalanced digraphs with nonidentical local constraints. We also provide a numerical example to illustrate the effectiveness of our algorithm.

Keywords Variational inequality · Distributed computation · Multi-agent system · Weight-unbalanced graph

1 Introduction

The variational inequality problem VI($\boldsymbol{K}$, $\boldsymbol{F}$) is to find a vector $\boldsymbol{x} \in \boldsymbol{K}$, such that

$$(\boldsymbol{y}-\boldsymbol{x})^{\mathrm{T}} \boldsymbol{F}(\boldsymbol{x}) \geq 0, \quad \forall\, \boldsymbol{y} \in \boldsymbol{K} \tag{1}$$

for a given mapping $\boldsymbol{F}\colon \boldsymbol{K} \to \mathbb{R}^N$ and set $\boldsymbol{K} \subset \mathbb{R}^N$. This problem provides us with a unified framework to study various kinds of practical problems, such as traffic equilibrium models, energy markets, and communication networks. Many important results including both the theoretical and algorithmic foundations have been delivered in [1, 2].

Classical algorithms for solving the variational inequality problem are centralized in the sense that all processing and storage operations of problem data are executed on a single computing unit. In the era of Big Data, the required amount of computations may be too large to be done by a centralized unit due to its limited capacity. Also, there are many scenarios where the problem data are distributed at widely dispersed locations, and the traditional parallel framework cannot be utilized. Thus, it is of particular importance to develop decentralized/distributed algorithms involving many networked computing units to solve the variational inequality problem.

Although pioneered attempts to cooperatively solve VI($\boldsymbol{K}$, $\boldsymbol{F}$) were made in the 1980s [3], this topic has gained renewed interests due to its application in many recent distributed computing problems. For example, as pointed out in [4], the intensively studied distributed optimization problem could be converted into some equivalent variational inequalities. Hence, different distributed optimization algorithms can be developed by extending classical variational inequality solvers to the distributed case. Although not often explicitly

This work was supported by the National Natural Science Foundation of China (No. 61973043) and Shanghai Municipal Science and Technology Major Project (No. 2021SHZDZX0100).

✉ Yutao Tang
yttang@bupt.edu.cn

Yichen Zhang
zhangyc930@163.com

Zhipeng Tu
tuzhipeng01@icloud.com

Yiguang Hong
yghong@iss.ac.cn

1 School of Artificial Intelligence, Beijing University of Posts and Telecommunications, Beijing 100876, China

2 Network Technology Laboratory, Huawei Technologies Co., Ltd., Beijing 100095, China

3 Department of Control Science and Engineering, Tongji University, Shanghai 200092, China

4 Shanghai Research Institute for Intelligent Autonomous Systems, Tongji University, Shanghai 200092, China

stated, the distributedness of the corresponding variational inequalities in these distributed optimization problems are from the separable structure of the mapping, where each agent is assumed to know only one component [5]. Then, the information-sharing graph becomes a crucial factor and has been extensively studied in designing efficient distributed algorithms. Following this line, [6] considered some Minty variational inequalities and developed decentralized algorithms by extending classical stochastic extragradient methods. They also showed the convergence of the proposed algorithms under time-varying undirected graphs. Additionally, [7] further incorporated compressed communication toward communication-efficient algorithms but for a special fixed star-like graph, which was later extended to any fixed connected undirected graph in [8].

Meanwhile, the distributed computation of VI($\boldsymbol{K}$, $\boldsymbol{F}$) may come from the partition of decision variables at different computing units, as shown in [3]. Recently, this computing model has found its applications in partial-decision information network games. For example, [9] related a non-cooperative game to some variational inequalities as that in [10] and then developed gossip-based algorithms to determine the Nash equilibrium of this game under connected undirected graphs. Typically, the variational inequalities thus obtained are non-separable with different local constraints. Hence, their solvability over general directed graphs can be more challenging than the preceding separable model from the perspective of distributed optimization. In fact, although many important algorithms have been developed for this type of variational inequality problem in the name of (generalized) Nash equilibrium seeking [11–14], most existing results are limited to fixed graphs except some very recent works [15, 16] considering time-varying undirected or weight-balanced directed graphs.

Motivated by these observations, we consider the variational inequality problem corresponding to multiple decision-makers (or agents). Here, each agent only knows a part of the full mapping and maintains a subset of the whole decision variables as well as nonidentical local set constraints. Moreover, we are interested in distributed algorithms over general time-varying directed graphs which can be weight-unbalanced. To solve these variational inequalities, we first decompose them into several coupled variational inequities of smaller dimensions. Then, we develop a consensus-based projection algorithm to solve them in a distributed way. By incorporating an auxiliary time-varying vector to compensate the graph imbalance, we show the convergence of our algorithm with different stepsizes under standard assumptions.

The contributions of this paper can be summarized as follows:

- We present a novel distributed computing model to solve the variational inequality problem. Different from existing formulations [6–8], the distributedness of our problem comes from the partition of decision variables rather than the summation structure of the corresponding mapping. We are not aware of any earlier general discussion on this model in distributed settings.
- We develop a novel distributed projection algorithm to solve the formulated variational inequality problem over uniformly jointly strongly connected interaction digraphs. Specially, the proposed algorithm can effectively solve the Nash equilibrium seeking problem considered in [14–17] without requiring the graph to be weight-balanced.

The rest of this paper is organized as follows: We first introduce some preliminaries in Sect. 2. Then, we give the formulation of our problem and present the main algorithm in Sect. 3. After that, the convergence analysis of our algorithm is provided in Sect. 4 under different stepsize conditions along with further discussions in Sect. 5. A simulation example and some concluding remarks are finally given in Sects. 6 and 7.

2 Preliminaries

This section introduces some preliminaries about convex analysis and graph theory for the following analysis.

2.1 Convex analysis

Let $\mathbb{R}^n$ be n-dimensional Euclidean space and $\mathbb{R}^{n\times m}$ be the set of all $n \times m$ matrices with all entries in $\mathbb{R}$. Denote the set containing all nonnegative integers by $\mathbb{Z}_+$. We always use bold math symbols to represent column vectors or vector-valued mappings. $\mathbf{0}_n$ (or $\mathbf{0}_{n\times m}$) represent the all-zero vector in $\mathbb{R}^n$ (or the all-zero matrix in $\mathbb{R}^{n\times m}$). We use $\mathcal{I}_n$ be the n-dimensional identity matrix. We may omit the subscript when it is self-evident. For a vector $\boldsymbol{x}$ (or a matrix $\boldsymbol{A}$), $\|\boldsymbol{x}\|$ (or $\|\boldsymbol{A}\|$) represents its Euclidean norm (or spectral norm). A square matrix W is said to be row-stochastic if it has nonnegative entries with its row summing to one. A matrix W is column-stochastic if its transpose is row-stochastic. W is said to be doubly stochastic if it is both row- and column-stochastic.

For a given set $\boldsymbol{K} \subset \mathbb{R}^n$, it is convex if, for any two points $\boldsymbol{x}$, $\boldsymbol{y} \in \boldsymbol{K}$, the line segment between them is still in $\boldsymbol{K}$. A function $f : \boldsymbol{K} \to \mathbb{R}$ is said to be convex on $\boldsymbol{K}$ if its epigraph $\operatorname{epi} f \triangleq \{(\boldsymbol{x}, t) | \boldsymbol{x} \in \boldsymbol{K}, t \in \mathbb{R}, t \geq f(x)\}$ is convex as a subset of $\mathbb{R}^{n+1}$. Define the distance between any point $\boldsymbol{x}$ and set $\boldsymbol{K}$ as $d_{\boldsymbol{K}}(\boldsymbol{x}) = \inf\{\|\boldsymbol{x} - \boldsymbol{y}\| \mid \boldsymbol{y} \in \boldsymbol{K}\}$. When $\boldsymbol{K}$ is closed and convex, it is well known that $d_{\boldsymbol{K}}(\boldsymbol{x}) = \|\boldsymbol{x} - \Pi_{\boldsymbol{K}}(\boldsymbol{x})\|$ with $\Pi_{\boldsymbol{K}}(\boldsymbol{x})$ the projection of $\boldsymbol{x} \in \mathbb{R}^n$ on $\boldsymbol{K}$.

Consider a vector-valued mapping $\boldsymbol{F}: \boldsymbol{K} \to \mathbb{R}^n$. We say it is Lipschitz continuous with $\ell > 0$ (or simply ℓ-Lipschitz) on $\boldsymbol{K}$ if $\|\boldsymbol{F}(\boldsymbol{x}) - \boldsymbol{F}(\boldsymbol{y})\| \leq \ell\|\boldsymbol{x} - \boldsymbol{y}\|$ for $\forall \boldsymbol{x}, \boldsymbol{y} \in \boldsymbol{K}$. The mapping is monotone on $\boldsymbol{K}$ if $[\boldsymbol{F}(\boldsymbol{x}) - \boldsymbol{F}(\boldsymbol{y})]^{\mathrm{T}}(\boldsymbol{x} - \boldsymbol{y}) \geq 0$ for $\forall \boldsymbol{x}, \boldsymbol{y} \in \boldsymbol{K}$. F is strictly monotone if $[\boldsymbol{F}(\boldsymbol{x}) - \boldsymbol{F}(\boldsymbol{y})]^{\mathrm{T}}(\boldsymbol{x} - \boldsymbol{y}) = 0$ only happens when $\boldsymbol{x} = \boldsymbol{y}$. If $\boldsymbol{F}$ is strictly monotone on $\boldsymbol{K}$, then $\mathrm{VI}(\boldsymbol{K}, \boldsymbol{F})$ has at most one solution. More details can be found in [2].

The following lemma is a direct application of Jensen's inequality and can also be found in the multi-agent literature, e.g., [18].

Lemma 1 *Let $\boldsymbol{K} \subset \mathbb{R}^n$ be a nonempty closed convex set and $a_1, \ldots, a_m$ be any scalars, such that $a_i \geq 0$ and $\sum_{i=1}^m a_i = 1$. Then, for any $\boldsymbol{x}^1, \ldots, \boldsymbol{x}^m \in \mathbb{R}^n$, $\|\sum_{i=1}^m a_i \boldsymbol{x}^i - \Pi_{\boldsymbol{K}}(\sum_{i=1}^m a_i \boldsymbol{x}^i)\| \leq \sum_{i=1}^m a_i \|\boldsymbol{x}^i - \Pi_{\boldsymbol{K}}(\boldsymbol{x}^i)\|$.*

2.2 Graph theory

A directed graph (digraph) is described by $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ with node set $\mathcal{V}$ and edge set $\mathcal{E} \subset \mathcal{V} \times \mathcal{V}$. The ordered pair $(i, j) \in \mathcal{E}$ is a directed edge of this digraph from node i to node j. A self-loop is an edge from any node to itself. A digraph is strongly connected if there exists a directed path between any two nodes. The composition of two digraphs $\mathcal{G}_p$ and $\mathcal{G}_q$ with the same node set is defined as $\mathcal{G}_p \circ \mathcal{G}_q = (\mathcal{V}, \mathcal{E})$ where $\mathcal{E} = \{(i, j) \mid \exists k \in \mathcal{V}$, such that $(i, k) \in \mathcal{E}_p, (k, j) \in \mathcal{E}_q\}$. The definition of graph composition can extend to any finite sequence of digraphs with the same node set. An infinite directed graph sequence $\{\mathcal{G}_k\}$ is said to be uniformly jointly strongly connected if there exists a positive integer B, such that the composed graph $\mathcal{G}_{k+B-1} \circ \cdots \mathcal{G}_{k+1} \circ \mathcal{G}_k$ is strongly connected for any $k \geq 0$.

3 Computing model and algorithm

Consider the variational inequality problem (1). Suppose that the set $\boldsymbol{K}$ is a Cartesian product $\boldsymbol{K} = K_1 \times \cdots \times K_n$ with each $K_i \subset \mathbb{R}^{N_i}$ a compact convex set and $\sum_{i=1}^n N_i = N$. Then, the mapping $\boldsymbol{F}$ and vector $\boldsymbol{x}$ can be accordingly decomposed as $\boldsymbol{F} = \mathrm{col}(F_1, \ldots, F_n)$ and $\boldsymbol{x} = \mathrm{col}(x_1, \ldots, x_n)$ with compatible dimensions.

The following assumption guarantees the well-posedness of $\mathrm{VI}(\boldsymbol{K}, \boldsymbol{F})$ [2].

Assumption 1 The mapping $\boldsymbol{F}$ is strictly monotone and l-Lipschitz for some $l > 0$.

Assume that we have n networked agents and each agent i only knows F_i and K_i. We are going to develop iterative rules for them to cooperatively solve (1). A key observation on our problem is that: A vector $\boldsymbol{x}^* = \mathrm{col}(x_1^*, \ldots, x_n^*)$ solves (1) if and only if for each $i \in \mathcal{N} \triangleq \{1, \ldots, n\}$, we have $x_i^* \in K_i$ and

$$(x_i - x_i^*)^{\mathrm{T}} F_i(\boldsymbol{x}^*) \geq 0, \quad \forall x_i \in K_i. \tag{2}$$

In this way, we can decompose the considered variational inequalities (1) into n coupled variational inequalities of smaller dimensions. This motivates us to augment classical algorithms with a consensus-based mechanism to solve the problem.

For this purpose, we assume that each agent maintains an estimate $\boldsymbol{x}^i(k) \in \mathbb{R}^N$ of the solution to $\mathrm{VI}(\boldsymbol{K}, \boldsymbol{F})$ at time instant $k \in \mathbb{Z}_+$. They can collect and share the estimates with other agents through a time-varying interaction network described by $\mathcal{G}_k = \{\mathcal{N}, \mathcal{E}(k)\}$. A directed edge $(i, j) \in \mathcal{E}(k)$ means that agent j can receive the estimate of agent i at time k. Denote the neighbor set of agent i at time k by $\mathcal{N}_i(k)$. It is defined by $\mathcal{N}_i(k) \triangleq \{j \in \mathcal{N} \mid (j, i) \in \mathcal{E}(k)\}$. Then, $\mathcal{N}_i(k)$ represents all the information source of agent i at time k. Certainly, the mapping F_i and set K_i are private to agent i and prohibitive to be shared with others.

Let $\hat{\boldsymbol{x}}^i(k) = \sum_{j=1}^n w_{ij}(k)\boldsymbol{x}^j(k)$ be the aggregate estimate of the solution to (1) at agent i with $w_{ij}(k)$ some nonnegative weight of agent i on agent j's estimate if there exists a directed edge $(j, i) \in \mathcal{E}(k)$. Correspondingly, we let $\check{\boldsymbol{x}}^i(k) = \sum_{j=1}^n w_{ji}(k)\boldsymbol{x}^i(k)$ be the overall estimate that agent i pushes into the network. These two variables represent how agent i collects and distributes the local recent computing progress in solving (1). Put all weights at time k into a matrix $W(k)$. Assume that matrix $W(k)$ is compatible with graph $\mathcal{G}_k$ for any $k \in \mathbb{Z}_+$ in the sense that $w_{ij}(k) > 0$ if and only if $j \in \mathcal{N}_i(k)$. For time-varying interaction graphs, most aforementioned results require the graph to be either undirected or directed but weight-balanced. Thus, we always have $\hat{\boldsymbol{x}}^i(k) = \check{\boldsymbol{x}}^i(k)$ for any k. Nevertheless, this implies that each matrix $W(k)$ should be doubly stochastic, which may be too restrictive in practical scenarios with possible interaction uncertainties and asymmetries.

In this paper, we focus on the solvability of the variational inequality problem (1) under general graph conditions as follows:

Assumption 2 $\{\mathcal{G}_k\}$ is a sequence of digraphs with self-loops and uniformly jointly strongly connected for some integer $\varrho > 0$.

Assumption 3 For any $k \in \mathbb{Z}_+$, matrix $W(k)$ is row-stochastic with $w_{ij}(k) \geq \eta$ if $w_{ij}(k) > 0$ for some scalar $0 < \eta < 1$.

Under Assumption 3, the collected and distributed estimates $\hat{\boldsymbol{x}}^i(k)$ and $\check{\boldsymbol{x}}^i(k)$ at agent i are generally not equal even when all agents have reached the expected solution. To address this issue, we modify classical projection methods to solve the local variational inequalities (2) and propose a distributed algorithm as follows:

$$\begin{aligned} \boldsymbol{x}^i(k+1) &= \Pi_{\boldsymbol{K}_i}\left(\hat{\boldsymbol{x}}^i(k) - \tau_k \frac{R_i F_i(\hat{\boldsymbol{x}}^i(k))}{z_i^i(k)}\right), \\ z^i(k+1) &= \sum_{j=1}^n w_{ij}(k) z^j(k), \end{aligned} \tag{3}$$

where $\boldsymbol{K}_i = \mathbb{R}^{\sum_{j=1}^{i-1} N_j} \times K_i \times \mathbb{R}^{\sum_{j=i+1}^{n} N_j}$, $R_i \in \mathbb{R}^{N \times N_i}$ is defined as $R_i = \text{col}(\boldsymbol{0}_{\sum_{j=1}^{k-1} N_j \times N_i}, \mathbb{I}_{N_i}, \boldsymbol{0}_{\sum_{j=k+1}^{n} N_j \times N_i})$, $\tau_k > 0$ is the stepsize, and $z^i(k) = \text{col}(z_1^i(k), \ldots, z_n^i(k)) \in \mathbb{R}^n$ with initial value $z_i^i(0) = 1$ and $z_j^i(0) = 0$. It can be verified that $\boldsymbol{K} = \cap_{i=1}^n \boldsymbol{K}_i$.

This rule is motivated by the existing consensus-based designs to solve special variational inequalities in the name of partial-decision information network games [14, 19]. To compensate the graph imbalance, we further introduce an extra vector z^i to rescale the iteration. Similar ideas have been partially discussed in the Nash equilibrium seeking literature [17, 20, 21]. By contrast, we consider more general variational inequalities under time-varying digraphs which may not be always connected and weight-balanced. These asymmetric and time-varying features of the interaction graphs certainly make the analysis of our algorithm much more challenging.

Define the transition matrix by $\Phi(k, s) \triangleq W(k-1)\cdots W(s)$ for any two integers $k > s \geq 0$ and let $\Phi(k, k) = \mathbb{I}_n$ for consistence. This matrix is verified to be row-stochastic for any $k \geq s \geq 0$ under Assumption 3. Here is a key lemma on its limit behaviors [22].

Lemma 2 *Suppose that Assumptions 2–3 hold. Consider the algorithm (3). Then, for any integer $k \geq 0$, there exists a normalized vector $\pi(k) = col(\pi_1(k), \ldots, \pi_n(k))$ (i.e., $\boldsymbol{1}_n^{\mathrm{T}}\pi(k) = 1$), such that the following statements hold:*

(1) $\pi(k) = W(k)^{\mathrm{T}}\pi(k+1)$.
(2) There exists a constant $\gamma \geq \eta^{(n-1)\varrho}$, such that $\pi_i(k) \geq \gamma$ for any $i \in \mathcal{N}$ and $k \in \mathbb{Z}_+$.
(3) For any $i, j \in \mathcal{N}$ and $k \geq s \geq 0$, $|[\Phi(k, s)]_{ij} - \pi_j(s)| \leq C\lambda^{k-s}$ for some constants $C > 0$ and $0 < \lambda < 1$.
(4) For any $i, j \in \mathcal{N}$, when $k - s \geq n\varrho + 2\varrho$, $[\Phi(k, s)]_{ij} > \xi$ for some constant $0 < \xi < 1$.
(5) For any $i \in \mathcal{N}$, $z_i^i(k) > 0$ for any $k \in \mathbb{Z}_+$ and satisfies that $|z_i^i(k) - \pi_i(0)| \leq C\lambda^k$.

Before the main results, two more supporting lemmas are extracted from [23, 24] for the following analysis.

Lemma 3 *Consider a nonnegative sequence $\{\gamma_k\}$ and a scalar $0 < \beta < 1$. Suppose that $\lim_{k\to\infty} \gamma_k = 0$. Then, $\lim_{k\to\infty} \sum_{l=0}^{k} \beta^{k-l}\gamma_l = 0$. In addition, if $\sum_{k=0}^{\infty} \gamma_k < \infty$, then $\sum_{k=0}^{\infty}\sum_{l=0}^{k} \beta^{k-l}\gamma_l < \infty$.*

Lemma 4 *Let $\{v_k\}$, $\{u_k\}$, $\{b_k\}$, and $\{c_k\}$ be nonnegative sequences, such that $\sum_{k=0}^{\infty} b_k < \infty$, $\sum_{k=0}^{\infty} c_k < \infty$, and*

$$v_{k+1} \leq (1 + b_k)v_k - u_k + c_k, \quad \forall k \geq 0.$$

Then, $\{v_k\}$ converges and $\sum_{k=0}^{\infty} u_k < \infty$.

4 Performance analysis

In this section, we analyze the convergence of algorithm (3) over time-varying unbalanced digraphs.

For convenience, we take

$$\epsilon_i(k) = \Pi_{\boldsymbol{K}_i}\left(\hat{\boldsymbol{x}}^i(k) - \tau_k \frac{R_i F_i(\hat{\boldsymbol{x}}^i(k))}{z_i^i(k)}\right) - \hat{\boldsymbol{x}}^i(k)$$

and rewrite (3) in the following perturbed form:

$$\begin{aligned} \boldsymbol{x}^i(k+1) &= \hat{\boldsymbol{x}}^i(k) + \epsilon_i(k), \\ z^i(k+1) &= \sum_{j=1}^n w_{ij}(k) z^j(k), \end{aligned} \tag{4}$$

where $\hat{\boldsymbol{x}}^i(k)$ and $\epsilon_i(k)$ are the linear nominal and nonlinear perturbed terms of (3), respectively.

We first show that each estimate $\boldsymbol{x}^i(k)$ converges to the global constrained set $\boldsymbol{K}$.

Lemma 5 *Let Assumptions 1–3 hold. Then, for any $i \in \mathcal{N}$ and $k \in \mathbb{Z}_+$, we have*

$$D^i(k) \leq (1 - \xi)^{\lfloor \frac{k}{n\varrho + 2\varrho} \rfloor} D^i(0) \tag{5}$$

with $D^i(k) \triangleq \max_{j\in\mathcal{N}} \mathrm{d}_{\boldsymbol{K}_i}(\boldsymbol{x}^j(k))$ and $\lfloor x \rfloor$ the maximum integer less than x.

Proof We start with the evolution of $d_{\boldsymbol{K}_i}(\boldsymbol{x}^j(k))$. Suppose that $j \neq i$. By definitions of $\Pi_{\boldsymbol{K}_i}$ and R_i, we have

$$\begin{aligned} d_{\boldsymbol{K}_i}(\boldsymbol{x}^j(k)) &= \|\boldsymbol{x}^j(k) - \Pi_{\boldsymbol{K}_i}(\boldsymbol{x}^j(k))\| \\ &= \|\boldsymbol{x}^j(k) - \Pi_{\boldsymbol{K}_i}(\hat{\boldsymbol{x}}^j(k-1) + \epsilon_j(k-1))\| \\ &= \|\hat{\boldsymbol{x}}^j(k-1) - \Pi_{\boldsymbol{K}_i}(\hat{\boldsymbol{x}}^j(k-1))\|. \end{aligned}$$

According to Lemma 1

$$\begin{aligned} & d_{\boldsymbol{K}_i}(\boldsymbol{x}^j(k)) \\ & \leq \sum_{l=1}^n w_{jl}(k-1)\|\boldsymbol{x}^l(k-1) - \Pi_{\boldsymbol{K}_i}(\boldsymbol{x}^l(k-1))\| \end{aligned}$$

$$= \sum_{l=1}^{n} w_{jl}(k-1)\, d_{\boldsymbol{K}_i}(\boldsymbol{x}^l(k-1)).$$

Repeating this iteration and recalling the definition of transition matrix Φ, we have

$$d_{\boldsymbol{K}_i}(\boldsymbol{x}^j(k)) \le \sum_{l=1}^{n} [\Phi(k,\, s)]_{jl} d_{\boldsymbol{K}_i}(\boldsymbol{x}^l(s))$$

for any $k \ge s \ge 0$. Due to the stochasticity of Φ, we obtain that $D^i(k) \le D^i(s)$ for any $k \ge s \ge 0$. Then, we recall the fact $d_{\boldsymbol{K}_i}(\boldsymbol{x}^i(k)) = 0$. This together with Lemma 2 implies

$$\begin{aligned} D^i(k) &\le D^i(\lfloor \frac{k}{n\varrho + 2\varrho} \rfloor (n\varrho + 2\varrho)) \\ &\le (1-\xi)^{\lfloor \frac{k}{n\varrho+2\varrho} \rfloor} D^i(0). \end{aligned}$$

The proof is thus complete. □

From $\boldsymbol{K} = \cap_{i=1}^{n} \boldsymbol{K}_i$ and its compactness, all estimates $\boldsymbol{x}^1(k), \ldots, \boldsymbol{x}^n(k)$ ultimately enter into set $\boldsymbol{K}$. Thus, they are uniformly bounded for any $k \in \mathbb{Z}_+$. Consequently, there exists a constant $L > 0$, such that $\|F_i(\boldsymbol{x}(k))\| \le \|\boldsymbol{F}(\boldsymbol{x}(k))\| \le L$ for each $i \in \mathcal{N}$ and $k \in \mathbb{Z}_+$.

Next, we move on to the limit behavior of sequence $\{\boldsymbol{x}^i(k)\}$. For this purpose, we let $\overline{\boldsymbol{x}}(k) = \sum_{i=1}^{n} \pi_i(k)\boldsymbol{x}^i(k)$ and denote $\Gamma(\overline{\boldsymbol{x}}(k),\, \boldsymbol{x}^*) = (\overline{\boldsymbol{x}}(k) - \boldsymbol{x}^*)^{\mathrm{T}}(\boldsymbol{F}(\overline{\boldsymbol{x}}(k)) - \boldsymbol{F}(\boldsymbol{x}^*))$ for short with $\pi(k)$ given in Lemma 2. According to Lemma 5, this weighted average $\overline{\boldsymbol{x}}(k)$ also converges to $\boldsymbol{K}$ as $k \to \infty$, while $\Gamma(\overline{\boldsymbol{x}}(k),\, \boldsymbol{x}^*) \ge 0$ under Assumption 1.

We are ready to present the first main result of this paper.

Theorem 1 *Let Assumptions 1–3 hold. Then, there exist three constants $C_1, C_2, C_3 > 0$, such that, for any $k \in \mathbb{Z}_+$, we have*

$$\begin{aligned} &\sum_{i=1}^{n} \pi_i(k+1)\|\boldsymbol{x}^i(k+1) - \boldsymbol{x}^*\|^2 \\ &\le \sum_{i=1}^{n} \pi_i(k)\|\boldsymbol{x}^i(k) - \boldsymbol{x}^*\|^2 - C_1\tau_k \Gamma(\overline{\boldsymbol{x}}(k),\, \boldsymbol{x}^*) \\ &\quad + C_2\tau_k^2 + C_3\tau_k \max_{i\in\mathcal{N}}\{\|\boldsymbol{x}^i(k) - \overline{\boldsymbol{x}}(k)\|\}. \end{aligned} \tag{6}$$

Proof At first, according to Proposition 1.5.8 in [1], $\boldsymbol{x}^* = \Pi_{\boldsymbol{K}}(\boldsymbol{x}^* - \tau \boldsymbol{F}(\boldsymbol{x}^*))$ for any $\tau > 0$. Or equivalently, $\boldsymbol{x}^* = \Pi_{\boldsymbol{K}_i}(\boldsymbol{x}^* - \tau R_i F_i(\boldsymbol{x}^*))$, $\forall i \in \mathcal{N}$ by the definitions of $\boldsymbol{K}_i$ and R_i. This combined with the nonexpansive property of projection operator implies

$$\begin{aligned} &\|\boldsymbol{x}^i(k+1) - \boldsymbol{x}^*\|^2 \\ &\le \left\| \Pi_{\boldsymbol{K}_i}\left(\hat{\boldsymbol{x}}^i(k) - \tau_k \frac{R_i F_i(\hat{\boldsymbol{x}}^i(k))}{z_i^i(k)}\right) \right. \\ &\quad \left. - \Pi_{\boldsymbol{K}_i}\left(\boldsymbol{x}^* - \tau_k \frac{R_i F_i(\boldsymbol{x}^*)}{z_i^i(k)}\right) \right\|^2 \\ &\le \|\hat{\boldsymbol{x}}^i(k) - \boldsymbol{x}^* - \frac{\tau_k R_i}{z_i^i(k)}[F_i(\hat{\boldsymbol{x}}^i(k)) - F_i(\boldsymbol{x}^*)]\|^2 \\ &\le \|\hat{\boldsymbol{x}}^i(k) - \boldsymbol{x}^*\|^2 + \frac{\tau_k^2}{(z_i^i(k))^2}\|F_i(\hat{\boldsymbol{x}}^i(k)) - F_i(\boldsymbol{x}^*)\|^2 \\ &\quad - \frac{2\tau_k}{z_i^i(k)}(\hat{\boldsymbol{x}}^i(k) - \boldsymbol{x}^*)^{\mathrm{T}} R_i[F_i(\hat{\boldsymbol{x}}^i(k)) - F_i(\boldsymbol{x}^*)]. \end{aligned} \tag{7}$$

Note that $\|\hat{\boldsymbol{x}}^i(k) - \boldsymbol{x}^*\|^2 \le \sum_{j=1}^{n} w_{ij}(k)\|\boldsymbol{x}^j(k) - \boldsymbol{x}^*\|^2$ under Assumption 3. We recall the boundedness of F_i and obtain

$$\begin{aligned} &\|\boldsymbol{x}^i(k+1) - \boldsymbol{x}^*\|^2 \\ &\le \sum_{j=1}^{n} w_{ij}(k)\|\boldsymbol{x}^j(k) - \boldsymbol{x}^*\|^2 + \frac{4L^2\tau_k^2}{(z_i^i(k))^2} \\ &\quad - \frac{2\tau_k}{z_i^i(k)}(\hat{\boldsymbol{x}}^i(k) - \boldsymbol{x}^*)^{\mathrm{T}} R_i[F_i(\hat{\boldsymbol{x}}^i(k)) - F_i(\boldsymbol{x}^*)]. \end{aligned}$$

Multiplying both side of this inequality by $\pi_i(k+1)$ and summing them together over $i \in \mathcal{N}$ imply

$$\begin{aligned} &\sum_{i=1}^{n} \pi_i(k+1)\|\boldsymbol{x}^i(k+1) - \boldsymbol{x}^*\|^2 \\ &\quad \le \sum_{i=1}^{n} \pi_i(k+1) \sum_{j=1}^{n} w_{ij}(k)\|\boldsymbol{x}^j(k) - \boldsymbol{x}^*\|^2 + \frac{4nL^2\tau_k^2}{\min_{i\in\mathcal{N}}(z_i^i(k))^2} \\ &\quad\quad - \sum_{i=1}^{n} \frac{2\tau_k}{z_i^i(k)}\pi_i(k+1)(\hat{\boldsymbol{x}}^i(k) - \boldsymbol{x}^*)^{\mathrm{T}} R_i[F_i(\hat{\boldsymbol{x}}^i(k)) \\ &\quad\quad - F_i(\boldsymbol{x}^*)] \\ &\quad \le \sum_{i=1}^{n} \pi_i(k)\|\boldsymbol{x}^i(k) - \boldsymbol{x}^*\|^2 + \frac{4nL^2\tau_k^2}{\min_{i\in\mathcal{N}}(z_i^i(k))^2} \\ &\quad\quad - \sum_{i=1}^{n} \frac{2\tau_k}{z_i^i(k)}\pi_i(k+1)(\hat{\boldsymbol{x}}^i(k) - \boldsymbol{x}^*)^{\mathrm{T}} R_i[F_i(\hat{\boldsymbol{x}}^i(k)) \\ &\quad\quad - F_i(\boldsymbol{x}^*)] \end{aligned}$$

with the property of $\pi(k)$ from Lemma 2.

Let us consider the upper bound of the right-hand side of the above inequality. We split the last term into three parts as follows:

$$\begin{aligned} &\sum_{i=1}^{n} \frac{2\tau_k}{z_i^i(k)}\pi_i(k+1)(\hat{\boldsymbol{x}}^i(k) - \boldsymbol{x}^*)^{\mathrm{T}} R_i[F_i(\hat{\boldsymbol{x}}^i(k)) - F_i(\boldsymbol{x}^*)] \\ &= \underbrace{\sum_{i=1}^{n} \frac{2\tau_k}{z_i^i(k)}\pi_i(k+1)(\overline{\boldsymbol{x}}(k) - \boldsymbol{x}^*)^{\mathrm{T}} R_i[F_i(\overline{\boldsymbol{x}}(k)) - F_i(\boldsymbol{x}^*)]}_{\Delta_1} \\ &\quad + \underbrace{\sum_{i=1}^{n} \frac{2\tau_k}{z_i^i(k)}\pi_i(k+1)(\hat{\boldsymbol{x}}^i(k) - \overline{\boldsymbol{x}}(k))^{\mathrm{T}} R_i[F_i(\overline{\boldsymbol{x}}(k)) - F_i(\boldsymbol{x}^*)]}_{\Delta_2} \\ &\quad + \underbrace{\sum_{i=1}^{n} \frac{2\tau_k}{z_i^i(k)}\pi_i(k+1)(\hat{\boldsymbol{x}}^i(k) - \boldsymbol{x}^*)^{\mathrm{T}} R_i[F_i(\hat{\boldsymbol{x}}^i(k)) - F_i(\overline{\boldsymbol{x}}(k))]}_{\Delta_3}. \end{aligned}$$

By Lemma 2, there exists a constant $\hat{\gamma} > 0$, such that $\hat{\gamma} \leq z_i^i(k) \leq 1$ for any $k \in \mathbb{Z}_+$ and $i \in \mathcal{N}$. Since the sequences generated by (3) are contained in the set $\{\boldsymbol{x} \mid d_{\boldsymbol{K}}(\boldsymbol{x}) \leq \sum_{i=1}^n D^i(0)\}$ and $\boldsymbol{K}$ is compact, all $\|\hat{\boldsymbol{x}}^i(k)\|$, $\|\boldsymbol{x}^*\|$, $\|\overline{\boldsymbol{x}}(k)\|$ can be upperly bounded by a large number $\Theta > 0$. Using Lemma 2, we can bound the first term Δ_1 from below as follows:

$$\Delta_1 \geq 2\gamma\tau_k(\overline{\boldsymbol{x}}(k) - \boldsymbol{x}^*)^{\mathrm{T}}(\boldsymbol{F}(\overline{\boldsymbol{x}}(k)) - \boldsymbol{F}(\boldsymbol{x}^*)).$$

Similarly, for the rest two parts, we have

$$-\Delta_2 \leq \frac{4L\tau_k}{\hat{\gamma}} \max_{i\in\mathcal{N}}\{\|\boldsymbol{x}^i(k) - \overline{\boldsymbol{x}}(k)\|\},$$
$$-\Delta_3 \leq \frac{4\Theta\ell\tau_k}{\hat{\gamma}} \max_{i\in\mathcal{N}}\{\|\boldsymbol{x}^i(k) - \overline{\boldsymbol{x}}(k)\|\}.$$

We put all inequalities together and obtain

$$\begin{aligned}&\sum_{i=1}^n \pi_i(k+1)\|\boldsymbol{x}^i(k+1) - \boldsymbol{x}^*\|^2 \\ &\quad\leq \sum_{i=1}^n \pi_i(k)\|\boldsymbol{x}^i(k) - \boldsymbol{x}^*\|^2 - 2\gamma\tau_k\Gamma(\overline{\boldsymbol{x}}(k), \boldsymbol{x}^*) \\ &\qquad + \frac{4nL^2\tau_k^2}{\hat{\gamma}^2} + \frac{4(L+\Theta\ell)\tau_k}{\hat{\gamma}} \max_{i\in\mathcal{N}}\{\|\boldsymbol{x}^i(k) - \overline{\boldsymbol{x}}(k)\|\}.\end{aligned}$$

Letting $C_1 = 2\gamma$, $C_2 = \frac{4nL^2}{\hat{\gamma}^2}$, and $C_3 = \frac{4(L+\Theta\ell)}{\hat{\gamma}}$ implies the expected inequality (6). □

According to Lemma 5 and Theorem 1, our distributed algorithm (3) may not converge to the expected solution $\boldsymbol{x}^*$ without further treatments. Precisely, we can rewrite (6) as follows:

$$\begin{aligned}&\sum_{i=1}^n \pi_i(k+1)\|\boldsymbol{x}^i(k+1) - \boldsymbol{x}^*\|^2 \\ &\quad\leq \sum_{i=1}^n \pi_i(k)\|\boldsymbol{x}^i(k) - \boldsymbol{x}^*\|^2 - C_1\tau_k\Xi_k\end{aligned}$$

with $\Xi_k \triangleq \Gamma(\overline{\boldsymbol{x}}(k), \boldsymbol{x}^*) - \frac{C_2}{C_1}\tau_k - \frac{C_3}{C_1}\max_{i\in\mathcal{N}}\{\|\boldsymbol{x}^i(k) - \overline{\boldsymbol{x}}(k)\|\}$. The following corollary summarizes the convergence error of $\overline{\boldsymbol{x}}(k)$ to $\boldsymbol{x}^*$ upon both the stepsize and consensus error.

Corollary 1 *Let Assumptions 1–3 hold. Suppose that* $\sum_{k=0}^{\infty}\tau_k = \infty$ *and* $\limsup_{k\to\infty}\tau_k = \overline{\tau} < \infty$. *Then, there exists a constant* $C_4 > 0$, *such that*

$$\begin{aligned}0 &\leq \liminf_{k\to\infty} \Gamma(\overline{\boldsymbol{x}}(k), \boldsymbol{x}^*) \\ &\leq C_4(\overline{\tau} + \limsup_{k\to\infty}\max_{i\in\mathcal{N}}\{\|\boldsymbol{x}^i(k) - \overline{\boldsymbol{x}}(k)\|\}).\end{aligned}$$

Its proof can be done by seeking a contradiction and is omitted to save space. With Corollary 1, diminishing stepsizes are required toward the exact solvability of our problem.

Next, we consider stepsizes $\{\tau_k\}$ satisfying the following conditions:

$$\sum_{k=0}^{\infty}\tau_k = \infty, \quad \sum_{k=0}^{\infty}\tau_k^2 < \infty. \tag{8}$$

Here is the second main result of this paper about the effectiveness of algorithm (4) under these stepsizes.

Theorem 2 *Let Assumptions 1–3 hold. Suppose that the stepsizes satisfy (8). Then, for any* $i \in \mathcal{N}$, $\lim_{k\to\infty}\|\boldsymbol{x}^i(k) - \overline{\boldsymbol{x}}(k)\| = 0$ *and* $\lim_{k\to\infty}\boldsymbol{x}^i(k) = \boldsymbol{x}^*$.

Proof To prove this theorem, we first claim that $\lim_{k\to\infty}\|\epsilon_i(k)\| = 0$. In fact, according to the definition of ϵ_i, it follows that:

$$\begin{aligned}\|\epsilon_i(k)\| &\leq \|\Pi_{\boldsymbol{K}_i}(\hat{\boldsymbol{x}}^i(k) - \tau_k\frac{R_iF_i(\hat{\boldsymbol{x}}^i(k))}{z_i^i(k)}) - \Pi_{\boldsymbol{K}_i}(\hat{\boldsymbol{x}}^i(k))\| \\ &\quad + \|\Pi_{\boldsymbol{K}_i}(\hat{\boldsymbol{x}}^i(k)) - \hat{\boldsymbol{x}}^i(k)\| \\ &\leq \tau_k\|\frac{R_iF_i(\hat{\boldsymbol{x}}^i(k))}{z_i^i(k)}\| + \sum_{l=1}^n w_{il}(k)d_{\boldsymbol{K}_i}(\boldsymbol{x}^l(k)) \\ &\leq \frac{L\tau_k}{\hat{\gamma}} + D^i(k),\end{aligned} \tag{9}$$

where we have used the boundedness of $\boldsymbol{F}$ and the definition of $D^i(k)$. From the last inequality, we can recall Lemma 5 and confirm our claim.

By Lemma 2, we obtain

$$\begin{aligned}\overline{\boldsymbol{x}}(k) &= \sum_{j=1}^n \pi_j(k)[\hat{\boldsymbol{x}}^j(k-1) + \epsilon_j(k-1)] \\ &= \sum_{j=1}^n \pi_j(k-1)\boldsymbol{x}^j(k-1) + \sum_{j=1}^n \pi_j(k)\epsilon_j(k-1) \\ &= \overline{\boldsymbol{x}}(k-1) + \sum_{j=1}^n \pi_j(k)\epsilon_j(k-1) \\ &= \overline{\boldsymbol{x}}(0) + \sum_{l=1}^k\sum_{j=1}^n \pi_j(l)\epsilon_j(l-1).\end{aligned}$$

Meanwhile, we rewrite algorithm (4) in the following form:

$$\boldsymbol{x}^i(k) = \sum_{j=1}^n[\Phi(k,0)]_{ij}\boldsymbol{x}^j(0) + \sum_{l=1}^k\sum_{j=1}^n[\Phi(k,l)]_{ij}\epsilon_j(l-1).$$

Since $\sum_{j=1}^n \pi_j(0) = 1$, we combine the above two equations and obtain

$$\boldsymbol{x}^i(k) - \overline{\boldsymbol{x}}(k) = \sum_{j=1}^n\{[\Phi(k,0)]_{ij} - \pi_j(0)\}[\boldsymbol{x}^j(0) - \overline{\boldsymbol{x}}(0)]$$

$$+\sum_{l=1}^{k}\sum_{j=1}^{n}\{[\Phi(k,\,l)]_{ij}-\pi_j(l)\}\epsilon_j(l-1).$$

Recalling that $|[\Phi(k,\,l)]_{ij}-\pi_j(l)|\le C\lambda^{k-l}$ under theorem assumptions by Lemma 5, we further have

$$\|\boldsymbol{x}^i(k)-\overline{\boldsymbol{x}}(k)\|\le C\lambda^k\sum_{j=1}^{n}\|\boldsymbol{x}^j(0)-\overline{\boldsymbol{x}}(0)\| + nC\sum_{l=1}^{k}\lambda^{k-l}\max_{j\in\mathcal{N}}\|\epsilon_j(l-1)\|. \quad (10)$$

Since $\lim_{k\to\infty}\|\epsilon_i(k)\|=0$, we recall Lemma 3 and obtain $\lim_{k\to\infty}\|\boldsymbol{x}^i(k)-\overline{\boldsymbol{x}}(k)\|=0$. This verifies the first part of this theorem.

Next, we are going to prove the second part. We multiply both sides of inequality (10) by τ_k and substitute the upper bound (9) into it. This gives

$$\begin{aligned}&\tau_k\|\boldsymbol{x}^i(k)-\overline{\boldsymbol{x}}(k)\|\\&\le\tau_kC\lambda^k\sum_{j=1}^{n}\|\boldsymbol{x}^j(0)-\overline{\boldsymbol{x}}(0)\|\\&\quad+\tau_knC\sum_{l=1}^{k}\lambda^{k-l}\Big[\frac{L\tau_{l-1}}{\hat{\gamma}}+\max_{j\in\mathcal{N}}D^j(l-1)\Big].\end{aligned}$$

By completing squares, we further have

$$\begin{aligned}&\tau_k\|\boldsymbol{x}^i(k)-\overline{\boldsymbol{x}}(k)\|\\&\le C^2\sum_{j=1}^{n}\|\boldsymbol{x}^j(0)-\overline{\boldsymbol{x}}(0)\|^2\tau_k^2+\lambda^{2k}+2n^2C^2\tau_k^2\\&\quad+\sum_{l=1}^{k}\lambda^{2(k-l)}\Big[\frac{L^2\tau_{l-1}^2}{\hat{\gamma}^2}+\max_{j\in\mathcal{N}}\|D^j(l-1)\|^2\Big].\end{aligned}$$

Summing up the above inequalities for $k\in\mathbb{Z}_+$ provides us

$$\begin{aligned}&\sum_{k=1}^{\infty}\tau_k\|\boldsymbol{x}^i(k)-\overline{\boldsymbol{x}}(k)\|\\&\le C^2\sum_{j=1}^{n}\|\boldsymbol{x}^j(0)-\overline{\boldsymbol{x}}(0)\|^2\sum_{k=1}^{\infty}\tau_k^2+\sum_{k=1}^{\infty}\lambda^{2k}\\&\quad+2n^2C^2\sum_{k=1}^{\infty}\tau_k^2\\&\quad+\sum_{k=1}^{\infty}\sum_{l=1}^{k}\lambda^{2(k-l)}\Big[\frac{L^2\tau_{l-1}^2}{\hat{\gamma}^2}+\max_{j\in\mathcal{N}}\|D^j(l-1)\|^2\Big].\end{aligned}$$

According to Lemma 5, $\sum_{l=1}^{\infty}\max_{i\in\mathcal{N}}\|D^i(l-1)\|^2<\infty$. This along with $\sum_{k=0}^{\infty}\tau_k^2<\infty$ implies $\sum_{k=1}^{\infty}\sum_{l=1}^{k}\lambda^{2(k-l)}[\frac{L^2\tau_{l-1}^2}{\hat{\gamma}^2}+\max_{j\in\mathcal{N}}\|D^j(l-1)\|^2]<\infty$ by Lemma 3. Consequently

$$\sum_{k=1}^{\infty}\tau_k\|\boldsymbol{x}^i(k)-\overline{\boldsymbol{x}}(k)\|<\infty,$$
$$\sum_{k=1}^{\infty}\tau_k\sum_{i=1}^{n}\|\boldsymbol{x}^i(k)-\overline{\boldsymbol{x}}(k)\|<\infty.$$

Note that $\max_{i\in\mathcal{N}}\{\|\boldsymbol{x}^i(k)-\overline{\boldsymbol{x}}(k)\|\le\sum_{j=1}^{n}\|\boldsymbol{x}^i(k)-\overline{\boldsymbol{x}}(k)\|\}$. Applying Lemma 4 to inequality (6), we have the convergence of $\sum_{i=1}^{n}\pi_i(k)\|\boldsymbol{x}^i(k)-\boldsymbol{x}^*\|^2$, or equivalently, the convergence of $\|\boldsymbol{x}^i(k)-\boldsymbol{x}^*\|$ for any $i\in\mathcal{N}$.

Meanwhile, in this case, $\liminf_{k\to\infty}\Gamma(\overline{\boldsymbol{x}}(k),\,\boldsymbol{x}^*)=0$ by Corollary 1. Thus, it has a subsequence $\{\overline{\boldsymbol{x}}(k_p)\}$, such that $\lim_{p\to\infty}\Gamma(\overline{\boldsymbol{x}}(k_p),\,\boldsymbol{x}^*)=0$. Since $\overline{\boldsymbol{x}}(k)$ is bounded, this subsequence again has a convergence subsequence. We abuse the notation and still denote it by $\{\overline{\boldsymbol{x}}(k_p)\}$. By the continuity of $\boldsymbol{F}$, $\Gamma(\lim_{p\to\infty}\overline{\boldsymbol{x}}(k_p),\,\boldsymbol{x}^*)=0$. According to the strict monotonicity of $\boldsymbol{F}$, $\lim_{p\to\infty}\overline{\boldsymbol{x}}(k_p)=\boldsymbol{x}^*$.

Since $\|\boldsymbol{x}^i(k)-\boldsymbol{x}^*\|\le\|\boldsymbol{x}^i(k)-\overline{\boldsymbol{x}}(k)\|+\|\overline{\boldsymbol{x}}(k)-\boldsymbol{x}^*\|$ by the triangle inequality, we take the limit inferiors of both sides and recall that $\lim_{k\to\infty}\|\boldsymbol{x}^i(k)-\overline{\boldsymbol{x}}(k)\|=0$. It follows that:

$$\lim_{k\to\infty}\|\boldsymbol{x}^i(k)-\boldsymbol{x}^*\|\le\liminf_{t\to\infty}\|\overline{\boldsymbol{x}}(k)-\boldsymbol{x}^*\|=0.$$

The proof is thus complete. □

Remark 1 Theorems 1 and 2 establish the basic convergence performance of (3) under different stepsize conditions. The algorithm (3) can be regarded as a distributed counterpart of classical projection rules for these constrained variational inequalities [1, 2].

5 Comparative discussion

In this section, we provide two examples to illustrate how the proposed algorithm (3) can solve different distributed computing problems in a unified fashion.

5.1 Seeking Nash equilibrium in noncooperative games

We start with the partial-information Nash equilibrium seeking problem, which has been extensively studied in the past few years.

Following the formulation in [9, 14], we consider an n-player noncooperative game $G=\{\mathcal{N},\,\{\theta_1,\,\ldots,\,\theta_N\},\,K_1\times\cdots\times K_n\}$, where $\theta_i(\boldsymbol{x}_i,\boldsymbol{x}_{-i})$ and $K_i\subset\mathbb{R}^{N_i}$ are the cost function and constrained strategy set of player i, respectively. Here, $\boldsymbol{x}_i\in\mathbb{R}^{N_i}$ is player i's strategy and $\boldsymbol{x}_{-i}$ represents the other players' strategies. In the partial-information setting, each player only knows its own cost function and constrained set. Meanwhile, the players are allowed to communicate with others through a network. The aim of each player is to minimize its cost function by selecting permissible strategy as follows:

$$\begin{aligned} &\min \quad \theta_i(\boldsymbol{x}_i, \boldsymbol{x}_{-i}), \\ &\text{s.t.} \quad \boldsymbol{x}_i \in K_i. \end{aligned} \tag{11}$$

Suppose that $\theta_i(\boldsymbol{x}_i, \boldsymbol{x}_{-i})$ is convex and continuously differentiable for each fixed $\boldsymbol{x}_{-i}$. We can resort to Proposition 1.4.2 in [1] and characterize the equilibria of this game as the solution to the following variational inequalities:

$$(\boldsymbol{x} - \boldsymbol{x}^*)^{\mathrm{T}} F(\boldsymbol{x}^*) \geq 0, \quad \forall \boldsymbol{x} \in \boldsymbol{K}, \tag{12}$$

where $F(\boldsymbol{x}) = \mathrm{col}(\nabla_{\boldsymbol{x}_1}\theta_1(\boldsymbol{x}_1, \boldsymbol{x}_{-1}), \ldots, \nabla_{\boldsymbol{x}_n}\theta_n(\boldsymbol{x}_n, \boldsymbol{x}_{-n}))$ and $\boldsymbol{K} = K_1 \times \cdots \times K_n$.

Since player i knows $\nabla\theta_i$ and K_i, this problem can be solved by our model. In fact, algorithm (3) in this case can be repeated as follows:

$$\begin{aligned} \boldsymbol{x}^i(k+1) &= \Pi_{\boldsymbol{K}_i}\left(\hat{\boldsymbol{x}}^i(k) - \tau_k \frac{R_i \nabla_{\boldsymbol{x}_i}\theta_i(\hat{\boldsymbol{x}}^i_i(k), \hat{\boldsymbol{x}}^i_{-i}(k))}{z^i_i(k)}\right), \\ \boldsymbol{z}^i(k+1) &= \sum_{j=1}^n w_{ij}(k)\boldsymbol{z}^j(k), \end{aligned} \tag{13}$$

where $\boldsymbol{x}^i(k)$, $\hat{\boldsymbol{x}}^i(k)$, $\boldsymbol{K}_i$, $\boldsymbol{z}^i$, and R_i are defined as above.

Here is the corresponding conclusion.

Corollary 2 *Suppose that the function θ_i is strictly convex in $\boldsymbol{x}_i$ for any fixed $\boldsymbol{x}_{-i}$ and its gradient $\nabla_{\boldsymbol{x}_i}\theta_i$ is l-Lipschitz for some constant $l > 0$. Suppose that the stepsizes satisfy (8). Then, under Assumptions 2 and 3, $\lim_{k\to\infty} \boldsymbol{x}^i(k) = \boldsymbol{x}^*$ for any $i \in \mathcal{N}$ with $\boldsymbol{x}^*$ the Nash equilibrium corresponding to game G.*

Similar algorithms have been developed in the literature for either fixed or switching interaction graphs [14, 16, 17, 21, 25]. However, most of these results require the interaction graphs be undirected or at least weight-balanced. Here, we remove this limitation and allow the graphs to be time-varying and weight-unbalanced.

Remarkably, a very recent work [26] has also developed a similar rule to solve this Nash equilibrium seeking problem under the same setting. However, to address the imbalance issue, the full mapping $\boldsymbol{F}$ is required to be strongly monotone while each local cost function θ_i should be strongly convex. Here, we only assume some mild strict monotonicity of $\boldsymbol{F}$ to ensure the convergence of the algorithm, which certainly includes their results as special cases.

5.2 Solving large-scale linear equations

We further consider the problem to solve large-scale linear equations by distributed designs.

Consider linear algebraic equations

$$\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b} \tag{14}$$

with $\boldsymbol{A} \in \mathbb{R}^{N\times N}$, $\boldsymbol{x} \in \mathbb{R}^N$, and $\boldsymbol{b} \in \mathbb{R}^N$. Without loss of generality, we assume that matrix $\boldsymbol{A}$ is positive definite. Developing effective algorithms to solve this problem is fundamental in scientific computing and has been found in many areas. Since the number N may be too large to over the capacity of single computing node, we shall partition these equations into some small groups and utilize a network of computing node to solve the problem in a cooperative way. This problem has been partially studied in many papers, e.g., [27–30]. Here, we show how this problem can be solved by the algorithm (3) under time-varying digraphs.

To be specific, we assume that each node $i \in \mathcal{N}$ knows N_i rows of $\boldsymbol{A}$ and also the corresponding portion of $\boldsymbol{b}$ with $\sum_{i=1}^n N_i = N$. Denote these portions by $\boldsymbol{A}_i$ and $\boldsymbol{b}_i$. Note that these equations are equivalent to the following unconstrained variational inequalities:

$$(\boldsymbol{x} - \boldsymbol{x}^*)^{\mathrm{T}}(\boldsymbol{A}\boldsymbol{x}^* - \boldsymbol{b}) \geq 0, \quad \forall \boldsymbol{x} \in \mathbb{R}^N. \tag{15}$$

Thus, these variational inequalities can be solved by our distributed computing model by letting $K_i = \mathbb{R}^{N_i}$ and $F_i(\boldsymbol{x}) = \boldsymbol{A}_i\boldsymbol{x} - \boldsymbol{b}_i$. The corresponding distributed solver is thus given as follows:

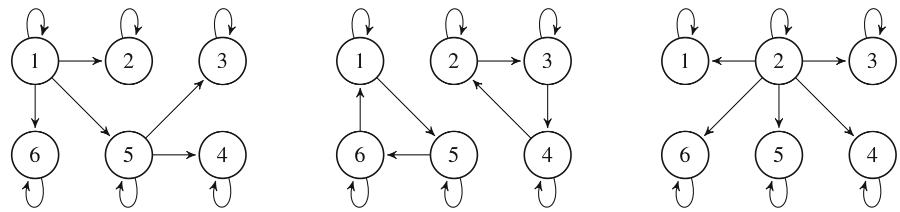
$$\begin{aligned} \boldsymbol{x}^i(k+1) &= \hat{\boldsymbol{x}}^i(k) - \tau_k \frac{R_i(\boldsymbol{A}_i\hat{\boldsymbol{x}}^i(k) - \boldsymbol{b}_i)}{z^i_i(k)}, \\ \boldsymbol{z}^i(k+1) &= \sum_{j=1}^n w_{ij}(k)\boldsymbol{z}^j(k), \end{aligned} \tag{16}$$

where $\boldsymbol{x}^i$ denote agent i's estimation of the solution $\boldsymbol{x}^*$.

Corollary 3 *Suppose that matrix $\boldsymbol{A}$ is positive definite and the stepsizes satisfy (8). Then, under Assumptions 2 and 3, $\lim_{k\to\infty} \boldsymbol{x}^i(k) = \boldsymbol{x}^*$ for any $i \in \mathcal{N}$, where $\boldsymbol{x}^*$ is the unique solution to linear equations (14).*

Different from the primal-dual type distributed solvers in [29–31], we develop a novel distributed solver from the perspective of variational inequalities for these linear equations over time-varying digraphs. Compared with similar rules in [28, 32], only a portion of the local estimate at each agent requires extra update other than local averaging. This definitely saves many computation resources.

Fig. 1 The interaction graphs in our example



(a) $\overline{\mathcal{G}}_1$ (b) $\overline{\mathcal{G}}_2$ (c) $\overline{\mathcal{G}}_3$

6 Simulation

In this section, we solve the River Basin Pollution game modified from [33] to illustrate the effectiveness of our preceding algorithm (3).

Suppose that there are six factories located along a river. Each factory engages in an economic activity that will cause pollution to the river (e.g., non-ferrous metal smelting). Two monitoring stations are located along the river to monitor the pollutant concentration levels. Besides, the government gives subsidies (or applies Pigouvian taxes) to these factories based on the overall pollutant concentration level. Each factory wants to minimize its overall function.

To be specific, the revenue and expenditure of player j are $R_j(\boldsymbol{x}) = [d_1 - d_2(\sum_{i=1}^{6} x_i)]x_j$ and $F_j(\boldsymbol{x}) = (c_{1j} + c_{2j}x_j)x_j$ with constants $d_1,\ d_2,\ c_{1j}$, and c_{2j}. Here, economic constants d_1 and d_2 determine the inverse demand law, while c_{1j} and c_{2j} are player j's private information. Moreover, the pollution concentration measured at the lth monitoring station is $q_l(\boldsymbol{x}) = \sum_{j=1}^{6} u_{jl}e_j x_j$ with parameters $u_{jl},\ e_j > 0$. The subsidies are given according to $T_l(\boldsymbol{x}) = \lambda_l(q_l(\boldsymbol{x}) - K_l)$ with K_l the government's maximum tolerance level of pollutant concentration and $\lambda_l > 0$ some penalty coefficient. Thus, the overall cost of each player j can be put down as $\phi_j(\boldsymbol{x}) = F_j(\boldsymbol{x}) + \sum_{l=1}^{2} T_l(\boldsymbol{x}) - R_j(\boldsymbol{x})$. Moreover, we assume that player j takes its economic activity level x_j within some strategy set Ω_j. In this way, the players play a noncooperative game.

Since the local cost functions contain private information of each factory, centralized algorithms requiring all these sensitive data in [33] are prohibitive in this case. Thus, we provide distributed design which relies on local computation and neighboring information to solve the problem. Let $\phi(\boldsymbol{x}) = \mathrm{col}(\phi_1(\boldsymbol{x}),\ \dots,\ \phi_6(\boldsymbol{x}))$. We can formulate the problem as some variational inequalities of the form (12) with $\boldsymbol{K} = \Omega_1 \times \cdots \times \Omega_6$. Suppose that these players can share information with other agents through a time-varying interaction graph except their private parameters in cost functions. Here, the interaction graph is time-varying and weight-unbalanced. Thus, existing Nash equilibrium seeking rules might fail to solve the problem. Meanwhile, we can resort to the arguments in Sect. 5 and confirm its solvability by algorithm (3) in a distributed way.

Table 1 Parameters in our game

Player j	c_{1j}	c_{2j}	e_j	u_{j1}	u_{j2}
1	0.1	0.01	0.5	6.5	4.5
2	0.2	0.05	0.25	5	6.25
3	0.15	0.01	0.55	5.5	3.75
4	0.25	0.05	0.5	4.5	5
5	0.2	0.02	0.25	4	5.5
6	0.1	0.03	0.25	5	6

For simulations, the interactions among these players are determined by digraphs in Fig. 1. Here, the interaction graph switches according to $\mathcal{G}(k) = \overline{\mathcal{G}}_{1+\mathrm{mod}(\lfloor k/3 \rfloor,\, 3)}$. Moreover, the nonzero elements in the corresponding weight matrix are given by $w_{ij} = \frac{1}{|\mathcal{N}_i|}$ for $i \in \mathcal{N}$ with $|\mathcal{N}_i|$ the cardinality of node i's neighboring set $\mathcal{N}_i$. We set $d_1 = 3,\ d_2 = 0.01$, and $K_1 = K_2 = 100$. Other parameters are given in Table 1. The strategy set Ω_j for player j is $[0,\ X_j]$ with X_j randomly chosen in $[30,\ 40]$ for all $j = 1,\ \dots,\ 6$. The Nash equilibrium to this game can be numerically derived as $\boldsymbol{x}^* \approx \mathrm{col}(13.22, 10.03, 15.68, 4.29, 24.68, 17.45)$. We set $\tau_k = 2/(k^{0.6} + 10)$. The simulation results are provided in Figs. 2 and 3. Figure 2 shows the profiles of each player's estimate about its corresponding expected strategy, while Fig. 3 illustrates the evolution of the relative convergence errors of the form $\|\boldsymbol{x}^i(k) - \boldsymbol{x}^*\|/\|\boldsymbol{x}^*\|$ at each player i. One can find that all estimates of these players quickly converge to the expected Nash equilibrium. These observations confirm the effectiveness of algorithm (3) in solving problem (1) over time-varying unbalanced digraphs.

7 Conclusion

In this paper, we studied a distributed model to solve the variational inequality problem when each agent only knows a part of the full mapping and also a subset of the whole decision variables. A consensus-based projection algorithm was developed for each agent and shown to be effective in solving the problem over time-varying unbalanced directed graphs. Future work may include how to design

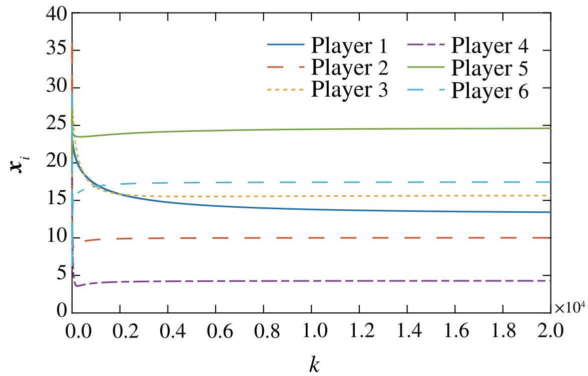


Fig. 2 Activity levels of all factories

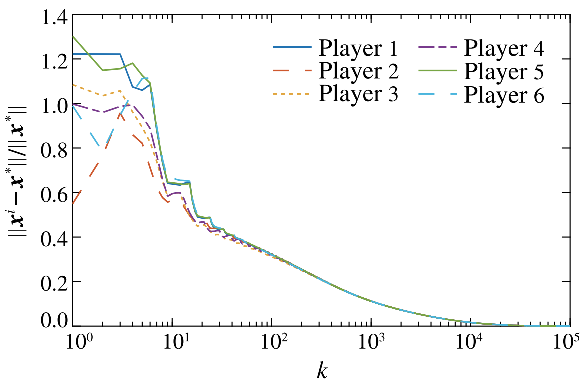


Fig. 3 Relative error between $\boldsymbol{x}^i$ and $\boldsymbol{x}^*$

communication-efficient algorithms for general mappings without the monotonicity condition.

References

1. Facchinei, F., & Pang, J.-S. (2003). *Finite-dimensional Variational Inequalities and Complementarity Problems*. New York: Springer.
2. Ansari, Q. H., Lalitha, C., & Mehta, M. (2013). *Generalized Convexity, Nonsmooth Variational Inequalities, and Nonsmooth Optimization*. Boca Raton, FL: CRC Press.
3. Bertsekas, D. P., & Tsitsiklis, J. N. (1989). *Parallel and Distributed Computation: Numerical Methods*. Englewood Cliffs, NJ: Prentice-Hall.
4. Koshal, J., Nedić, A., & Shanbhag, U. V. (2011). Multiuser optimization: Distributed algorithms and error analysis. *SIAM Journal on Optimization, 21*(3), 1046–1081.
5. Yang, T., Yi, X., Wu, J., Yuan, Y., Wu, D., Meng, Z., Hong, Y., Wang, H., Lin, Z., & Johansson, K. H. (2019). A survey of distributed optimization. *Annual Reviews in Control, 47*, 278–305.
6. Beznosikov, A., Dvurechenskii, P., Koloskova, A., Samokhin, V., Stich, S. U., & Gasnikov, A. (2022). Decentralized local stochastic extra-gradient for variational inequalities. *Advances in Neural Information Processing Systems, 35*, 38116–38133.
7. Beznosikov, A., Richtárik, P., Diskin, M., Ryabinin, M., & Gasnikov, A. (2022). Distributed methods with compressed communication for solving variational inequalities, with theoretical guarantees. *Advances in Neural Information Processing Systems, 35*, 14013–14029.
8. Tu, Z. (2023). *Research on communication-efficient distributed optimization algorithms*. Ph.D. thesis. Beijing: University of Chinese Academy of Sciences.
9. Salehisadaghiani, F., & Pavel, L. (2016). Distributed nash equilibrium seeking: A gossip-based algorithm. *Automatica, 72*, 209–216.
10. Scutari, G., Palomar, D. P., Facchinei, F., & Pang, J.-S. (2010). Convex optimization, game theory, and variational inequality theory. *IEEE Signal Processing Magazine, 27*(3), 35–49.
11. Liang, S., Yi, P., & Hong, Y. (2017). Distributed nash equilibrium seeking for aggregative games with coupled constraints. *Automatica, 85*, 179–185.
12. Parise, F., & Ozdaglar, A. (2019). A variational inequality framework for network games: Existence, uniqueness, convergence and sensitivity analysis. *Games and Economic Behavior, 114*, 47–82.
13. Yi, P., & Pavel, L. (2019). An operator splitting approach for distributed generalized nash equilibria computation. *Automatica, 102*, 111–121.
14. Tatarenko, T., Shi, W., & Nedić, A. (2021). Geometric convergence of gradient play algorithms for distributed nash equilibrium seeking. *IEEE Transactions on Automatic Control, 66*(11), 5342–5353.
15. Belgioioso, G., Nedic, A., & Grammatico, S. (2021). Distributed generalized nash equilibrium seeking in aggregative games on time-varying networks. *IEEE Transactions on Automatic Control, 66*(5), 2061–2075.
16. Bianchi, M., & Grammatico, S. (2021). Fully distributed nash equilibrium seeking over time-varying communication networks with linear convergence rate. *IEEE Control Systems Letters, 5*(2), 499–504.
17. Tang, Y., Yi, P., Zhang, Y., & Liu, D. (2022). Nash equilibrium seeking over directed graphs. *Autonomous Intelligent Systems, 2*(1), 7.
18. Lin, P., Ren, W., & Song, Y. (2016). Distributed multi-agent optimization subject to nonidentical constraints and communication delays. *Automatica, 65*, 120–131.
19. Gadjov, D., & Pavel, L. (2019). A passivity-based approach to nash equilibrium seeking over networks. *IEEE Transactions on Automatic Control, 64*(3), 1077–1092.
20. Lou, Y., Hong, Y., Xie, L., Shi, G., & Johansson, K. H. (2016). Nash equilibrium computation in subnetwork zero-sum games with switching communications. *IEEE Transactions on Automatic Control, 61*(10), 2920–2935.
21. Zhu, Y., Yu, W., Wen, G., & Chen, G. (2021). Distributed nash equilibrium seeking in an aggregative game on a directed graph. *IEEE Transactions on Automatic Control, 66*(6), 2746–2753.
22. Li, H., Lü, Q., & Huang, T. (2018). Distributed projection subgradient algorithm over time-varying general unbalanced directed graphs. *IEEE Transactions on Automatic Control, 64*(3), 1309–1316.
23. Nedic, A., Ozdaglar, A., & Parrilo, P. A. (2010). Constrained consensus and optimization in multi-agent networks. *IEEE Transactions on Automatic Control, 55*(4), 922–938.
24. Bauschke, H. H., & Combettes, P. L. (2011). *Convex Analysis and Monotone Operator Theory in Hilbert Spaces*. New York: Springer.
25. Arefizadeh, S., Nedić, A. (2022). *A distributed algorithm for aggregative games on directed communication graphs*. In: *Proceedings of IEEE 61st Conference on Decision and Control (CDC)*, pp. 6407–6412. IEEE
26. Nguyen, D.T.A., Nguyen, D.T., Nedić, A. (2023). Distributed Nash equilibrium seeking over time-varying directed communication networks. arXiv:2201.02323
27. Mou, S., Liu, J., & Morse, A. S. (2015). A distributed algorithm for solving a linear algebraic equation. *IEEE Transactions on Automatic Control, 60*(11), 2863–2878.
28. Wang, P., Ren, W., & Duan, Z. (2018). Distributed algorithm to solve a system of linear equations with unique or multiple solutions from arbitrary initializations. *IEEE Transactions on Control of Network Systems, 6*(1), 82–93.

29. Zeng, X., Liang, S., Hong, Y., & Chen, J. (2018). Distributed computation of linear matrix equations: An optimization perspective. *IEEE Transactions on Automatic Control, 64*(5), 1858–1873.
30. Liu, Y., Lageman, C., Anderson, B. D., & Shi, G. (2019). An Arrow-Hurwicz-Uzawa type flow as least squares solver for network linear equations. *Automatica, 100*, 187–193.
31. Tang, Y., Zhang, Y., Li, R., & Wang, X. (2023). On measurement disturbances in distributed least squares solvers for linear equations. arXiv:2305.05512
32. Wang, L., Fullmer, D., & Morse, A.S. (2016). A distributed algorithm with an arbitrary initialization for solving a linear algebraic equation. In: *2016 American Control Conference* (ACC), pp. 1078–1081. IEEE
33. Krawczyk, J. B., & Uryasev, S. (2000). Relaxation algorithms to find nash equilibria with economic applications. *Environmental Modeling and Assessment, 5*, 63–73.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.

Yichen Zhang received the B.Eng. degree in Measurement & Control Technology and Instruments from Beijing University of Posts and Telecommunications, Beijing, China, where he is currently working toward the M.Sc. degree in Control Science and Engineering. His research interests include distributed optimization and multi-agent systems.

Yutao Tang received the B.Sc. degree in Applied Mathematics from Harbin Engineering University, Harbin, China, and the Ph.D. degree in Operations Research and Cybernetics from the Chinese Academy of Sciences, Beijing, China. He is currently a Professor with the School of Artificial Intelligence, Beijing University of Posts and Telecommunications. His research interests include nonlinear control, distributed optimization, mobile robotics, and machine learning.

Zhipeng Tu received the B.Sc. degree in Mathematics and Applied Mathematics from Xi'an Jiaotong University and the Ph.D. degree in Systems Theory from Academy of Mathematics and Systems Science, Chinese Academy of Sciences. He is currently with the Network Technology Laboratory, Huawei Technologies. His research interests include distributed optimization, machine learning, and parallel training.

Yiguang Hong received his B.Sc. and M.Sc. degrees from Department of Mechanics of Peking University, China, and the Ph.D. degree from the Chinese Academy of Sciences (CAS), China. He is currently a Professor and Deputy Director of Shanghai Research Institute for Intelligent Autonomous Systems, Tongji University, Shanghai. He was a Professor of Academy of Mathematics and Systems Science, CAS. Also, he is a Fellow of IEEE, a Fellow of Chinese Association for Artificial Intelligence, and a Fellow of Chinese Association of Automation. His current research interests include nonlinear control, multi-agent systems, distributed optimization and game, machine learning, and social networks.