Distributed Optimal Variational GNE Seeking in Merely Monotone Games

Wangli He, *Senior Member, IEEE*, and Yanzhen Wang

***Abstract*—In this paper, the optimal variational generalized Nash equilibrium (v-GNE) seeking problem in merely monotone games with linearly coupled cost functions is investigated, in which the feasible strategy domain of each agent is coupled through an affine constraint. A distributed algorithm based on the hybrid steepest descent method is first proposed to seek the optimal v-GNE. Then, an accelerated algorithm with relaxation is proposed and analyzed, which has the potential to further improve the convergence speed to the optimal v-GNE. Some sufficient conditions in both algorithms are obtained to ensure the global convergence towards the optimal v-GNE. To illustrate the performance of the algorithms, numerical simulation is conducted based on a networked Nash-Cournot game with bounded market capacities.**

***Index Terms*—Distributed algorithms, equilibria selection, generalized Nash equilibrium (GNE), merely monotone games.**

I. Introduction

GENERALIZED games are a modeling paradigm employed to investigate the direct interaction among self-interested decision makers, also known as agents in multi-agent systems (MAS), where agents individually aim to optimize their respective cost functions according to the strategies employed by others. Generalized games are ubiquitous in wide engineering applications, including lane-changing control for autonomous vehicles [1], demand-side response [2]–[4], charging and discharging of electric vehicles [5]. A particularly favourable and rational solution is the generalized Nash equilibrium (GNE), in which no one can independently modify its strategy to decrease its own cost function [6]. However, if the GNE is non-unique, each agent has to decide which GNE has to follow individually. Sometimes this will lead to be inconsistent. The problem of non-unique GNE has been found in autonomous driving [7], and distributed energy resources field [8], [9].

When multiple GNEs exist, agents may have their own preference for different GNEs. For example, in the field of demand-side management, the GNE, redistributing loads to certain off-peak hours, may violate users' electricity consumption habits. In this case, coordination between agents' strategies becomes challenging [9]. The work presented in [10] and [11] investigated the optimal GNE seeking problem in the form of the minimum norm based on the Tikhonov method. Further, the work in [12] proposed a semi-decentralized method to seek the specific GNE with minimum Euclidean distance to the reference point for aggregative games, where agents' cost functions are contingent upon their individual strategies as well as an aggregation value. However, the current research lacks a sufficient characterization of optimality among multiple GNEs (e.g., the magnitude of norms sometimes fails to represent the superiority of GNEs). Furthermore, the semi-decentralized manner, assuming there is a reliable central coordinator that broadcasts/receives global information is unrealistic due to the heavy communication burden and the lack of privacy protection for agents in the networked scenario. Therefore, selecting the desired GNE, among potentially numeroues, in a distributed manner is still an open issue.

While there are few works on the distributed seeking of the optimal GNE, a recent part of research work focuses on the distributed seeking of GNE by introducing additional constraints to guarantee the existence of a unique GNE. The distributed framework and seeking strategies designed offer valuable insights and inspiration to our investigation into the optimal GNE selection. In [13], a distributed algorithm was proposed by synthesizing the projection operator with a gradient search method, which can converge to a neighborhood of the NE. The game on a directed graph was studied in [14], in which an approach for seeking NE was presented, achieved through the interconnection of projected gradient-play with average consensus dynamics. The work in [11] further considered a distributed strategy for NE over a time-varying graph via iterative Tikhonov regularization method. Considering a game with coupled constrains, a continuous-time algorithm based on a projection map method and variational inequalities was proposed in [15]. Recently, the distributed GNE seeking based on the monotone operator theory is appealing because it can ensure the global convergence with fixed update step size and its proof of convergence is concise, see [16] for a general overview. The monotone operator-based methods in [17]–[21] aim at transforming the GNE-seeking problem into a monotone operator zero-point search problem, the case in which each agent holds an estimate of other agents and maintains an

Manuscript received November 27, 2023; accepted January 29, 2024. This work was supported by the National Natural Science Foundation of China (Basic Science Center Program) (61988101), the Joint Fund of Ministry of Education for Equipment Pre-research (8091B022234), Shanghai International Science and Technology Cooperation Program (21550712400), Shanghai Pilot Program for Basic Research (22TQ1400100-3), the Fundamental Research Funds for the Central Universities, and Shanghai Artifcial Intelligence Laboratory. Recommended by Associate Editor Maojiao Ye. *(Corresponding author: Wangli He.)*

Citation: W. He and Y. Wang, "Distributed optimal variational GNE seeking in merely monotone games," *IEEE/CAA J. Autom. Sinica*, vol. 11, no. 7, pp. 1621–1630, Jul. 2024.

The authors are with the Key Laboratory of Smart Manufacturing in Energy Chemical Process, Ministry of Education, East China University of Science and Technology, Shanghai 200237, China (e-mail: wanglihe@ecust.edu.cn; yanzhenwang@mail.ecust.edu.cn).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/JAS.2024.124284

auxiliary variable used to ensure consensus of local estimations. However, most of the above works consider a particular strongly-monotone game, which assumes the game mapping F (see Section II for details) is strongly-monotone to ensure the uniqueness of GNE. Furthermore, the aforementioned algorithms either fail to converge or converge to a certain GNE, thereby falling short of fulfilling the requirement of converging to a specified optimal GNE.

Motivated by the above research results and issues, the optimal GNE seeking problem in a distributed manner is investigated. Contributions of this paper are summarized as follows:

1) To address the issue of seeking potential multiple GNEs in merely monotone games with linearly coupled cost functions, a convex function named global evaluation function is formulated to assess the quality of GNEs. Among the potential numerous GNEs, the objective is to seek the optimal GNE concerning the global evaluation function minimization.

2) A distributed optimal GNE seeking algorithm is firstly proposed. It is noteworthy that differently from [16], [21], [22], the proposed algorithm is distributed, and each agent adjusts its update step size solely based on its local information. Based on the operator theory, some convergence conditions are derived for the proposed algorithm, and convergence towards the optimal GNE with respect to the global evaluation function can be achieved.

3) To expedite the convergence speed, an algorithm, augmented with relaxation acceleration scheme, is also proposed and analyzed. While retaining the merits of the original algorithm, it is shown that, the iteration steps are significantly reduced in numerical simulation (see Section VI).

This paper is organized as follows. Section II gives the necessary preliminaries. The formulation of the generalized game and the distributed algorithm are provided in Section III. Section IV presents the motivation behind the algorithm and the convergence analysis. Then an accelerated algorithm with relaxation is presented in Section V, and simulation and analysis based on networked Nash-Cournot game are given in Section VI. Finally, Section VII provides a conclusive summary of the paper.

II. Preliminary

A. Notations

$\mathbb{R}$ ($\mathbb{R}_{\geq 0}$) is the set of (non-negative) real numbers, and let $\emptyset$ denote the empty set. $\mathbf{0}_n$ ($\mathbf{1}_n$) is a vector consisting of zeros (ones), and $I_n \in \mathbb{R}^{n\times n}$ represents an identity matrix. For a vector $x \in \mathbb{R}^n$, x^T denotes its transpose and $\|x\| = \sqrt{x^T x}$. Given a symmetric positive definite matrix Γ, the Γ-induced inner product is $\langle x, y\rangle_\Gamma = \langle \Gamma x, y\rangle$, the Γ-induced norm is $\|x\|_\Gamma = \sqrt{\langle \Gamma x, x\rangle}$. The stacked vector of N vectors $x_1, \ldots, x_N$ is represented as $\mathrm{col}(x_1, \ldots, x_N) = [x_1^T, \ldots, x_N^T]^T$. For a matrix Γ, The notations $\|\Gamma\|$ and $\|\Gamma\|_\infty$ represent the Euclidean norm and the maximum absolute row sum of Γ, respectively. The kernel space and range space of Γ is Ker(Γ) and Range(Γ). Denote the eigenvalue of Γ as $s(\Gamma)$. For $\mathcal{I} = \{1, \ldots, N\}$, $\mathrm{diag}\{(\Gamma_i)_{i\in\mathcal{I}}\}$ represents the block diagonal matrix constructed using $\Gamma_1, \ldots, \Gamma_N$ as its diagonal components. $\otimes$ denotes the kronecker product.

B. Operator Theory

Let Id be the identity operator, i.e., $Id(x) = x$. For a closed convex set $\Theta \subset \mathbb{R}^n$, the normal cone of Θ is denoted as

$$N_\Theta(v) = \begin{cases} \emptyset & v \notin \Theta \\ \{v \mid (u-v)^T w,\ \forall u \in \Theta\} & v \in \mathrm{bd}(\Theta) \\ \mathbf{0} & v \in \mathrm{int}(\Theta) \end{cases} \tag{1}$$

where bd(Θ) denotes the boundary set of Θ, and int(Θ) denotes the interior of Θ. The projection of x onto Θ is denoted as $\mathrm{proj}_\Theta(x) = \arg\min_{y\in\Theta}\|x - y\|$. Let $\mathcal{T}$ be a set-valued operator, its domain is $\mathrm{dom}(\mathcal{T}) = \{x | \mathcal{T}x \neq \emptyset\}$, and the range of $\mathcal{T}$ is $\mathrm{ran}(\mathcal{T}) = \{y | \exists\, x \in \mathrm{dom}(\mathcal{T}), y \in \mathcal{T}x\}$, the graph of $\mathcal{T}$ is $\mathrm{gra}(\mathcal{T}) = \{(x,u) | u \in \mathcal{T}x\}$, we denote its zero set as $\mathrm{zer}(\mathcal{T}) = \{x \in \mathrm{dom}(\mathcal{T}) | 0 \in \mathcal{T}x\}$ and the fixed point set is $\mathrm{fix}(\mathcal{T}) = \{x \in \mathrm{dom}(\mathcal{T}) | x \in \mathcal{T}x\}$. $\mathcal{T}$ is called monotone if $\forall (w,u)$, $(x,y) \in \mathrm{gra}(\mathcal{T})$, we have $(w-x)^T(u-y) \geq 0$. Futhermore, if $\mathrm{gra}(\mathcal{T})$ is not strictly contained within the graph of any other monotone operator, $\mathcal{T}$ is called maximally monotone. $J_\mathcal{T} = (Id + \mathcal{T})^{-1}$ is the resolvent operator of $\mathcal{T}$. $\mathcal{T}$ is Lipschitz continuous if there exists a positive constant L such that, for all pairs (w,u) and $(x,y) \in \mathrm{gra}(\mathcal{T})$, $\|u - y\| \leq L\|w - x\|$, moreover if $L = 1$, $\mathcal{T}$ is called nonexpansive. $\mathcal{T}$ is attracting nonexpansive, if $\mathrm{fix}(T) \neq \emptyset$ and $\|u - z\| < \|w - z\|$, $\forall (w,u) \in \mathrm{gra}(\mathcal{T})$, $z \in \mathrm{fix}(\mathcal{T})$. Take any (w,u), $(x,y) \in \mathrm{gra}(\mathcal{T})$, $\mathcal{T}$ is η−averaged if $\|u - y\|^2 \leq \|w - x\|^2 - \frac{1-\eta}{\eta}\|w - x - u + y\|^2$. Moreover, if $\|u - y\|^2 + \|w - x - u + y\|^2 \leq \|w - x\|^2$ holds, then $\mathcal{T}$ is firmly nonexpansive.

C. Graph Theory

The graph $\mathcal{G}(\mathcal{I}, \mathcal{E})$ represents the information exchange among a set of agents with $\mathcal{I} = \{1, 2, \ldots, N\}$, and $\mathcal{E}$ denoting the set of edges. Regarding a connected undirected graph $\mathcal{G}$, its weighted adjacency matrix is $W = [w_{ij}] \in \mathbb{R}^{N\times N}$, where $w_{i,i} = 0$, $w_{ij} \geq 0$, and $W = W^T$. The degree matrix is defined as $D = \mathrm{diag}\{(d_i)_{i\in\mathcal{I}}\}$, $d_i = \sum_{j\in\mathcal{I}} w_{ij}$, and the weighted Laplacian of $\mathcal{G}$ is $L = D - W$. It can be deducted that 0 is a simple eigenvalue of L, $L\mathbf{1}_N = \mathbf{0}_N$, $\mathbf{1}_N^T L = \mathbf{0}_N^T$ and all the remaining eigenvalues are positive. Let the eigenvalues of L in ascending order be $0 < s_2(L) \leq \cdots \leq s_N(L)$, it has $d^* \leq s_N(L) \leq 2d^*$ with $d^* = \max_{i\in\mathcal{I}}(d_i)$.

III. Game Formulation & Distributed Algorithm

A. Game Formulation

For a group of agents, in which agent $i \in \mathcal{I}$ selects its strategy from its local strategy set $\Theta_i \subseteq \mathbb{R}^{n_i}$. Define the strategy variable of agent i by $x_i \in \mathbb{R}^{n_i}$, and $\Theta = \prod_{i=1}^N \Theta_i \subseteq \mathbb{R}^n$ represents the combined strategy domain, $\boldsymbol{x} = \mathrm{col}((x_i)_{i\in\mathcal{I}})$ denotes the strategy profile, i.e., the stacked vector of all the agents' strategies where $n = \sum_{i=1}^N n_i$. The objective of each agent is to minimize its cost function $J_i(x_i, x_{-i}) : \Theta \to \mathbb{R}$, where the cost function depends on the strategy variables of the other agents, denoted as $x_{-i} = \mathrm{col}(x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_N)$, as well as the local strategy variable. This paper specifically focuses on a class of linearly coupled cost functions, which is a commonly used form in the problem of seeking GNE [23]−[25].

$$J_i(x_i, x_{-i}) = \ell_i(x_i) + \left\langle \sum_{j=1, j\neq i}^{N} C_{ij}x_j, x_i \right\rangle \tag{2}$$

in which $\ell_i(x_i): \mathbb{R}^{n_i} \to \mathbb{R}$ denotes the local cost function of agent i, and $C_{ij} \in \mathbb{R}^{n_i \times n_j}$ signifies the impact of agent j's strategy on the cost function of agent i. In addition, we assume strategies are coupled together via a shared affine constraints and the overall feasible space can be written as

$$\mathcal{X} = \Theta \cap \{\boldsymbol{x} \in \mathbb{R}^n \mid A\boldsymbol{x} \leq b\} \tag{3}$$

in which $A = [A_1, A_2, \ldots, A_N]$ and $b = \sum_{i=1}^{N} b_i$. Specifically, $A_i \in \mathbb{R}^{m \times n_i}$, $b_i \in \mathbb{R}^m$ and Θ_i are the local information of agent i. The GNE seeking problem is then formulated that the agent finds the optimal response from the feasible space after giving the strategies of other agents, which may not be known by any agents. Therefore, this interconnected coupled optimization problem is given by

$$\min_{x_i \in \Theta_i} J_i(x_i, x_{-i}) \quad \text{s.t. } A_i x_i \leq b - \sum_{j=1, j\neq i}^{N} A_j x_j,\ \forall i \in \mathcal{I}. \tag{4}$$

Assumption 1: The local function $\ell_i(x_i)$ is lower semicontinuous and convex. The set Θ_i in (3) is nonempty, compact, and convex. Furthermore, Θ satisfies the Slater's constraint qualification condition, for all $i \in \mathcal{I}$.

Then, the pseudo gradient mapping (game mapping) is defined as

$$F(\boldsymbol{x}) = \begin{bmatrix} \partial_{x_1} J_1(x_1, \boldsymbol{x}_{-1}) \\ \vdots \\ \partial x_N J_N(x_N, \boldsymbol{x}_{-N}) \end{bmatrix}. \tag{5}$$

Definition 1: Generalized Nash equilibrium (GNE) $\boldsymbol{x}^* = col((x_i^*)_{i\in\mathcal{I}})$ ensures that each agent satisfies the following condition:

$$f_i\left(x_i^*, x_{-i}^*\right) \leq \inf\left\{f_i\left(y, x_{-i}^*\right) \mid y \in \mathcal{X}_i\left(x_{-i}^*\right)\right\}. \tag{6}$$

The local Lagrangian for agent i can be denoted as $L_i(x_i, \lambda_i, x_{-i}) = J_i(x_i, x_{-i}) + \lambda_i^T(A\boldsymbol{x} - b)$ with $\lambda_i \in \mathbb{R}_+^m$, $x_i \in \Theta_i$. Under Assumption 1, $\boldsymbol{x}^*$ is the optimal solution to problem (4), if there is a λ^* that satisfies the Karush-Kuhn-Tucher (KKT) condition: $\mathbf{0}_{n_i} = \nabla_{x_i} L_i(x_i^*, \lambda_i^*, x_{-i}^*)$, $x_i^* \in \Theta_i$, $\langle \lambda_i^*, A\boldsymbol{x}^* - b\rangle = 0$, $-A\boldsymbol{x}^* - b \geq \mathbf{0}$. Applying normal cone, KKT condition can be rewritten as

$$\begin{cases} \mathbf{0}_{n_i} \in \mathrm{N}_{\Theta_i}\left(x_i^*\right) + A_i^T\lambda_i^* + \nabla_{x_i} J_i\left(x_i^*, x_{-i}^*\right) \\ \mathbf{0}_m \in -(A\boldsymbol{x}^* - b) + \mathrm{N}_{\mathbb{R}_+^m}\left(\lambda_i^*\right). \end{cases} \tag{7}$$

In this paper, our primary focus lies on a specific subset of GNE, namely variational GNE (v-GNE). v-GNE solves the variational inequality (VI): $(F(\boldsymbol{x}))^T(u - \boldsymbol{x}) \geq 0,\ \forall u \in \mathcal{X}$. The KKT condition of v-GNE can be concisely written in a compact form as follows:

$$\begin{cases} \mathbf{0}_n \in F(\boldsymbol{x}^*) + A^T\lambda^* + \mathrm{N}_\Theta(\boldsymbol{x}^*) \\ \mathbf{0}_m \in -(A\boldsymbol{x}^* - b) + \mathrm{N}_{\mathbb{R}_+^m}(\lambda^*) \end{cases} \tag{8}$$

in which $F(\boldsymbol{x}^*)$ denotes the pesudo gradient defined in (5) and $\mathrm{N}_\Theta(\boldsymbol{x}^*) = \prod_{i=1}^{N} \mathrm{N}_{\Theta_i}(x_i^*)$.

By comparing (7) with (8), when $\boldsymbol{x}^*$ and λ^* satisfy the KKT condition (7) and all multipliers are the same, i.e., $\mathrm{col}((\lambda_i)_{i\in\mathcal{I}}) = \mathbf{1}_N \otimes \lambda^*$, they also satisfy (8), hence v-GNE is a subclass of GNE. Furthermore, the v-GNE set is nonempty, compact and convex under Assumption 1 [26, proposition 2.2].

Assumption 2: The matrices $C_{ij} \in \mathbb{R}^{n_i \times n_j}$ in (2) satisfy $C_{ij} = C_{ji}^T, \forall i, j \in \mathcal{I}$.

Remark 1: In recent literature, it is commonly assumed that the pseudo gradient exhibits Lipschitz continuity and strongly monotonicity, which can ensure the uniqueness of v-GNE in the game [27]. However, under Assumptions 1, only the existence of v-GNE can be ensured, while the uniqueness is not enough to be guaranteed, which results in the existence of multiple solutions to (8).

Among the potential v-GNEs, one needs to select one with more ideal characteristics (according to the strategy variable $\boldsymbol{x}$). Naturally, a global evaluation function, denoted as ϕ, to measure the performance of different v-GNE is required. Selecting the optimal v-GNE according to the global evaluation function is called the optimal v-GNE selection problem in this paper, which can be formulated as

$$\begin{cases} \arg\min_{\boldsymbol{x}\in\mathbb{R}^n} \phi(\boldsymbol{x}) \\ \text{s.t. } \exists(\boldsymbol{x}, \lambda) \text{ satisfying (8).} \end{cases} \tag{9}$$

Assumption 3: The global evaluation function ϕ is convex, differential, and separable with respect to the variable $\boldsymbol{x}$, that is, $\phi(\boldsymbol{x}) = \sum_{i\in\mathcal{I}} \phi_i(x_i)$. Additionally, the gradient with respect to $\boldsymbol{x}$ is $L_{\nabla\phi}$-Lipschitz continuous.

Remark 2: Problem (8) can be translated, through the definition of the normal cone, into the task of finding the solution to $VI(\mathcal{F}, \Delta)$, denoted as $\mathrm{Sol}(\mathcal{F}, \Delta)$, where $\mathcal{F} = \mathrm{col}(F(x) + A^T\lambda, b - Ax)$ and $\Delta = \Theta \times \mathbb{R}_+^m$. Furthermore, problem (9) can be equivalently reformulated as the search for a solution to $VI(\nabla\phi, Sol(\mathcal{F}, \Delta))$ by following a similar set of steps. Once again, making use of proposition 2.2 in [26], and under the Assumption 3, it can be established that $\mathrm{Sol}(\nabla\phi, Sol(\mathcal{F}, \Delta))$ is a non-empty, compact and convex set, implying that the optimal v-GNE set is not empty.

B. Distributed Optimal Equilibrium Seeking

In this section, we introduce the fundamental algorithm developed for distributed searching of the optimal v-GNE. Specifically, assuming that each agent possesses only the knowledge of its individual cost function J_i and feasible space Θ_i, and the corresponding portion of coupling constrains (A_i, b_i). Furthermore, each agent updates its own strategy by communicating with neighbors through the two communication topological networks denoted by $\mathcal{G}_f$ and $\mathcal{G}_\lambda$. Graph $\mathcal{G}_f$, called interference graph, is defined by the relationship between the agent's cost function and the strategies of agents. For $\mathcal{G}_f(\mathcal{I}, \mathcal{E}_f)$, $(j,i) \in \mathcal{E}_f$ only if the strategy of agent j will directly affect the agent i's cost function, the interference neighbours of agent i can be denoted as $\mathcal{N}_i^f = \{j \mid (j,i) \in \mathcal{E}_f\}$. We adopt the classical framework established in [28] and [29], which assumes that each agent has access to the strategies of other agents whose cost function directly influences its own. The

agent can also exchange local information through multiplier graphs $\mathcal{G}_\lambda$, agent i can receive information from agent j if $j \in \mathcal{N}_i^\lambda : \{j | (j,i) \in \mathcal{E}_f\}$.

Assumption 4: The multiplier graph $\mathcal{G}_\lambda$ is undirected and connected.

Remark 3: The interference graph $\mathcal{G}_f$ may be a sparse graph when the agent's cost function depends on only a fraction of the agents (e.g., the Nash-Cournot game in [28]). If each agent's cost function is contingent upon the strategies chosen by all other agents, the interference graph $\mathcal{G}_f$ is a complete graph (e.g., aggregative game in [23])

The proposed distributed optimal v-GNE algorithm is outlined in Algorithm 1, in which each agent independently manages its local strategy x_i and Lagragian multiplier λ_i. Furthermore, to handle the limited information about other agents, agent i also maintains a local auxiliary variable σ_i to ensure compliance with the global affine constraint and achieve consensus among λ_i, $i \in \mathcal{I}$. Here, γ_i, β_i, and δ_i represent the constant step-sizes for agent i. Additionally, $\mu^{(k)}$ corresponds to the global decay step, which will be elaborated later. The variables x_i^k, σ_i^k, and λ_i^k denote the values of x_i, σ_i, and λ_i at iteration k respectively. Moreover, $[w_{ij}]$ represents the weighted adjacency matrix of $\mathcal{G}_\lambda$.

Algorithm 1 Distributed Optimal v-GNE Seeking

Intialization: $\forall i \in \mathcal{I}$, $x_i^0 \in \Theta_i$, $\sigma_i^0 \in \mathbf{0}_m$, $\lambda_i^0 \in \mathbb{R}_{\geq 0}^m$

Local update, for $i \in \mathcal{I}$:

1: *Phase 1:* Receive $x_j^k, j \in \mathcal{N}_i^f, \lambda_j^k, j \in \mathcal{N}_i^\lambda$ and update

2: $x_i^{k+\frac{1}{2}} = \arg\min_{y \in \Theta_i} \left(\frac{\gamma_i}{2} \left\| y - x_i^k \right\|^2 + l_i(y) + \left\langle y, A_i^T \lambda_i^k + \sum_{j \in \mathcal{I} \setminus \{i\}} C_{ij} x_j^k \right\rangle \right)$

3: $x_i^{k+1} = x_i^{k+\frac{1}{2}} - \mu^{(k)} \nabla_{x_i} \phi \left(x_i^{k+\frac{1}{2}} \right)$

4: $\sigma_i^{k+1} = \sigma_i^k + \beta_i \sum w_{ij} \left(\lambda_i^k - \lambda_j^k \right)$

5: *Phase 2:* Receive $\sigma_j^{k+1}, j \in \mathcal{N}_i^\lambda$ and update

6: $\lambda_i^{k+1} = \text{proj}_{\mathbb{R}_{\geq 0}^m} \left(\lambda_i^k + \delta_i \left(A_i \left(2x_i^{k+\frac{1}{2}} - x_i^k \right) - \sum w_{ij} \left(2\sigma_i^{k+1} - \sigma_i^k - 2\sigma_j^{k+1} + \sigma_j^k \right) + b_i \right) \right)$

Assumption 5: The sequence $(\mu^{(k)})_{k \in \mathbb{N}} \in \mathbb{R}_{\geq 0}$ satisfies $\lim_{k \to \infty} \mu^{(k)} = 0$, $\sum_{k>0} \mu^{(k)} = \infty$, and $\sum_{k>0} (\mu^{(k)})^2 < \infty$.

Remark 4: Algorithm 1 is called a double-layer algorithm because a complete state update requires two rounds of communication among agents. Each iteration k alternates between "communication" and "computation", in phase 1, agent i gets x_{-i}^k from the neighbors $j \in \mathcal{N}_i^f$ in interference graph $\mathcal{G}_f$, meanwhile, agent i gets λ_j^k, $j \in \mathcal{N}_i^\lambda$ and update δ_j^k. Then agent i observes latest information σ_j^{k+1}, $j \in \mathcal{N}_i^\lambda$ and updates λ_i^k by one-step projection.

The auxiliary variable σ_i can be interpreted as a discrete-time integrator for the consensual errors, ensuring the ultimate consensus of Lagrangian multipliers λ_i. In addition, σ_i can also reflect the contribution of agent i to the coupling constraints, and plays the role of global affine constraint decoupling. Algorithm 1 is called distributed because: 1) Each agent is required to store its own local information. 2) No central node is required to receive, broadcast, and update multiplier information (differently from [12], [22]).

IV. Algorithm Development & Convergence Analysis

A. Algorithm Development

In this section, we will explore the design motivation behind Algorithm 1 and systematically illustrate that how to write Algorithm 1 in the form of the preconditioned proximal point algorithm (PPPA) and the hybrid steepest descent method (HSDM).

Lemma 1: Algorithm 1 is equivalent to

$$\begin{cases} \mathcal{T} \varpi^k = J_{\Gamma^{-1} \mathcal{T}_e} \left(\varpi^k \right) \\ \varpi^{k+1} = \mathcal{T} \varpi^k - \mu^{(k)} \nabla \phi \left(\mathcal{T} \varpi^k \right), \ \forall k \in \mathbb{N} \end{cases} \tag{10}$$

with $\varpi^k = col(\boldsymbol{x}^k, \sigma^k, \lambda^k)$, $x^k = col((x_i^k)_{i \in \mathcal{I}})$, $\sigma^k = col((\sigma_i^k)_{i \in \mathcal{I}})$ and $\lambda^k = col((\lambda_i^k)_{i \in \mathcal{I}})$. Γ is called the precondition matrix which is defined as

$$\Gamma = \begin{bmatrix} \bar{\gamma} - C & \mathbf{0} & -\mathfrak{A}^T \\ \mathbf{0} & \bar{\beta}^{-1} & L_m \\ -\mathfrak{A} & L_m & \bar{\delta}^{-1} \end{bmatrix} \tag{11}$$

where $\bar{\gamma} = \text{diag}\{(\gamma_i I_{n_i})_{i \in \mathcal{I}}\}$, $\bar{\beta} = \text{diag}\{(\beta_i I_m)_{i \in \mathcal{I}}\}$ and $\bar{\delta} = \text{diag}\{(\delta_i I_m)_{i \in \mathcal{I}}\}$. $\mathfrak{A} = \text{diag}\{(A_i)_{i \in \mathcal{I}}\}$, $\bar{b} = col((b_i)_{i \in \mathcal{I}})$ and $L_m = L \otimes I_m$. Define operation $\mathcal{T}_e$ as

$$\mathcal{T}_e \varpi = \begin{bmatrix} N_\Theta(\boldsymbol{x}) + F(\boldsymbol{x}) + \mathfrak{A}^T \lambda \\ -L_m \lambda \\ N_{\mathbb{R}_+^{Nm}}(\lambda) - \left(\mathfrak{A} \boldsymbol{x} - \bar{b} \right) + L_m \sigma \end{bmatrix}. \tag{12}$$

Proof: Firstly, (10) can be split in two-step iterations

$$\varpi^{k+\frac{1}{2}} = J_{\Gamma^{-1} \mathcal{T}_e} \left(\varpi^k \right) \tag{13}$$

$$\varpi^{k+1} = \varpi^{k+\frac{1}{2}} - \mu^{(k)} \nabla \phi \left(\varpi^{k+\frac{1}{2}} \right). \tag{14}$$

By the definition of the resolvent operator, (13) can be rewritten as $\Gamma(\varpi^k - \varpi^{k+\frac{1}{2}}) \in \mathcal{T}_e \varpi^{k+\frac{1}{2}}$. Then, we simplify each row in operator $\mathcal{T}_e$. In the first row of the inclusion, one has that

$$\begin{aligned} &(\bar{\gamma} - c) \left(\boldsymbol{x}^k - \boldsymbol{x}^{k+\frac{1}{2}} \right) - \mathfrak{A}^T \left(\lambda^k - \lambda^{k+\frac{1}{2}} \right) \in N_\Theta \left(\boldsymbol{x}^{k+\frac{1}{2}} \right) \\ &\quad + (G + C) \boldsymbol{x}^{k+1} + \mathfrak{A}^T \lambda^{k+\frac{1}{2}} \\ &\Leftrightarrow \mathbf{0} \in \bar{\gamma} \left(\boldsymbol{x}^{k+\frac{1}{2}} - \boldsymbol{x}^k \right) + C \boldsymbol{x}^k + \mathfrak{A}^T \lambda^k \\ &\quad + N_\Theta \left(\boldsymbol{x}^{k+\frac{1}{2}} \right) + G \boldsymbol{x}^{k+\frac{1}{2}} \\ &\underset{\forall i \in \mathcal{I}}{\Leftrightarrow} \mathbf{0} \in \gamma_i \left(x_i^{k+\frac{1}{2}} - x_i^k \right) + \sum_{j \in \mathcal{I} \setminus i} C_{ij} x_j^k + A_i^T \lambda_i^k \\ &\quad + N_{\Theta_i} \left(x_i^{k+\frac{1}{2}} \right) + \nabla_{x_i} l_i \left(x_i^{k+\frac{1}{2}} \right) \\ &\underset{\forall i \in \mathcal{I}}{\Leftrightarrow} \mathbf{0} \in \nabla_{\substack{y \\ y \in \Theta_i}} \left(\frac{\gamma_i}{2} \| y - x_i^k \|^2 + \left\langle y, \sum_{i \in \mathcal{I} \setminus i} C_{ij} x_j^k \right. \right. \\ &\quad \left. \left. + A_i^T \lambda_i^k \right\rangle + l_i(y) \right). \end{aligned}$$

To maintain a concise proof process, operator G denotes $\mathrm{diag}\{(\nabla_{x_i} l_i)_{i\in\mathcal{I}}\}$, and $C\in\mathbb{R}^{n\times n}$ is a block matrix with C_{ij} defined in (2). Obviously, the above problem is to search the zero points of the gradient in a strongly convex function, which corresponds to the unique minima and can be rewritten as

$$x_i^{k+\frac{1}{2}} = \arg\min_{y\in\Theta_i}\left(\frac{\gamma_i}{2}\left\|y-x_i^k\right\|^2 + \left\langle y, \sum_{j\in I\setminus i} C_{ij}x_j^k + A_i^T\lambda_i^k\right\rangle + l_i(y)\right). \tag{15}$$

Similarly, simplify the second row of the operator $\mathcal{T}_e$ as

$$\begin{aligned}&\bar{\beta}^{-1}\left(\sigma^k-\sigma^{k+\frac{1}{2}}\right)+L_m\left(\lambda^k-\lambda^{k+\frac{1}{2}}\right)=-L_m\lambda^{k+\frac{1}{2}}\\ &\Leftrightarrow_{i\in\mathcal{I}}\ \sigma_i^{k+\frac{1}{2}}=\sigma_i^k+\beta_i\sum_{j\in\mathcal{I}}w_{ij}\left(\lambda_i^k-\lambda_j^k\right).\end{aligned} \tag{16}$$

Using the normal cones and projection operators, the last row of the operation $\mathcal{T}_e$ can be simplified as

$$\begin{aligned}&\mathbf{0}\in\bar{\delta}\left(\mathfrak{A}\left(\boldsymbol{x}^k-2\boldsymbol{x}^{k+\frac{1}{2}}\right)+\bar{b}+L_m\left(2\sigma^{k+\frac{1}{2}}-\sigma^k\right)\right)\\ &\quad-\lambda^k+\lambda^{k+\frac{1}{2}}+N_{\mathbb{R}_+^{Nm}}\left(\lambda^{k+\frac{1}{2}}\right)\\ &\Leftrightarrow_{i\in\mathcal{I}}\ \lambda_i^{k+\frac{1}{2}}=\mathrm{proj}_{\mathbb{R}_+^m}\left(\lambda_i^k-\delta_i\left(A_i\left(x_i^k-x_i^{k+\frac{1}{2}}\right)+b_i\right.\right.\\ &\quad\left.\left.+\sum_{j\in\mathcal{I}}w_{ij}\left(2\sigma_i^{k+\frac{1}{2}}-\sigma_i^k-2\sigma_j^{k+\frac{1}{2}}+\sigma_j^k\right)\right)\right).\end{aligned} \tag{17}$$

Finally, because the global evaluation function ϕ is set to search the optimal v-GNE dedicated to strategy variable $\boldsymbol{x}$, (14) is equivalent to

$$\begin{cases}\boldsymbol{x}^{k+1}=\boldsymbol{x}^{k+\frac{1}{2}}-\mu^{(k)}\nabla_{\boldsymbol{x}}\phi\left(\boldsymbol{x}^{k+\frac{1}{2}}\right)\\ \sigma^{k+1}=\sigma^{k+\frac{1}{2}}\\ \lambda^{k+1}=\lambda^{k+\frac{1}{2}}.\end{cases} \tag{18}$$

The conclusion follows by substituting (15)−(17) into (18). ■

Remark 5: It is essential to emphasize that matrix Γ plays two important roles: 1) Since the operator $\mathcal{T}_e$ contains a skew-symmetric matrix, its resolvent cannot be computed without matrix Γ. 2) It can enable the algorithm to be executed in a distributed manner, in which each agent chooses its own step size only based on its local information, which can be realized in practice.

Next, we will discuss the properties of operator $\mathcal{T}$ and operator $\mathcal{T}_e$, and illustrate that the problem (9) to be solved can be further transformed by operators.

Lemma 2: Assuming that Assumptions 1−4 are satisfied, the statements below can be obtained.

1) $\forall\ \mathrm{col}(x^*,\sigma^*,\bar{\lambda}^*)\in\mathrm{zer}(\mathcal{T}_e)$, $\bar{\lambda}^*=\mathbf{1}_N\otimes\lambda^*$ and x^*, λ^* solves (8), hence x^* is a v-GNE of the game (4). Moreover, $\mathrm{col}(\boldsymbol{x}^*,\sigma^*,\bar{\lambda}^*)\in\mathrm{fix}(\mathcal{T})$.

2) $\mathrm{zer}(\mathcal{T}_e)\neq\emptyset$ and $\mathrm{fix}(\mathcal{T})\neq\emptyset$.

Proof:

1) Given $\mathrm{col}(\boldsymbol{x}^*,\sigma^*,\bar{\lambda}^*)$ is a zero point of the operator $\mathcal{T}_e$, by the definition of the operator $\mathcal{T}_e$ in (12), one has that

$$\begin{bmatrix} N_\Theta(\boldsymbol{x}^*)+F(\boldsymbol{x}^*)+\mathfrak{A}^T\bar{\lambda}^*\\ -L_m\bar{\lambda}^*\\ N_{\mathbb{R}_+^{Nm}}\left(\bar{\lambda}^*\right)-\left(\mathfrak{A}\boldsymbol{x}-\bar{b}\right)+L_m\sigma^*\end{bmatrix}=\begin{bmatrix}\mathbf{0}\\ \mathbf{0}\\ \mathbf{0}\end{bmatrix}. \tag{19}$$

The second line of (19) implies $\bar{\lambda}^*\in\mathrm{Ker}(L_m)\Leftrightarrow\bar{\lambda}^*=\mathbf{1}_N\otimes\lambda^*,\lambda^*\in\mathbb{R}^m$. Then by the fact that $\lambda_1^*=\cdots=\lambda_N^*=\lambda^*$ and $\mathfrak{A}=\mathrm{diag}\{(A_i)_{i\in\mathcal{I}}\}$, the first line of (19) is equivalent to the first line of the v-GNE KKT condition (8). Finally, combining the fact that $\sum_N^{i=1}N_{\Theta_i}\Leftrightarrow N_{\cap_{i=1}^N\Theta_i}$ when $\cap_{i=1}^N\Theta_i\neq\emptyset$ and multiplying $\mathbf{1}_N^T\otimes I_m$ in the last line of (19), it is equivalent to the last line of the v-GNE KKT condition (8). Hence, $\mathrm{col}(\boldsymbol{x}^*,\lambda^*)$ is a solution to the $VI(F,\mathcal{X})$ and $\boldsymbol{x}^*$ is a v-GNE. Furthermore, by the definition of resolvent operator, one has that

$$\begin{aligned}\mathbf{0}\in\mathcal{T}_e(\varpi^*)&\Leftrightarrow\Gamma(\varpi^*-\varpi^*)\in\mathcal{T}_e(\varpi^*)\\ &\Leftrightarrow\varpi^*=J_{\Gamma^{-1}\mathcal{T}_e}(\varpi^*)\end{aligned} \tag{20}$$

which implies that $\varpi^*=\mathrm{col}(\boldsymbol{x}^*,\sigma^*,\bar{\lambda}^*)$ is also a fixed point of operation $\mathcal{T}$.

2) In Remark 1, we have concluded that the v-GNE set of the game (4) is non-empty. Hence, there is at least one point $\mathrm{col}(x^*,\lambda^*)$ satisfying the KKT condition (8). Taking $\lambda_1^*=\lambda_2^*=\cdots=\lambda_N^*=\lambda^*$, the first and second lines in (18) naturally hold. Next, we will prove that there exists $\sigma^*=\mathrm{col}((\sigma_i^*)_{i\in\mathcal{I}})$ satisfying the last line of (19). Firstly, there must exist $z^*\in N_{\mathbb{R}_+^m}(\lambda^*)$ such that $-(A\boldsymbol{x}^*-b)+z^*=\mathbf{0}$. Take $z_1^*=z_2^*=\cdots=z_N^*=\frac{1}{N}z^*$, then one has that

$$\begin{aligned}&\left(\mathbf{1}_N^T\otimes I_m\right)\left[-\left(\mathfrak{A}\boldsymbol{x}^*-\bar{b}\right)+L_m\bar{\lambda}^*+\mathrm{col}\left(\left(z_i^*\right)_{i\in\mathcal{I}}\right)\right]\\ &\quad=-(A\boldsymbol{x}^*-b)+z^*=\mathbf{0}\end{aligned} \tag{21}$$

which means $-(\mathfrak{A}\boldsymbol{x}^*-\bar{b})+L_m\bar{\lambda}^*+\mathrm{col}((z_i^*)_{i\in\mathcal{I}})\in\mathrm{Ker}(\mathbf{1}_N^T\otimes I_m)$. By the fact that $\mathrm{Ker}(\mathbf{1}_N^T\otimes I_m)=\mathrm{Range}(L_m)$ and $\mathrm{col}((z_i^*)_{i\in\mathcal{I}})\in N_{\mathbb{R}_{\geq0}^{Nm}}$, there exists a $\sigma^*\in\mathrm{Range}(L_m)$ satisfying the last line of (19). ■

Based on Lemma 2, the optimal v-GNE selection problem (9) can be reformulated as

$$\begin{cases}\arg\min_{\boldsymbol{x}\in\mathbb{R}^n}\phi(\boldsymbol{x})\\ \text{s.t. }\mathrm{col}(\boldsymbol{x},\sigma,\lambda)\in\mathrm{fix}(\mathcal{T})\end{cases} \tag{22}$$

and feasible space of the problem (22) is non-empty. The form of problem (22) is called the fixed-point selection problem that was investigated in [30]. We will illustrate that the problem can be solved via HSDM in the next section.

B. Convergence Analysis

Before proving that Algorithm 1 can converge towards the solution to (22), it is imperative to discuss the properties of operators $\mathcal{T}$ and $\mathcal{T}_e$ required in the HDSM based on the rearrangement of Algorithm 1.

Lemma 3: Assuming Assumptions 1−4 hold, the operator $\Gamma^{-1}\mathcal{T}_e$ is the maximally monotone, and $\mathcal{T}$ is attracting operator in the Γ-induced norm if matrix Γ is positive definite.

Proof: Operator $\mathcal{T}_e$ can be split into three terms

$$\mathcal{T}_e(\varpi)=\begin{bmatrix} N_\Theta(\boldsymbol{x}) \\ \mathbf{0} \\ N_{\mathbb{R}_+^{Nm}} \end{bmatrix}+\begin{bmatrix} F(\boldsymbol{x}) \\ \mathbf{0} \\ \bar{b} \end{bmatrix}+\begin{bmatrix} \mathbf{0} & \mathbf{0} & \mathfrak{A}^T \\ \mathbf{0} & \mathbf{0} & -L_m \\ -\mathfrak{A} & L_m & \mathbf{0} \end{bmatrix}\varpi.$$

For the first term on the right-hand side (RHS), because norm crone with respect to a closed convex set is maximally monotone [31, Example 20.26], operators N_Θ and $N_{\mathbb{R}_+^{Nm}}$ are both the maximally monotone. Moreover the constant vector is also maximally monotone, and the direct sum of maximally monotone operators remains maximally monotone [31, Proposition 20.23], so the first term on the RHS is the maximally monotone. Under Assumption 1, the epigraph of $J_i, i\in\mathcal{I}$ is closed. Consider the fact that the subgradient of the convex, closed and proper function is the maximally monotone [31, Theorem 20.25], the second term on the RHS is also maximally monotone. Maximally monotonicity of the third term on the RHS follows from the [31, Example 20.35]. Finally, as the domains of the second and third terms on the RHS are $\mathbb{R}^n$, $\mathbb{R}^{n+2Nm}$ respectively, their sum results in a maximally monotone operator.

Assume Γ is positive definite. $\forall\ (x,u)\in \mathrm{gra}(\Gamma^{-1}\mathcal{T}_e)$ and $(y,v)\in \mathrm{gra}(\Gamma^{-1}\mathcal{T}_e)$, we have that $\Gamma u\in\Gamma^{-1}\Gamma\mathcal{T}_e x$ and $\Gamma v\in\Gamma^{-1}\Gamma\mathcal{T}_e y$, which implies that

$$\langle x-y,u-v\rangle_\Gamma=\langle x-y,\Gamma(u-v)\rangle\geq 0. \tag{23}$$

Equation (23) holds due to the maximally monotonicity of $\mathcal{T}_e$, hence $\Gamma\mathcal{T}_e$ is the maximally monotone in the Γ-induced norm.

Consider $\mathcal{T}$ as the resolvent of the maximally monotone operator $\Gamma^{-1}\mathcal{T}_e$, then $\mathcal{T}$ is firmly nonexpansive [31, Proposition 23.7], and $\mathcal{T}\in\mathcal{A}\left(\frac{1}{2}\right)$ from [31, Proposition 4.25]. Therefore, Lemma 3 holds by the fact that firmly nonexpansive operators are also attracting nonexpansive [32, Lemma 2.4]. ■

Lemma 4: The matrix Γ is positive definite if every agent follows Algorithm 1 with fixed step-sizes β_i, δ_i, and γ_i in Algorithm 1 satisfying

$$\begin{cases} 0<\beta_i<(2d_i)^{-1} \\ \gamma_i>\displaystyle\sum_{j\in\mathcal{I}}\|C_{ij}\|_\infty+\|A_i\|_\infty \\ 0<\delta_i<(2d_i+\|A_i\|_\infty)^{-1}. \end{cases} \tag{24}$$

Proof: Obviously, matrix Γ is symmetric. Conclusion can be easily drawn by constructing a strictly diagonally dominant matrix, its detailed proof follows the Gershgorin's theorem [33, Corollary 6.2.27]. ■

Remark 6: It should be noted that, in [17], [22], the setting of step-sizes for each agent required the coercivity constant of the pseudo-gradient, indicating that the cost function of each agent could not be safeguarded. However, according to Lemma 4, agent i determines its own step-sizes only based on local information (A_i, C_{ij}, and d_i), which ensures Algorithm 1 can be executed in a distributed manner.

The main result of the convergence of Algorithm 1 can be obtained as follows.

Theorem 1: Suppose Assumptions 1−5 hold, and step-sizes γ_i, β_i, δ_i satisfy Lemma 4. Let Θ^* denote the solution set of the problem (22), and $(\varpi^k)_{k\in\mathbb{N}}=\mathrm{col}(\boldsymbol{x}^k,\sigma^k,\lambda^k)$ is the sequence generated by Algorithm 1. Then one has

$$\lim_{k\to\infty} dist\left(\varpi^k,\Theta^*\right)=0. \tag{25}$$

Proof: First, (22) can be transformed into the general form that can be solved by HSDM using the definition of variational inequality. Consider

$$\boldsymbol{x}^*=\arg\min_{\boldsymbol{x}\in\mathbb{R}^n}\Leftrightarrow\langle v-\boldsymbol{x}^*,\nabla\phi(\boldsymbol{x}^*)\rangle\geq 0,\ \forall v\in\mathbb{R}^n \tag{26}$$

one has the equivalent form of problem (22)

$$\begin{cases} \text{Find } \varpi^*=\mathrm{col}(\boldsymbol{x}^*,\sigma^*,\lambda^*), & \varpi^*\in\mathrm{fix}(\mathcal{T}) \\ \text{s.t.}\ \langle v-\varpi^*,\nabla_{\boldsymbol{x}}\phi(\varpi^*)\rangle\geq 0, & \forall v\in\mathrm{fix}(\mathcal{T}). \end{cases} \tag{27}$$

Under Assumption 1, Θ is compact, which ensures that $\mathrm{fix}(\mathcal{T})$ is nonempty and bounded. Moreover, $\nabla\phi$ is Lipschitz continuous and monotone over the domain of the operator $\mathcal{T}$ under Assumption 3. From Lemma 3, $\mathcal{T}$ is attracting operator. Algorithm 1 (or (10)) is the standard form of the HSDM, and fulfills all assumptions of [34, Theorem 3]. The result immediately ensues. ■

V. Accelerations

In this section, we employ the relaxation in Algorithm 1 to accelerate the convergence speed. Motivated by the convergence property of Algorithm 1, which relies on the monotone operator $\mathcal{T}$, various modifications have been proposed to enhance its convergence speed. These modifications involve constructing the next iterate by combining the operator's output with previous outputs. Specifically, we focus on a modification scheme called relaxation, which has shown pretty acceleration performance in distributed optimization algorithms [35]. While the introduction of a relaxation scheme poses challenges to the convergence of algorithms based on HSDM, we firstly prove that the convergence can be ensured by properly selecting the relaxation step-size within the operator framework.

Algorithm 2 Distributed Optimal v-GNE Seeking With Relaxation

Intialization: $\forall i\in\mathcal{I}$, $x_i^0\in\Theta_i$, $\sigma_i^0\in\mathbf{0}_m$, $\lambda_i^0\in\mathbb{R}_+^m$

Local update, for $i\in\mathcal{I}$:

1: *Phase 1:* Receive $x_j^k, j\in\mathcal{N}_i^f, \lambda_j^k, j\in\mathcal{N}_i^\lambda$ and update

2: $x_i^{k+\frac{1}{2}}=\arg\min_{y\in\Theta_i}\left(\frac{\gamma_i}{2}\left\|y-x_i^k\right\|^2+l_i(y)\right.$
$\left.+\left\langle y,A_i^T\lambda_i^k+\sum_{j\in\mathcal{I}\setminus i}C_{ij}x_j^k\right\rangle\right)$

3: $\hat{x}_i^{k+1}=x_i^k+\eta\left(x_i^{k+\frac{1}{2}}-x_i^k\right)$

4: $x_i^{k+1}=\hat{x}_i^{k+1}-\mu^{(k)}\nabla_{x_i}\phi\left(\hat{x}_i^{k+1}\right)$

5: $\sigma_i^{k+1}=\sigma_i^k+\eta\beta_i\sum w_{ij}\left(\lambda_i^k-\lambda_j^k\right)$

6: *Phase 2:* Receive $\sigma_j^{k+1}, j\in\mathcal{N}_i^\lambda$ and update

7: $\lambda_i^{k+\frac{1}{2}}=\mathrm{proj}_{\mathbb{R}_+^m}\left(\lambda_i^k+\delta_i\left(A_i\left(2x_i^{k+\frac{1}{2}}-x_i^k\right)\right.\right.$
$\left.\left.-\sum w_{ij}\left(2\sigma_i^{k+1}-\sigma_i^k-2\sigma_j^{k+1}+\sigma_j^k\right)+b_i\right)\right)$

8: $\lambda_i^{k+1}=\lambda_i^k+\eta\left(\lambda_i^{k+\frac{1}{2}}-\lambda_i^k\right)$

Remark 7: Compared to Algorithm 1, Algorithm 2 is established through adding an additional relaxation step. Following the exact proof steps in Lemma 1, $\forall k\in\mathbb{N}$, Algorithm 2

can be rewritten in a compact form of

$$\begin{cases} \varpi^{k+\frac{1}{2}} = \mathcal{T}\varpi^k \\ \mathring{\varpi}^{k+1} = \varpi^k + \eta\left(\varpi^{k+\frac{1}{2}} - \varpi^k\right) \\ \varpi^{k+1} = \mathring{\varpi}^{k+1} - \mu^{(k)}\nabla\phi\left(\mathring{\varpi}^{k+1}\right) \end{cases} \tag{28}$$

where operator $\mathcal{T}$ is defined in (10), and η is a positive relaxation parameter.

Currently, relaxation scheme is primarily investigated for the case of firmly nonexpansive operators. However, in this paper, the attracting property of operator $\mathcal{T}$ is essential to ensure the convergence of the HSDM-based algorithms. Introducing relaxation mechanisms poses challenges in maintaining attracting property of $\mathcal{T}$. The following result demonstrates the convergence of Algorithm 2 under the correct choice for step size η as well as γ_i, β_i, δ. We omit part of the proof because it is similar to that of Theorem 1.

Theorem 2: Suppose Assumptions 1−5 hold. Take the $\eta \in (0, 3/2)$ and each agent choose its step-sizes γ_i, β_i, δ_i to satisfy Lemma 4. Θ^* denotes the set of the solutions to problem (22), and $(\varpi^k)_{k\in\mathbb{N}} = \mathrm{col}(\boldsymbol{x}^k, \sigma^k, \lambda^k)$ is the sequence generated by Algorithm 2 such that

$$\lim_{k\to\infty} dist\left(\varpi^k, \Theta^*\right) = 0. \tag{29}$$

Proof: Recall the compact form (28) of Algorithm 2. Introduce operator $\mathcal{T}' = \eta\mathcal{T} + (1-\eta)Id$ to simplify (28) as

$$\begin{cases} \mathcal{T}'\varpi^k = \eta\mathcal{T}\varpi^k + (1-\eta)\varpi^k \\ \varpi^{k+1} = \mathcal{T}'\varpi^k - \mu^{(k)}\nabla\phi\left(\mathcal{T}\varpi^k\right),\ \forall k \in \mathbb{N} \end{cases} \tag{30}$$

which is the standard form of HSDM based on the operator $\mathcal{T}'$. We need to prove that the operator $\mathcal{T}'$ satisfies the properties required by HSDM.

Given any $\varpi^* = \mathrm{col}(\boldsymbol{x}^*, \delta^*, \lambda^*) \in \mathrm{fix}(\mathcal{T}')$, one can obtain

$$\mathcal{T}'\varpi^* = \eta\mathcal{T}\varpi^* + \varpi^* - \eta\varpi^* = \varpi^* \Leftrightarrow \mathcal{T}\varpi^* = \varpi^* \tag{31}$$

hence ϖ^* is v-GNE of game (4) from Lemma 2. Moreover, considering $\mathcal{T}$ is firmly nonexpansive from Lemma 3, there exists a nonexpansive $\breve{\mathcal{T}}$ satisfying $\mathcal{T} = \frac{1}{2}Id + \frac{1}{2}\breve{\mathcal{T}}$. Therefore $\mathcal{T}'$ can be rewritten as

$$\begin{aligned} \mathcal{T}' &= \eta\left(\frac{1}{2}Id + \frac{1}{2}\breve{\mathcal{T}}\right) + (1-\eta)Id \\ &= \frac{\eta}{2}\breve{\mathcal{T}} + \left(1 - \frac{\eta}{2}\right)Id \end{aligned} \tag{32}$$

which implies $\mathcal{T}' \in \mathcal{A}(\eta/2)$ if $\eta \in (0, 3/2)$, or $\mathcal{T}'$ is attracting nonexpansive. The remainder of the proof follows a similar approach as in Theorem 1. ■

Remark 8: Algorithm 2 or (30) can alternatively be viewed as the robust hybrid steepest descent method (RHSDM), which has been proven that it is gifted with notable numerical robustness in [36]. The relaxation method is also applied to the preconditioned-forward-backward (P-FB) method [17]. In particular, the advantages of our algorithm is that relaxation parameter η can be correctly chosen globally and independently i.e., the choice of η does not require any knowledge of game mapping F in (8) (e.g., Lipschitz constant), which ensures Algorithm 2 is in a distributed manner.

Remark 9: In addition, we consider a special case where the derivative of the global evaluation function is strongly monotonic with respect to $\eta_{\nabla\phi}$, and each agent has access to $\eta_{\nabla\phi}$ and $L_{\nabla\phi}$. If the algorithm incurs a high computational cost and the optimal relaxation parameter is unknown, we consider adopting time-varying relaxation parameter, denoted as $\eta_k \in \{\eta_1, \eta_2, \ldots, \eta_N\}$, where $\eta_i \in (0, 3/2)$, $\forall i \in \{1, 2, \ldots, N\}$, to achieve a trade-off in acceleration effects. Specifically, $\eta_k = \eta_{[k]_N}$, where $[k]_N = \{k - iN | i = 0, 1, \ldots, \} \cap \{1, 2, \ldots, N\}$. Similarly to the proof of Theorem 2, at each update instant, $\mathcal{T}'_k = \eta_k\mathcal{T} + (1-\eta_k)Id$ is a nonexpansive operator, and $\mathrm{fix}(\mathcal{T}'_N, \mathcal{T}'_{N-1}, \ldots, \mathcal{T}'_1) = \mathrm{fix}(\mathcal{T}'_{N-1}, \mathcal{T}'_{N-2}, \ldots, \mathcal{T}'_1\mathcal{T}'_N) = \cdots = \mathrm{fix}(\mathcal{T}'_1, \mathcal{T}'_2, \ldots, \mathcal{T}'_N) = \mathrm{fix}(\mathcal{T})$. According to Theorem 2, it can be easily demonstrated that under the Assumptions 1−4 and $\mu^{(k)} = \mu_0/k$, where $\mu_0 \in \left(0, \frac{2\eta\nabla\phi}{L\nabla\phi^2}\right)$, Algorithm 2 with the adoption of time-varying relaxation parameter satisfies all assumptions of [36, Theorem 3.3], and thus still converges to the unique optimal v-GNE.

VI. Numerical Simulations

In this section, the advantages of our algorithms with the modified simulation in [12] are demonstrated. In order to ensure the results more convincing, the aggregation game is extended to the more general game model. A networked Nash-Cournot game with $N = 6$ companies (agents) participating in the production of a homogeneous commodity, competing across $M = 3$ markets is considered (see Fig. 1).

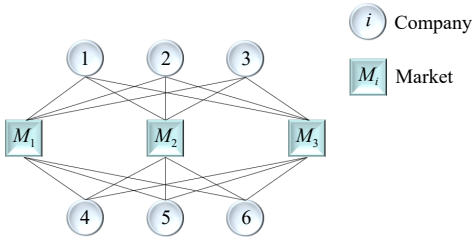


Fig. 1. Networked cournot game: A undirected edge from i to M_j in the graph represent company i participate in the competition in market M_j.

The multiplier graph $\mathcal{G}_\lambda$ of the companies refers to Fig. 2. For company i, the number of markets it is engaged in is represented as $n_i = M = 3$. Company i should decide on the quantities $x_i \in \mathbb{R}^{n_i}$ to minimize its cost function J_i, which is set as

$$J_i(x_i, x_{-i}) = q_i^T x_i + \frac{1}{N}x_i^T \bar{C}_i x_i + \left\langle \sum_{j\in\mathcal{I}\setminus i} C_{ij}x_j, x_i \right\rangle. \tag{33}$$

$\bar{C}_i$ and C_{ij} are both set as diagonal matrices, whose non-zero elements are randomly sampled in [0, 1], and q_i is randomly sampled in [−10, 0]. The local constraint $\Theta_i = \prod_{j=1}^{3}[a_{ij}, 100]$ represents the upper and lower bounds of goods production for company i, in which $a_{i,j}$ is set from [−1, 1]. Moreover, the global constraint in (3) is set as $A_i = I_3$ and $b_i = 5\mathbf{1}_3$, which denotes each market has the same maximal capacity. We set the global evaluation function as

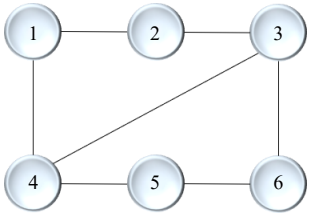


Fig. 2. Multiplier graph $\mathcal{G}_\lambda$: An edge from i to j on this graph implies that information can be received and transmitted between i and j.

$$\phi(x)=\sum_{i=1}^{N} x_i^T Q_i x_i+\tilde{q}_i^T x_i \tag{34}$$

in which Q_i is set as a diagonal matrix whose nonzero elements are randomly chosen in [1, 2], and the components of $\tilde{q}_i$ are stochastically sampled from the interval $[-1, 1]$. Finally, we take the step-sizes for each company as $\beta_i=0.15$, $\gamma_i=2$, $\delta_i=2$ and $\eta=1.4$, decay step $\mu^{(k)}$ is defined by $\mu^{(k)}=\mu_0/k^{0.8}$, with μ_0 =0.1.

The global evaluation function in (34) is set to reduce the sum of the distances between the strategy variables of each agent and randomly generated reference points. Furthermore, the pseudo-gradient of the game model can be written as $F(\boldsymbol{x})=\mathrm{col}(F_i(\boldsymbol{x}))_{i\in\mathcal{I}}$, where $F_i(\boldsymbol{x})=q_i^T+\frac{1}{N}\overline{C}_i x_i+\sum_{j\in\mathcal{I}\setminus i}C_{ij}x_j$. It is easy to prove that the setting of the parameters satisfies Assumptions 1−5.

We conduct a performance comparison of Algorithms 1 and 2 versus PPPA in [19] under the same parameters. The primary outcome is demonstrated in Fig. 3, where all three algorithms are initialized with an identical initial condition. It is shown the v-GNE computed by the PPPA is suboptimal with respect to the global evaluation function ϕ, and our Algorithms can converge to the optimal v-GNE with lower ϕ value. Moreover, Fig. 3 shows Algorithm 2 has the superior convergence speed with help of relaxation, which can also be seen from the convergence curve of agents' strategy variables in Fig. 4.

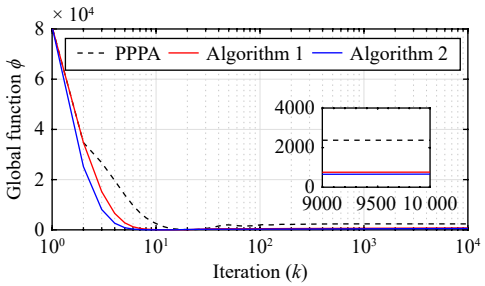
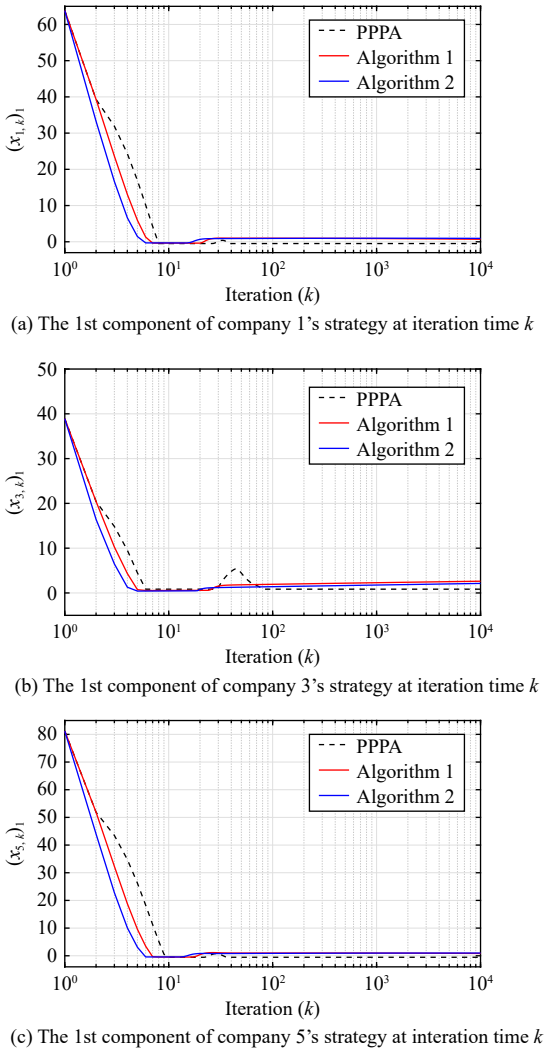


Fig. 3. Comparison of Algorithms 1 and 2 with PPPA method in terms of the global function ϕ.

In Fig. 5, the trajectory of $\|L\otimes I_m*\lambda^k\|^2$ represents consensus of each company's local multiplier, it can be observed Algorithms 1 and 2 can ensure the local multiplier of each company tends to be same. Finally, the strategy variable x_i^k convergence of each company can be obtained from Fig. 6.

The convergence speed of the algorithm is significantly determined by the rate at which the Lagrange multipliers



(a) The 1st component of company 1's strategy at iteration time k

(b) The 1st component of company 3's strategy at iteration time k

(c) The 1st component of company 5's strategy at interation time k

Fig. 4. The trajectories of the strategies of company 1, 3 and 5.

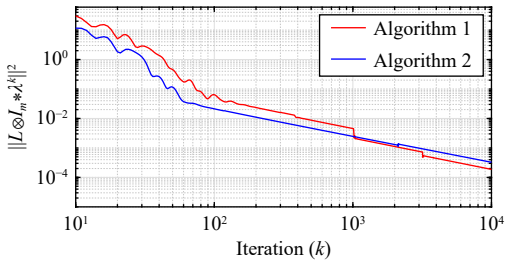
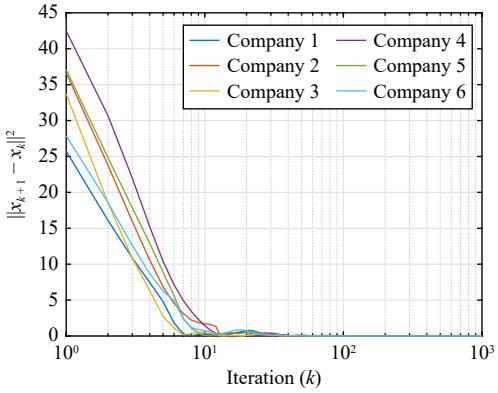


Fig. 5. The trajectory of $\|L\otimes I_m*\lambda^k\|^2$ shows that λ^k converge to the Ker$(L\otimes I_m)$, which implies all the multipliers eventually converge to a consensus.

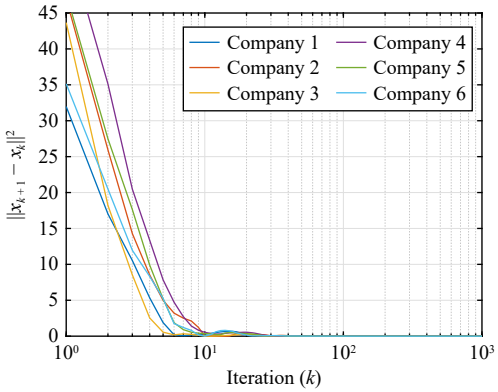
achieve consensus. Therefore, we compared the impact of different relaxation factors on the convergence of the multipliers. The results are depicted in Fig. 7, where it can be observed that within the permissible range, larger relaxation factors lead to a faster achievement of consensus among the multipliers of agents, resulting in a faster convergence of the algorithm.

VII. Conclusion

The optimal GNE seeking problem with affine coupling constraints for merely monotone games is studied in this



(a) The trajectories of strategy variables generated by Algorithm 1



(b) The trajectories of strategy variables generated by Algorithm 2

Fig. 6. Evolution of each company's strategy variable x_i^k.

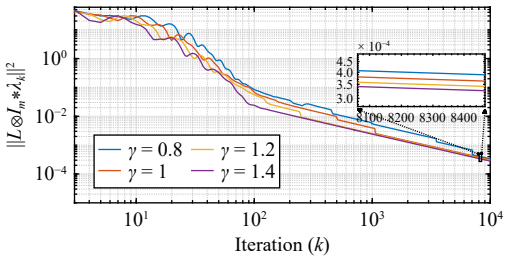


Fig. 7. The effect of Algorithm 2 employing different relaxation factors on the achievement of consensus among multipliers.

paper. A distributed algorithm, based on HSDM, is proposed to select the optimal GNE with respect to the given global evaluation function. Further, a relaxation scheme is introduced on the basis of the original algorithm for faster convergence. We have proved its global convergence under mild assumptions based on monotone operator theory, and verified the effectiveness in numerical simulation.

In the future, we will study the fully-distributed methods, where the utilization of strategy information that influences its own cost function is not permissible for each agent. The most of existing fully-distributed algorithms assume that their game mappings are restricted-monotone, which cannot ensure the global convergence of HSDM-based algorithms.

REFERENCES

[1] X. Na and D. J. Cole, "Modelling of a human driver s interaction with vehicle automated steering using cooperative game theory," *IEEE/CAA J. Autom. Sinica*, vol. 6, no. 5, pp. 1095–1107, Sept. 2019.

[2] M. Ye and G. Hu, "Game design and analysis for price-based demand response: An aggregate game approach," *IEEE Trans. Cybernetics*, vol. 47, no. 3, pp. 720–730, Mar. 2017.

[3] B. Li, Q. Yang, and I. Kamwa, "A novel Stackelberg-game-based energy storage sharing scheme under demand charge," *IEEE/CAA J. Autom. Sinica*, vol. 10, no. 2, pp. 462–473, Feb. 2023.

[4] N. Li, L. Chen, and M. A. Dahleh, "Demand response using linear supply function bidding," *IEEE Trans. Smart Grid*, vol. 6, no. 4, pp. 1827–1838, Jul. 2015.

[5] D. Paccagnan, B. Gentile, F. Parise, M. Kamgarpour, and J. Lygeros, "Distributed computation of generalized Nash equilibria in quadratic aggregative games with affine coupling constraints," in *Proc. 55th IEEE Decision and Control*, Las Vegas, USA, Dec. 2016, pp. 6123–6128.

[6] J. Wang, Y. Hong, J. Wang, J. Xu, Y. Tang, Q.-L. Han, and J. Kurths, "Cooperative and competitive multi-agent systems: From optimization to games," *IEEE/CAA J. Autom. Sinica*, vol. 9, no. 5, Apr. 2022.

[7] S. Le Cleac'h, M. Schwager, and Z. Manchester, "ALGAMES: A fast augmented Lagrangian solver for constrained dynamic games," *Autonomous Robots*, vol. 46, no. 1, pp. 201–215, Jan. 2022.

[8] X. Chen, E. Dall'Anese, C. Zhao, and N. Li, "Aggregate power flexibility in unbalanced distribution systems," *IEEE Trans. Smart Grid*, vol. 11, no. 1, pp. 258–269, Jan. 2020.

[9] C. Li, X. Yu, W. Yu, G. Chen, and J. Wang, "Efficient computation for sparse load shifting in demand side management," *IEEE Trans. Smart Grid*, vol. 8, no. 1, pp. 250–261, Jan. 2017.

[10] H. Yin, U. V. Shanbhag, and P. G. Mehta, "Nash equilibrium problems with scaled congestion costs and shared constraints," *IEEE Trans. Autom. Control*, vol. 56, no. 7, pp. 1702–1708, Jul. 2011.

[11] J. Lei, U. V. Shanbhag, and J. Chen, "Distributed computation of Nash equilibria for monotone aggregative games via iterative regularization," in *Proc. 59th IEEE Decision and Control*, Jeju, Korea (South), Dec. 2020, pp. 2285–2290.

[12] E. Benenati, W. Ananduta, and S. Grammatico, "On the optimal selection of generalized Nash equilibria in linearly coupled aggregative games," in *Proc. 61st IEEE Decision and Control*. Cancun, Mexico, Dec. 2022, pp. 6389–6394.

[13] M. Ye, Q.-L. Han, L. Ding, S. Xu, and G. Jia, "Distributed Nash equilibrium seeking strategies under quantized communication," *IEEE/CAA J. Autom. Sinica*, vol. 11, no. 1, pp. 103–112, Jan. 2024.

[14] Y. Zhu, W. Yu, G. Wen, and G. Chen, "Distributed Nash equilibrium seeking in an aggregative game on a directed graph," *IEEE Trans. Autom. Control*, vol. 66, no. 6, pp. 2746–2753, Jun. 2021.

[15] G. Chen, Y. Ming, Y. Hong, and P. Yi, "Distributed algorithm for ε-generalized Nash equilibria with uncertain coupled constraints," *Automatica*, vol. 123, p. 109313, Jan. 2021.

[16] G. Belgioioso, P. Yi, S. Grammatico, and L. Pavel, "Distributed generalized Nash equilibrium seeking: An operator-theoretic perspective," *IEEE Control Systems Magazine*, vol. 42, no. 4, pp. 87–102, Aug. 2022.

[17] P. Yi and L. Pavel, "An operator splitting approach for distributed generalized Nash equilibria computation," *Automatica*, vol. 102, pp. 111–121, Apr. 2019.

[18] S. Liang, P. Yi, and Y. Hong, "Distributed Nash equilibrium seeking for aggregative games with coupled constraints," *Automatica*, vol. 85, pp. 179–185, Nov. 2017.

[19] M. Bianchi, G. Belgioioso, and S. Grammatico, "Fast generalized Nash equilibrium seeking under partial-decision information," *Automatica*, vol. 136, p. 110080, Feb. 2022.

[20] L. Shi and W. He, "Generalized Nash equilibrium seeking for networked noncooperative games with a dynamic event-triggered mechanism," *Applied Mathematical Modelling*, vol. 118, pp. 39–52, Jun. 2023.

[21] D. Gadjov and L. Pavel, "Single-timescale distributed GNE seeking for aggregative games over networks via forward backward operator splitting," *IEEE Trans. Autom. Control*, vol. 66, no. 7, pp. 3259–3266, Jul. 2021.

[22] E. Benenati, W. Ananduta, and S. Grammatico, "Optimal selection and tracking of generalized Nash equilibria in monotone games," *IEEE Trans. Autom. Control*, vol. 68, no. 12, pp. 7644–7659, Dec. 2023.

[23] G. Belgioioso and S. Grammatico, "Semi-decentralized generalized Nash equilibrium seeking in monotone aggregative games," *IEEE Trans. Autom. Control*, vol. 68, no. 1, pp. 140–155, Jan. 2023.

[24] G. Belgioioso and S. Grammatico, "A distributed proximal-point algorithm for Nash equilibrium seeking in generalized potential games with linearly coupled cost functions," in *Proc. 18th European Control Conf.*, Naples, Italy, Aug. 2019, pp. 1–6.

[25] B. Franci, M. Staudigl, and S. Grammatico, "Distributed forward-backward (half) forward algorithms for generalized Nash equilibrium seeking," in *Proc. 19th European Control Conf.*, St. Petersburg, Russia, Jul. 2020, pp. 1274–1279.

[26] F. Facchinei, A. Fischer, and V. Piccialli, "On generalized Nash games and variational inequalities," *Operations Research Letters*, vol. 35, no. 2, pp. 159–164, Mar. 2007.

[27] M. Ye, Q.-L. Han, L. Ding, and S. Xu, "Distributed Nash equilibrium seeking in games with partial decision information: A survey," *Proc. the IEEE*, vol. 111, no. 2, pp. 140–157, 2023.

[28] C.-K. Yu, M. van der Schaar, and A. H. Sayed, "Distributed learning for stochastic generalized Nash equilibrium problems," *IEEE Trans. Signal Processing*, vol. 65, no. 15, pp. 3893–3908, Apr. 2017.

[29] M. Zhu and E. Frazzoli, "Distributed robust adaptive equilibrium computation for generalized convex games," *Automatica*, vol. 63, pp. 82–91, Jan. 2016.

[30] A. Cegielski, A. Gibali, S. Reich, and R. Zalas, "An algorithm for solving the variational inequality problem over the fixed point set of a quasi-nonexpansive operator in euclidean space," *Numerical Functional Analysis and Optimization*, vol. 34, no. 10, pp. 1067–1096, Aug. 2013.

[31] H. H. Bauschke and P. L. Combettes, *Convex Analysis and Monotone Operator Theory in Hilbert Spaces*. Springer, 2011, vol. 408.

[32] H. H. Bauschke and J. M. Borwein, "On projection algorithms for solving convex feasibility problems," *SIAM Review*, vol. 38, no. 3, pp. 367–426, Jan. 1996.

[33] R. A. Horn and C. R. Johnson, *Matrix Analysis*. Cambridge University press, 2012.

[34] N. Ogura and I. Yamada, "Nonstrictly convex minimization over the bounded fixed point set of a nonexpansive mapping," *Numerical Functional Analysis and Optimization*, vol. 24, no. 1–2, pp. 129–135, Aug. 2003.

[35] F. Iutzeler and J. M. Hendrickx, "A generic online acceleration scheme for optimization algorithms via relaxation and inertia," *Optimization Methods and Software*, vol. 34, no. 2, pp. 383–405, Nov. 2019.

[36] I. Yamada, N. Ogura, and N. Shirakawa, "A numerically robust hybrid steepest descent method for the convexly constrained generalized inverse problems," *Contemporary Mathematics*, vol. 313, pp. 269–305, Jan. 2002.

Wangli He (Senior Member, IEEE) received the B.Sc. degree in information and computing science and the Ph.D. degree in applied mathematics from Southeast University in 2005 and 2010, respectively. From 2010 to 2017, she held several visiting positions with Central Queensland University Australia, the University of Hong Kong, Hong Kong, China, the City University of Hong Kong, Hong Kong, China, Potsdam Institute for Climate Research Institute, Germany, and Tokyo Metropolitan University, Japan. She is currently a Professor with the Key Laboratory of Smart Manufacturing in Energy Chemical Process, Ministry of Education, East China University of Science and Technology. Her current research interests include distributed coordination control and optimization, networked multiagent systems, autonomous intelligent unmanned systems, and industrial cyber-physical systems. Dr. He was a recipient of the Sixth Young Scientist Award of Chinese Association of Automation in 2020, and the first prize of Shanghai Natural Science Award in 2019. She was the Chair of the Technical Committee on Networked-based Control Systems and Applications of IES from 2018 to 2019. She is an Associate Editor for several international journals including the *IEEE Transactions on Industrial Informatics* and the *IEEE Journal of Emerging and Selected Topics in Industrial Electronics*.

Yanzhen Wang received the B.Sc. degree in automation from East China University of Science and Technology in 2022, where he is currently pursuing the M.Sc. degree in control science and engineering. His current research interests include multiagent systems, game theory, and operator theory.