

Online distributed NE seeking for multi-agent noncooperative games with time-varying delays

Lingfei Su¹, Yongzhao Hua^{1,*}, Xiwang Dong^{1,2,3}, Jinhu Lü^{3,4}, Danwei Wang⁵

¹School of Artificial Intelligence (Institute of Artificial Intelligence), Beihang University, Beijing 100191, China

²Institute of Unmanned System, Beihang University, Beijing 100191, China

³School of Automation Science and Electrical Engineering, Beihang University, Beijing 100191, China

⁴Zhongguancun Laboratory, Beijing 100094, China

⁵School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore 639798, Singapore

ABSTRACT

This paper investigates noncooperative online games in multi-agent systems with time-delay feedback constraints. In these games, all players individually minimize their own time-varying costs, yet these costs are coupled through the joint strategies of the entire group. The gradient feedback available to the player is subject to unknown, heterogeneous delays that vary across players and time instants. To address these challenges, a distributed online gradient-descent algorithm is developed to seek the time-varying Nash equilibrium in the game. The algorithm is proven to achieve a sublinear dynamic regret bound via an appropriate learning rate, provided that the path variations grow sublinearly. Finally, an online electricity market game with time-varying delays is provided to illustrate the effectiveness of the algorithm.

KEYWORDS

noncooperative games; Nash equilibrium; online distributed algorithm; multi-agent systems; time-varying delays

Distributed algorithms in multi-agent systems have garnered substantial research attention across numerous domains, including formation control^[1], wireless sensor networks^[2], and smart grids^[3]. Within the realm of multi-agent noncooperative games, the Nash equilibrium (NE)^[4] is a strategy combination such that, given the fixed strategies of all other players, no single player can reduce its own cost by deviating from its chosen strategy. Extensive research has investigated distributed NE and generalized NE (GNE) seeking algorithms under varied problem settings, including coupled constraints^[5-6], multi-coalition structures^[7-8], and prescribed-time convergence requirements^[9-10].

The above works^[5-11] address offline game scenarios under the assumptions of known analytical cost functions and a stationary game environment. These assumptions are restrictive in dynamic settings such as online mobile sensor networks^[12], where sensors require real-time position adaptation to unpredictable target motion and a priori unknown environmental disturbances. Recent work^[13-17] on online noncooperative games generally relies on the simplification of a fixed one-step feedback delay, where players receive cost feedback after each decision. Multi-step delays have been addressed in Refs. [18-19], but they are assumed to be fixed, uniform across players, and known a priori. In practice, delays fluctuate over time due to environmental shifts and uneven player performance, while heterogeneity further introduces asymmetric delay profiles. This motivates the recent work^[20] on unknown time-varying delays, in which static regret analysis for

online optimization problems has been provided.

In this paper, online noncooperative games for multi-agent systems, where the feedback delays are unknown and time-varying, are investigated. A distributed algorithm for seeking the NE is designed, together with an accompanying analysis of its dynamic regret. The primary contributions of this work are outlined in the following content.

1) Noncooperative online games with unknown time-varying delays are considered over directed unbalanced topologies. Classical online game problems^[14-17] assume a one-step feedback delay, and fixed multi-step delays are considered in Refs. [18-19]. The unknown time-varying delay setting has wider applicability to real online scenarios.

2) An online distributed NE seeking algorithm is developed with an appropriate learning rate to ensure sublinear dynamic regret. The derived regret upper bound explicitly captures the effects of delays and consensus estimation errors. When reduced to the one-step delay case, the sublinear regret order matches that in Ref. [13].

Notations: The set of real numbers is defined by $\mathbb{R}$, the q -dimensional Euclidean space is represented by $\mathbb{R}^q$, and real matrices are denoted by $\mathbb{R}^{q \times q}$. The set of nonnegative integers is written as $\mathbb{Z}_{\geq 0}$. $\mathbb{E}$ denotes the expectation. $\text{col}(\cdot)$ stacks its arguments into a column vector. For a given matrix $\mathbf{A}$ and a vector $\mathbf{x}$, $\mathbf{A}^T$ and $\mathbf{x}^T$ denote their transpose. $\text{diag}(\cdot)$ treats its arguments as the diagonal elements of a diagonal matrix. The inner product in the

Received: June 15, 2026; Revised: July 10, 2026; Accepted: July 20, 2026

*Yongzhao Hua, yongzhaohua@buaa.edu.cn

Citation: Su L F, Hua Y Z, Dong X W, et al. Online distributed NE seeking for multi-agent noncooperative games with time-varying delays[J]. *C2 Science and Technology*, 2026. <https://doi.org/10.26599/C2ST.2026.9700002>.

© Chinese Institute of Command and Control. Publishing services by Tsinghua University Press. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

Euclidean space is formulated as $\langle \cdot, \cdot \rangle$. The cardinality of a set or the absolute value of a scalar constant is represented by $|\cdot|$. The Euclidean norm is $\|\cdot\|$. The Kronecker product is indicated by $\otimes$. $\max$ means the maximum. $\lfloor \cdot \rfloor$ represents the floor function. The gradient of a differentiable function $J(\mathbf{a}, \mathbf{b})$ with respect to $\mathbf{a}$ is denoted by $\nabla J(\mathbf{a}, \mathbf{b}) = \partial J(\mathbf{a}, \mathbf{b}) / \partial \mathbf{a}$, where ∂ stands for the partial derivative operator. The projection onto a set X is represented by $\mathcal{P}_X$. The $q \times q$ identity matrix is $\mathbf{I}_q$, and $\mathbf{1}_q$ denotes the q -dimensional all-one vector, where their subscripts can be omitted without causing any confusion. For a square matrix $\mathbf{A}$, $\rho(\mathbf{A})$ stands for its spectral radius. $\mathcal{O}(f)$ characterizes the upper bound of the asymptotic growth rate of f , $o(f)$ describes a strictly slower asymptotic growth rate relative to f , and $\Theta(f)$ describes the tight asymptotic order of f . For any $a \geq 0$, $\lfloor 1/a \rfloor_+ = \lfloor 1/a \rfloor$ if $a > 0$, and $\lfloor 1/a \rfloor_+ = 0$ otherwise.

1 Preliminaries

1.1 Graph theory

Let $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ be a directed network, with $\mathcal{V} = \{1, 2, \dots, N\}$ being the collection of N players and $\mathcal{E} \subseteq \mathcal{V} \times \mathcal{V}$ being the collection of directed links. A link $(i, j) \in \mathcal{E}$ is interpreted as player j sending data to player i . Strong connectivity of the network $\mathcal{G}$ is defined by the existence of a directed link sequence from i to j for any players in $\mathcal{G}$. The set of in-neighbors of player i is then expressed as $\mathcal{N}_i = \{j \in \mathcal{V} : (i, j) \in \mathcal{E}\}$.

1.2 Problem formulation

Consider an N -player online noncooperative game over the directed network $\mathcal{G}$. At time $t \in \mathbb{Z}_{\geq 0}$, every player $i \in \mathcal{V}$ chooses a decision $\mathbf{x}_{i,t} \in \mathcal{M}_i$, with $\mathcal{M}_i \in \mathbb{R}^q$ its feasible set and q its decision dimension. Let $\mathbf{x}_{-i,k} = \text{col}(\mathbf{x}_{1,k}, \mathbf{x}_{2,k}, \dots, \mathbf{x}_{i-1,k}, \mathbf{x}_{i+1,k}, \mathbf{x}_{i+2,k}, \dots, \mathbf{x}_{N,k})$ denote the joint decisions of all players except i . After the decisions are made, an adversary reveals to each player i delayed feedback regarding the local cost function $J_{i,k}(\mathbf{x}_{i,k}, \mathbf{x}_{-i,k}) : \mathbb{R}^{Nq} \rightarrow \mathbb{R}$ for $k \in \mathcal{T}_{i,t} = \{0 \leq k \leq t : k + \tau_{i,k} - 1 = t\}$, where $\tau_{i,k} \in \mathbb{Z}_{>0}$ is the unknown time delay during which player i makes a decision at time k . Each player i then tackles the following minimization problem:

$$\min_{\mathbf{x}_{i,t} \in \mathcal{M}_i} J_{i,t}(\mathbf{x}_{i,t}, \mathbf{x}_{-i,t}) \quad (1)$$

A decision profile $\mathbf{x}_t^* = (\mathbf{x}_{1,t}^*, \mathbf{x}_{2,t}^*, \dots, \mathbf{x}_{N,t}^*)$ is an NE of the game (1) if

$$J_{i,t}(\mathbf{x}_{i,t}^*, \mathbf{x}_{-i,t}^*) \leq J_{i,t}(\mathbf{x}_{i,t}, \mathbf{x}_{-i,t}^*) \quad (2)$$

holds for any $(\mathbf{x}_{i,t}, \mathbf{x}_{-i,t}^*) \in \mathcal{M} = \mathcal{M}_1 \times \dots \times \mathcal{M}_N$. In line with Ref. [21], the variational NE is defined as any $\mathbf{x}_t^*$ that resolves the following inequality problem

$$(\mathbf{F}_t(\mathbf{x}_t^*))^\top (\mathbf{x} - \mathbf{x}_t^*) \geq 0, \forall \mathbf{x} \in \mathcal{M} \quad (3)$$

where $\mathbf{F}_t(\mathbf{x}_t) = \text{col}(\nabla J_{1,t}(\mathbf{x}_{1,t}, \mathbf{x}_{-1,t}), \nabla J_{2,t}(\mathbf{x}_{2,t}, \mathbf{x}_{-2,t}), \dots, \nabla J_{N,t}(\mathbf{x}_{N,t}, \mathbf{x}_{-N,t}))$ is the pseudo-gradient, and $\mathbf{x}_t = \text{col}(\mathbf{x}_{1,t}, \mathbf{x}_{2,t}, \dots, \mathbf{x}_{N,t})$.

Motivated by Refs. [13–16], the dynamic regret is introduced to measure the performance of online algorithms as follows.

Definition 1: Consider the game (1) over the time horizon $T \in \mathbb{Z}_{>0}$. The dynamic regret of player $i \in \mathcal{V}$ is defined as

$$R_{i,T} = \sum_{t=1}^T (J_{i,t}(\mathbf{x}_{i,t}, \mathbf{x}_{-i,t}^*) - J_{i,t}(\mathbf{x}_{i,t}^*, \mathbf{x}_{-i,t}^*)) \quad (4)$$

In Eq. (4), $R_{i,T}$ measures the accumulated difference between

the cost incurred by the online decisions and the cost incurred by the time-varying NE over the horizon T .

A no-regret online-decision-making procedure is defined when the regret exhibits sublinear growth with respect to T , namely $\lim_{T \rightarrow \infty} R_{i,T}/T = 0$.

Assumption 1: The network $\mathcal{G}_t$ is guaranteed to be strongly connected.

Assumption 2: For each player $i \in \mathcal{V}$, the corresponding cost function and the feasible set are required to meet the properties listed below:

- 1) The feasible set $\mathcal{M}_i$ is nonempty, compact, and convex.
- 2) The cost function $J_{i,t}(\mathbf{x}_{i,t}, \mathbf{x}_{-i,t})$ is differentiable with respect to $\mathbf{x}_t$ and convex in $\mathbf{x}_{i,t}$ under any fixed $\mathbf{x}_{-i,t}$.
- 3) The gradient $\nabla J_{i,t}(\mathbf{x}_{i,t}, \mathbf{x}_{-i,t})$ is characterized by L -Lipschitz continuity.
- 4) The pseudo-gradient $\mathbf{F}_t(\mathbf{x}_t)$ satisfies the condition of μ -strong monotonicity.

Remark 1: Assumption 1 ensures that information can be transmitted across the network at any time. Assumption 2 is standard and widely adopted in NE seeking problems (see Refs. [13–16, 22–24]).

2 Main results

An online distributed algorithm executed by each player $i \in \mathcal{V}$ is presented in Algorithm 1. The core idea is as follows. Each player i receives the set of delayed local gradients $\{\nabla J_{i,k} : k \in \mathcal{T}_{i,t}\}$, the neighboring decisions $\{\mathbf{x}_{j,t} : j \in \mathcal{N}_{i,t}\}$, and the neighboring estimates $\{\hat{\mathbf{x}}_{j,t} : s \in \mathcal{V} \setminus \{j\}, j \in \mathcal{N}_{i,t}\}$. If the delayed feedback is nonempty, i.e., $|\mathcal{T}_{i,t}| \neq 0$, player i updates its decision $\mathbf{x}_{i,t+1}$ via gradient descent using the delayed feedback; otherwise, $\mathbf{x}_{i,t+1}$ remains unchanged at $\mathbf{x}_{i,t}$. Finally, player i updates its estimate $\hat{\mathbf{x}}_{j,t+1}$ of each other player $j \in \mathcal{V} \setminus \{i\}$ via a leader-follower consensus-based mechanism using the neighboring information.

Algorithm 1 Online distributed seeking algorithm with delayed gradient feedback

Input: Initial decisions $\mathbf{x}_{i,0} \in \mathcal{M}_i$ and estimates $\hat{\mathbf{x}}_{j,0} \in \mathcal{M}_j, \forall j \in \mathcal{V} \setminus \{i\}$; time horizon T ; learning rate $\alpha_{i,t}$; tracking gains $k_1, k_2 > 0$; weight matrix $\mathbf{A} = [a_{ij}]$.

Output: $\mathbf{x}_{i,t}, t = 1, 2, \dots, T$.

1: for $t = 0, 1, \dots, T-1$ do

2: Receive delayed gradients $\{\nabla J_{i,k}(\mathbf{x}_{i,k}, \hat{\mathbf{x}}_{-i,k}) : k \in \mathcal{T}_{i,t}\}$ and neighboring information $\{\mathbf{x}_{j,t}, \hat{\mathbf{x}}_{j,t} : j \in \mathcal{N}_{i,t}\}$.

3: if $|\mathcal{T}_{i,t}| \neq 0$ then

$$4: \quad \mathbf{z}_{i,t+1} = \mathcal{P}_{\mathcal{M}_i} \left[\mathbf{x}_{i,t} - \frac{\alpha_{i,t}}{|\mathcal{T}_{i,t}|} \sum_{k \in \mathcal{T}_{i,t}} \nabla J_{i,k}(\mathbf{x}_{i,k}, \hat{\mathbf{x}}_{-i,k}) \right]$$

5: else

$$6: \quad \mathbf{z}_{i,t+1} = \mathbf{x}_{i,t}$$

7: end if

$$8: \quad \mathbf{x}_{i,t+1} = (1 - \alpha_{i,t})\mathbf{x}_{i,t} + \alpha_{i,t}\mathbf{z}_{i,t+1}$$

$$9: \quad \hat{\mathbf{x}}_{j,t+1} = \hat{\mathbf{x}}_{j,t} - \left[k_1 \sum_{s=1}^N a_{is}(\hat{\mathbf{x}}_{j,t} - \hat{\mathbf{x}}_{s,t}) + k_2 a_{ij}(\hat{\mathbf{x}}_{j,t} - \mathbf{x}_{j,t}) \right], \forall j \in \mathcal{V} \setminus \{i\}$$

In Algorithm 1, $0 < \alpha_{i,t} < 1$ is a learning rate to be determined. $k_1, k_2 > 0$ is tracking gains to be determined. $a_{ij} \geq 0$ is the (i, j) th element of the weight matrix $\mathbf{A} \in \mathbb{R}^{N \times N}$, where $a_{ij} > 0$ if $(i, j) \in \mathcal{E}$; otherwise, $a_{ij} = 0$.

Let $\hat{\mathbf{x}}_{i,t} = \text{col}(\hat{\mathbf{x}}_{i1,t}, \hat{\mathbf{x}}_{i2,t}, \dots, \hat{\mathbf{x}}_{iN,t})$ be the joint estimates of all players except i , and $\hat{\mathbf{x}}_t = \text{col}(\hat{\mathbf{x}}_{1,t}, \hat{\mathbf{x}}_{2,t}, \dots, \hat{\mathbf{x}}_{N,t})$ be the joint estimates of all players. $\hat{\mathbf{x}}_{i,T,t} = \text{col}\{\hat{\mathbf{x}}_{i,k} : k \in \mathcal{T}_{i,t}\}$ represents the joint delayed estimates received at

time t by player i , and $\hat{\mathbf{x}}_t = \text{col}(\hat{\mathbf{x}}_{1,t}, \hat{\mathbf{x}}_{2,t}, \dots, \hat{\mathbf{x}}_{N,t})$ represents the joint delayed estimates of all players, where $\mathcal{T}_t = \bigcup_{i=1}^N \mathcal{T}_{i,t}$ is the union of all players' time delays. Moreover, $\bar{F}_t(\hat{\mathbf{x}}_{\mathcal{T}_t}) = \text{col}\left(\left[\frac{1}{|\mathcal{T}_{1,t}|}\right]_+ \sum_{k \in \mathcal{T}_{1,t}} \nabla J_{1,k}(\mathbf{x}_{1,k}, \hat{\mathbf{x}}_{-1,k}), \left[\frac{1}{|\mathcal{T}_{2,t}|}\right]_+ \sum_{k \in \mathcal{T}_{2,t}} \nabla J_{2,k}(\mathbf{x}_{2,k}, \hat{\mathbf{x}}_{-2,k}), \dots, \left[\frac{1}{|\mathcal{T}_{N,t}|}\right]_+ \sum_{k \in \mathcal{T}_{N,t}} \nabla J_{N,k}(\mathbf{x}_{N,k}, \hat{\mathbf{x}}_{-N,k})\right)$ is the joint averaged delayed gradients of all players. Then, let $\alpha_t = \text{diag}\{\alpha_{1,t}, \alpha_{2,t}, \dots, \alpha_{N,t}\}$, $\Gamma = \text{diag}\{a_{11}, a_{12}, \dots, a_{1N}, a_{21}, a_{22}, \dots, a_{2N}, \dots, a_{N1}, a_{N2}, \dots, a_{NN}\}$, $\Gamma_i = \text{diag}\{a_{1i}, a_{2i}, \dots, a_{Ni}\}$, then the compact form of Algorithm 1 is

$$\begin{aligned} \mathbf{x}_{t+1} &= \mathcal{P}_{\mathcal{M}}(\mathbf{x}_t - \alpha_t \bar{F}_t(\hat{\mathbf{x}}_{\mathcal{T}_t})) \\ \hat{\mathbf{x}}_{t+1} &= ((\mathbf{I}_{N^2} - k_1(\mathcal{L} \otimes \mathbf{I}_N) - k_2 \Gamma) \otimes \mathbf{I}_q) \hat{\mathbf{x}}_t + k_2(\Gamma \otimes \mathbf{I}_q)(\mathbf{I}_N \otimes \mathbf{x}_t) \end{aligned} \quad (5)$$

where $\mathcal{L} = \mathcal{D}^+ - \mathbf{A}$ denotes the Laplacian matrix associated with $\mathcal{G}$, $\mathcal{D}^+ = \text{diag}\{d_1^+, d_2^+, \dots, d_N^+\}$ is the diagonal in-degree matrix, and $d_i^+ = \sum_{j=1}^N a_{ij}$ represents the number of incoming links to player i .

Concerning the estimation error, the following lemma is derived.

Lemma 1^[44]: Under Assumptions 1 and 2, if k_1 and k_2 satisfy

$$\begin{cases} 0 < k_1 < \frac{1}{\max_{i,j \in \mathcal{V}} \{a_{ij}\} + \max_{i \in \mathcal{V}} \{d_i^+\}} \\ 0 < k_2 < \frac{1}{\max_{i,j \in \mathcal{V}} \{a_{ij}\} + \max_{i \in \mathcal{V}} \{d_i^+\}} \end{cases} \quad (6)$$

then the estimation error $\mathbf{e}_t = \hat{\mathbf{x}}_t - \mathbf{1}_N \otimes \mathbf{x}_t$ generated by Algorithm 1 satisfies

$$\|\mathbf{e}_t\| \leq m_0^{t+1} \|\mathbf{e}_0\| + 2NC_M \sum_{s=0}^{t-1} m_0^s \alpha_{\max,t-s-1} \quad (7)$$

where $C_M > 0$, $m_0 \in (0, 1)$, and $\alpha_{\max,t} = \max_{i \in \mathcal{V}} \alpha_{i,t}$.

Using the above lemma, the bound of the dynamic regret is obtained as follows.

Theorem 1: Under Assumptions 1 and 2, if k_1 and k_2 satisfy Eq. (10), and $\alpha_{i,t}$ satisfies $0 < \alpha_{i,t+1} \leq \alpha_{i,t} < 1$, then the dynamic regret (4) generated by Algorithm 1 satisfies the following conditions:

$$\begin{aligned} R_{i,T} &\leq \\ L_J &\sqrt{T \left(\frac{2NC_M^2 + 2C_M \sqrt{N} \phi_T}{\mu \alpha_{\max,T}^2} + \frac{2C_M \sqrt{N} (\tau_{\max} - 1) \vartheta_T}{\mu} + \psi_{1,T} + \psi_{2,T} \right)} \end{aligned} \quad (8)$$

where $\phi_T = \sum_{t=1}^T \|\mathbf{x}_{t+1}^* - \mathbf{x}_t^*\|$, $\vartheta_T = \sum_{t=1}^T \|F_t(\mathbf{x}_t) - F_{t+1}(\mathbf{x}_t)\|$, and

$$\begin{aligned} \psi_{1,T} &= \left(2C_M L \sqrt{N} \sum_{t=1}^T \beta_{0,t} + L^2 \sum_{t=1}^T \alpha_{\max,t} \beta_{0,t}^2 \right) / \mu, \\ \psi_{2,T} &= \left(4L_J \sum_{t=1}^T \beta_{1,t} + 2LC_M L_J \sum_{t=1}^T \beta_{2,t} \right) / \mu, \\ \beta_{0,t} &= m_0^{t-\tau_{\max,t}+1} \|\mathbf{e}_0\| + 2NC_M \sum_{s=0}^{t-1} m_0^s \alpha_{\max,t-s-1}, \\ \beta_{1,t} &= \alpha_{\max,t} \sum_{i=1}^N \beta_{3,t}, \\ \beta_{2,t} &= (\tau_{\max} - 1) N \sum_{i=1}^N \alpha_{i,t}^2 \beta_{3,t}, \\ \beta_{3,t} &= [1/|\mathcal{T}_{i,t}|]_+ |\mathcal{T}_{i,t}|, \end{aligned}$$

where C_M , $\alpha_{\max,t}$ and m_0 are defined in Lemma 1, $L_J > 0$, $\bar{\tau}_{i,t} = \max_{k \in \mathcal{T}_{i,t}} \tau_{i,k}$, $\tau_{i,\max} = \max_{t \in \{0,1,\dots,T\}} \bar{\tau}_{i,t} > 0$, and $\tau_{\max} = \max_{i \in \mathcal{V}} \tau_{i,\max}$.

Proof: Based on Assumption 2.(1), it can be derived as

$$\|\mathbf{x}_{t+1} - \mathbf{x}_{t+1}^*\|^2 \leq 4C_M \sum_{i=1}^N \|\mathbf{x}_{i,t+1}^* - \mathbf{x}_{i,t}^*\| + \sum_{i=1}^N \|\mathbf{x}_{i,t+1} - \mathbf{x}_{i,t}^*\|^2 \quad (9)$$

and from the mathematical relationship in Line 10 of Algorithm 1, it can be further obtained

$$\sum_{i=1}^N \|\mathbf{x}_{i,t+1} - \mathbf{x}_{i,t}^*\|^2 \leq \sum_{i=1}^N (1 - \alpha_{i,t}) \|\mathbf{x}_{i,t} - \mathbf{x}_{i,t}^*\|^2 + \sum_{i=1}^N \alpha_{i,t} \|\mathbf{z}_{i,t+1} - \mathbf{x}_{i,t}^*\|^2 \quad (10)$$

With the utilization of the variation inequality (3) and the inherent projection characteristics, Eq. (10) can be rearranged and rewritten as

$$\begin{aligned} \sum_{i=1}^N \|\mathbf{x}_{i,t+1} - \mathbf{x}_{i,t}^*\|^2 &\leq \alpha_{\max,t}^3 \sum_{i=1}^N \left[\frac{1}{|\mathcal{T}_{i,t}|} \right]_+^2 \left\| \sum_{k \in \mathcal{T}_{i,t}} (\nabla J_{i,k}(\mathbf{x}_{i,t}^*, \mathbf{x}_{-i,t}^*) - \nabla J_{i,k}(\mathbf{x}_{i,k}, \hat{\mathbf{x}}_{-i,k})) \right\|^2 + \sum_{i=1}^N \|\mathbf{x}_{i,t} - \mathbf{x}_{i,t}^*\|^2 - \\ &2\alpha_{\max,t}^2 \sum_{i=1}^N \left\langle \left[\frac{1}{|\mathcal{T}_{i,t}|} \right]_+ \sum_{k \in \mathcal{T}_{i,t}} \nabla J_{i,k}(\mathbf{x}_{i,k}, \hat{\mathbf{x}}_{-i,k}) - \nabla J_{i,t}(\mathbf{x}_{i,t}^*, \mathbf{x}_{-i,t}^*), \mathbf{x}_{i,t} - \mathbf{x}_{i,t}^* \right\rangle \end{aligned} \quad (11)$$

In accordance with Assumption 2, a positive constant L_J is guaranteed to exist such that the corresponding inequality $\|\nabla J_{i,t}(\mathbf{x}, \mathbf{y})\| \leq L_J$ holds for any $(\mathbf{x}, \mathbf{y}) \in \mathcal{M}$ and any $i \in \mathcal{V}$. Combined with the Cauchy-Schwarz inequality, the following inequality can be established:

$$\sum_{i=1}^N \left[\frac{1}{|\mathcal{T}_{i,t}|} \right]_+^2 \left\| \sum_{k \in \mathcal{T}_{i,t}} (\nabla J_{i,t}(\mathbf{x}_{i,t}^*, \mathbf{x}_{-i,t}^*) - \nabla J_{i,k}(\mathbf{x}_{i,k}, \hat{\mathbf{x}}_{-i,k})) \right\|^2 \leq 2\Delta_{1,t} + 2\Delta_{2,t} \quad (12)$$

where the relevant parameter definition is provided as follows:

$$\begin{aligned} \Delta_{1,t} &= \sum_{i=1}^N \sum_{k \in \mathcal{T}_{i,t}} \left[\frac{1}{|\mathcal{T}_{i,t}|} \right]_+ \|\nabla J_{i,k}(\mathbf{x}_{i,k}, \mathbf{x}_{-i,k}) - \nabla J_{i,k}(\mathbf{x}_{i,k}, \hat{\mathbf{x}}_{-i,k})\|^2, \\ \Delta_{2,t} &= \sum_{i=1}^N \sum_{k \in \mathcal{T}_{i,t}} \left[\frac{1}{|\mathcal{T}_{i,t}|} \right]_+ \|\nabla J_{i,t}(\mathbf{x}_{i,t}^*, \mathbf{x}_{-i,t}^*) - \nabla J_{i,k}(\mathbf{x}_{i,k}, \mathbf{x}_{-i,k})\|^2. \end{aligned}$$

By virtue of the strong monotonicity property specified in Assumption 2, the following relation can be validated:

$$\begin{aligned} -2\alpha_{\max,t}^2 \sum_{i=1}^N \left\langle \left[\frac{1}{|\mathcal{T}_{i,t}|} \right]_+ \sum_{k \in \mathcal{T}_{i,t}} \nabla J_{i,k}(\mathbf{x}_{i,k}, \hat{\mathbf{x}}_{-i,k}) - \nabla J_{i,t}(\mathbf{x}_{i,t}^*, \mathbf{x}_{-i,t}^*), \mathbf{x}_{i,t} - \mathbf{x}_{i,t}^* \right\rangle &\leq \\ -2\alpha_{\max,t}^2 \mu \|\mathbf{x}_t - \mathbf{x}_t^*\|^2 + 4C_M \alpha_{\max,t}^2 (\Delta_{3,t} + \Delta_{4,t}) \end{aligned} \quad (13)$$

and

$$\begin{aligned} \Delta_{3,t} &= \sum_{i=1}^N \sum_{k \in \mathcal{T}_{i,t}} \left[\frac{1}{|\mathcal{T}_{i,t}|} \right]_+ \|\nabla J_{i,k}(\mathbf{x}_{i,t}, \mathbf{x}_{-i,t}) - \nabla J_{i,k}(\mathbf{x}_{i,k}, \hat{\mathbf{x}}_{-i,k})\|^2, \\ \Delta_{4,t} &= \sum_{i=1}^N \sum_{k \in \mathcal{T}_{i,t}} \left[\frac{1}{|\mathcal{T}_{i,t}|} \right]_+ \|\nabla J_{i,t}(\mathbf{x}_{i,t}, \mathbf{x}_{-i,t}) - \nabla J_{i,k}(\mathbf{x}_{i,t}, \mathbf{x}_{-i,t})\|^2. \end{aligned}$$

Meanwhile, the Cauchy-Schwarz inequality is adopted to derive the following inequality:

$$\Delta_{3,t} \leq L \sum_{i=1}^N \sum_{k \in \mathcal{T}_{i,t}} \left[\frac{1}{|\mathcal{T}_{i,t}|} \right]_+ \|\mathbf{e}_{i,k}\| + L \sum_{i=1}^N \left[\frac{1}{|\mathcal{T}_{i,t}|} \right]_+ \phi_{i,t} \quad (14)$$

where $\phi_{i,t} = |\mathcal{T}_{i,t}| \sum_{s=t+1-\bar{\tau}_{i,t}}^{t-1} \|\mathbf{x}_s - \mathbf{x}_{s+1}\|$ if $\bar{\tau}_{i,t} > 1$, otherwise $\phi_{i,t} = 0$.

A summation operation is implemented on Eq. (13) with yields $\sum_{t=1}^T \sum_{i=1}^N \|\mathbf{x}_{i,t+1}^* - \mathbf{x}_{i,t}^*\| \leq \sqrt{N}\phi_T$. By substituting Eqs. (10)-(14) into Eq. (9) and performing summation on both sides of Eq. (9) with respect to t , the following inequality can be derived:

$$\begin{aligned} \mu \sum_{t=1}^T \|\mathbf{x}_t - \mathbf{x}_t^*\|^2 &\leq \sum_{t=1}^T \frac{\|\mathbf{x}_t - \mathbf{x}_t^*\|^2 - \|\mathbf{x}_{t+1} - \mathbf{x}_{t+1}^*\|^2}{2\alpha_{\max,t}^2} + \\ &2C_M \sum_{t=1}^T \frac{1}{\alpha_{\max,t}^2} \sum_{i=1}^N \|\mathbf{x}_{i,t+1}^* - \mathbf{x}_{i,t}^*\| + L^2 \sum_{t=1}^T \alpha_{\max,t} |\mathcal{T}_t| \beta_{0,t}^2 + \\ &2C_M L \sqrt{N} \sum_{t=1}^T |\mathcal{T}_t| \beta_{0,t} + 4L^2 \sum_{t=1}^T \beta_{1,t} + 2C_M \sum_{t=1}^N (\tau_{i,\max} - 1) \times \\ &\sum_{t=1}^T (\|\nabla J_{i,t}(\mathbf{x}_{i,t}, \mathbf{x}_{-i,t}) - \nabla J_{i,t+1}(\mathbf{x}_{i,t}, \mathbf{x}_{-i,t})\| + L \|\mathbf{x}_t - \mathbf{x}_{t+1}\|) \end{aligned} \quad (15)$$

and then, the following result can be acquired:

$$\begin{aligned} \sum_{t=1}^T \frac{\|\mathbf{x}_t - \mathbf{x}_t^*\|^2 - \|\mathbf{x}_{t+1} - \mathbf{x}_{t+1}^*\|^2}{2\alpha_{\max,t}^2} &\leq \frac{2NC_M^2}{\alpha_{\max,1}^2} + \\ \frac{1}{2} \left(\frac{1}{\alpha_{\max,T}^2} - \frac{1}{\alpha_{\max,1}^2} \right) 4C_M^2 &= \frac{2NC_M^2}{\alpha_{\max,T}^2} \end{aligned} \quad (16)$$

By Jensen's inequality and the non-increasing property of $\alpha_{i,t}$, it is derived as

$$\sum_{t=1}^T \frac{1}{\alpha_{\max,t}^2} \sum_{i=1}^N \|\mathbf{x}_{i,t+1}^* - \mathbf{x}_{i,t}^*\| \leq \frac{\sqrt{N}}{\alpha_{\max,T}^2} \phi_T \quad (17)$$

In combination with the mathematical relations in Eqs. (5)-(7) as well as the projection property and Assumption 2, one gets that

$$\|\mathbf{x}_{t+1} - \mathbf{x}_t\| \leq L_j \sum_{i=1}^N [1/|\mathcal{T}_{i,t}|]_+ |\mathcal{T}_{i,t}| \alpha_{i,t}^2 \quad (18)$$

Finally, combining Eqs. (15)-(18) and Lemma 1 with Assumption 2 gets Eq. (8). The proof is therefore completed. $\square$

Remark 2: The dynamic regret bound in (12) depends on four variables, i.e., ϕ_T , ϑ_T , $\psi_{1,T}$, and $\psi_{2,T}$. Specifically, ϕ_T refers to the NE-sequence path variation, a commonly used measure in online optimization and game problems^[13-16,25-27]. ϑ_T reflects the path variation of the pseudo-gradient sequence. $\psi_{1,T}$ captures the cumulative consensus estimation error, while $\psi_{2,T}$ represents the cumulative error incurred by time delays. When the considered problem reduces to the classical online game with $\tau_{i,k} = 1$, the bound in Eq. (8) remains applicable and simplifies to

$\mathcal{O}\left(\sqrt{T\left((1+\phi_T)/\alpha_{\max,T}^2 + \sum_{t=1}^T \alpha_{\max,t}^2\right)}\right)$ given $\tau_{\max} = 1$ and $\psi_{2,T} = 0$, in full agreement with the conclusion of Ref. [13].

Based on Theorem 1, the sufficient conditions for the sublinear dynamic regret are derived as follows.

Theorem 2: Under Assumptions 1 and 2, if $\phi_T = \mathcal{O}(T^\vartheta)$ and $\vartheta_T = \mathcal{O}(T^\theta)$ with $0 \leq \theta < 1$, k_1 and k_2 satisfy Eq. (10), $\alpha_{i,t}$ satisfies $\alpha_{i,t} = \alpha_t = \mathcal{O}(t^{-\delta})$ with $0 < \delta < (1-\theta)/2$, then the dynamic regret (4) generated by Algorithm 1 is sublinear in T . Moreover, the optimal choice $\delta = \delta^* = (1-\theta)/3$ leads $R_{i,T} = \mathcal{O}(T^{(5+\theta)/6})$.

Proof: One obtains that $0 < \alpha_{i,t} \leq \alpha_{i,t+1} < 1$ with $\alpha_{i,t} = \alpha_t = \mathcal{O}(t^{-\delta})$, and thus the conditions of Theorem 1 are all satisfied. Then, there is

$$\begin{aligned} \beta_{0,t} &= \Theta(\alpha_t) = \mathcal{O}(t^{-\delta}), \\ \beta_{1,t} &= \Theta(\alpha_t) = \mathcal{O}(t^{-\delta}), \\ \beta_{2,t} &= \Theta(\alpha_t^2) = \mathcal{O}(t^{-2\delta}) \end{aligned} \quad (19)$$

where the derivation of $\beta_{0,t}$ follows $0 < \lambda_0 < 1$. Thus, one gets that

$$\begin{aligned} \psi_{1,T} &= \mathcal{O}(T^{\max\{1-\delta, 1-3\delta\}}) = \mathcal{O}(T^{4-\delta}) \\ \psi_{2,T} &= \mathcal{O}(T^{\max\{1-\delta, 1-2\delta\}}) = \mathcal{O}(T^{4-\delta}) \end{aligned} \quad (20)$$

Substituting Eq. (20) into Eq. (8) yields that

$$\begin{aligned} R_{i,T} &= \mathcal{O}(T^{0.5+\delta} \sqrt{\phi_T} + T^{0.5} \sqrt{\vartheta_T} + T^{4-0.5\delta}) \\ &= \mathcal{O}(T^{\max\{0.5+\delta+0.5\theta, 1-0.5\delta\}}) \end{aligned} \quad (21)$$

Combining with $0.5 + \delta + 0.5\theta < 1$ and $1 - 0.5\delta < 1$, one concludes that $R_{i,T} = o(T)$. One knows that the order of $R_{i,T}$ in Eq. (21) is the smallest when $0.5 + \delta + 0.5\theta$ equals $1 - 0.5\delta$, and thus $\delta^* = (1-\theta)/3$. Substituting $\delta = \delta^*$ into Eq. (21) yields $R_{i,T} = \mathcal{O}(T^{(5+\theta)/6})$. The proof is therefore completed. $\square$

Remark 3: As implied by Theorem 2, no-regret property can be ensured by Algorithm 1 when path variations exhibit sublinear growth with respect to T , and the learning rate is determined based on path variations in the absence of time delay information.

3 Simulation results

Consider the online electricity market game with $N=5$ wind power companies. The communication topology is directed and strongly connected, as shown in Fig. 1.

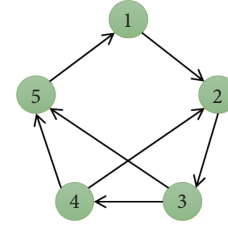


Figure 1 The directed and strongly connected topology of 5 wind power companies.

The time horizon is set to $T=1000$ time steps. Each wind power enterprise $i \in \mathcal{V}$ acts as an independent participant and determines its real-time power generation output $x_{i,t} \in \mathcal{M}_i \subseteq \mathbb{R}$, so the decision dimension is $q=1$. Referring to the game modeling method adopted in Ref. [28], the mathematical expression of the aforementioned online market game is established as Eq. (1). Meanwhile, the local cost function for each wind power producer is formulated as

$$J_{i,t}(\mathbf{x}_{i,t}, \mathbf{x}_{-i,t}) = J_{i1,t}(\mathbf{x}_{i,t}) + J_{i2,t}(\mathbf{x}_{i,t}, \mathbf{x}_{-i,t}) \quad (22)$$

where the first term $J_{i1,t}(\mathbf{x}_{i,t})$ stands for the time-dependent operational and generation cost, which dynamically fluctuates with the change of operating time slots. The second term $J_{i2,t}(\mathbf{x}_{i,t}, \mathbf{x}_{-i,t})$ indicates the real-time market electricity price, which serves as the core market transaction parameter affecting the profit of power generation enterprises. Particularly, the function $J_{i1,t}$ is

$$J_{i1,t}(\mathbf{x}_{i,t}) = a_{i,t} + b_{i,t} \mathbf{x}_{i,t} + c_{i,t} \mathbf{x}_{i,t}^2 \quad (23)$$

and the function $J_{i2,t}$ is

$$J_{i2,t}(\mathbf{x}_{i,t}, \mathbf{x}_{-i,t}) = g_t - \mathbf{x}_{i,t} - 0.25 \sum_{s=1, s \neq i}^N \mathbf{x}_{s,t} \quad (24)$$

where $g_t = 80 + 0.5 \sin(t/12)$. The detailed settings of $a_{i,t}$, $b_{i,t}$, $c_{i,t}$ and $\mathcal{M}_i$ follow Table 1.

To evaluate the robustness of the proposed algorithm against

Table 1 The detailed settings of parameters

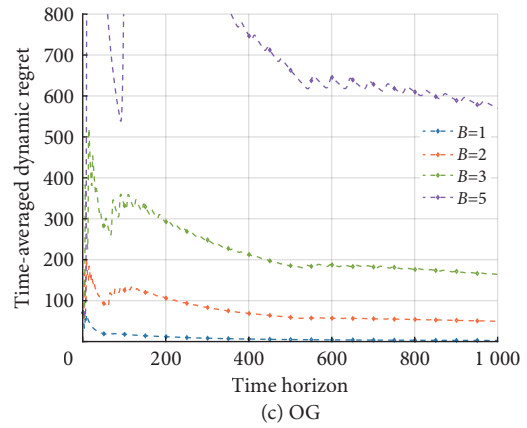
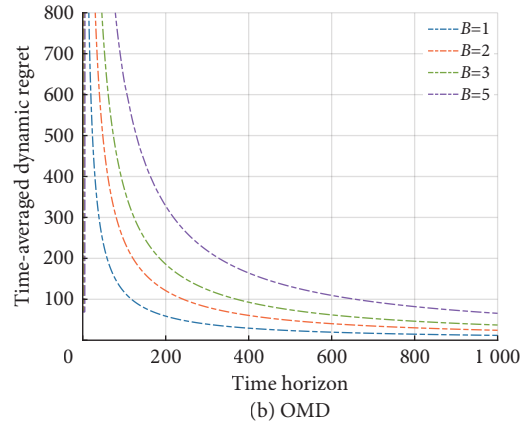
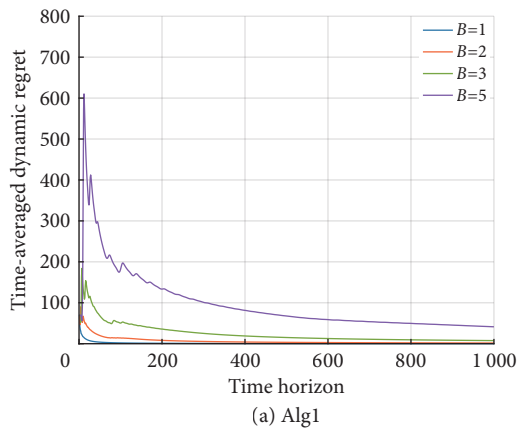
company index	$a_{i,t}$	$b_{i,t}$	$c_{i,t}$	$\mathcal{M}_i$
1	2	$1.5 \sin(t/2)$	6	[2,45]
2	9	$2 \sin(t/6)$	4	[5,50]
3	6	$2.5 \sin(t/5)$	5	[4,35]
4	7	$3 \sin(t/9)$	6	[3,30]
5	5	$3.5 \sin(t/12)$	3	[8,60]

delay variation, the time-varying delay is set as

$$\begin{cases} \tau_{i,t} = B, t \leq 50 \\ \tau_{i,t} = 2B, 51 \leq t < 500 \\ \tau_{i,t} = 3B, \tau_{i,t} = 3B \end{cases} \quad (25)$$

and is capped by $T+1-t$. Four delay levels, i.e., $B \in \{1, 2, 3, 5\}$, are tested in separate trials. The algorithm parameters are set as $k_1 = 0.1$, $k_2 = 0.2$, $\alpha_{i,t} = (t+1)^{-0.3667}$.

Algorithm 1 (referred to as “Alg1”) is compared with two classic online algorithms. The comparison benchmarks include the online mirror descent algorithm with delays proposed in Ref. [19] (referred to as “OMD”) and the online gradient algorithm proposed in Ref. [14] (referred to as “OG”). In the simulation setup of OMD, all time-varying feedback delays (i.e., $\tau_{i,t}$) are assumed to be completely observable and known in advance. For both benchmark algorithms OMD and OG, single delayed feedback is randomly selected at each independent time step. It is worth noting that Alg1 does not require prior knowledge of the delays, while OMD is provided with oracle delay information, which makes the comparison conservative and favorable to OMD. Fig. 2 presents the time-averaged dynamic regrets of all compared algorithms, i.e., $R_{i,T}/T$, under diverse delay scenarios. The numerical results and curve comparisons show that the proposed Alg1 achieves the best overall performance. In the early stage, the time-averaged regrets of all algorithms are relatively high because of the transient adaption process. As T increases, the regret curves of Alg1 decrease rapidly and gradually flatten, which is consistent with the sublinear dynamic regret established in Theorem 2. When B increases from 1 to 5, the regret curves of OMD and OG shift upward more noticeably, indicating that their performance is more sensitive to delay variation. In contrast, Alg1 maintains relatively stable and low time-averaged dynamic regret across all tested delay levels. These results fully validate that Alg1 possesses outstanding adaptive performance and stronger robustness in complex variable-delay environments than the two conventional benchmark algorithms.

**Figure 2** The time-averaged dynamic regrets of all compared algorithms.

4 Conclusions

This paper proposed an online distributed algorithm for NE seeking to handle unknown time-varying delays among players. By integrating leader-follower consensus with projected gradient descent over unbalanced topologies, an upper bound on dynamic regret was established. An appropriate learning rate, tailored to the path variation of NE sequence and that of pseudo-gradient sequence, was derived to ensure sublinear dynamic regret.

Funding

This work was supported by the National Natural Science Foundation of China (U2241217, 62473027, 62473029, 62403038, 62388101), and the Beijing Natural Science Foundation (JQ23019).

Data availability

All data needed to support the conclusions in the paper are presented in the manuscript. Additional data related to this paper may be requested from the corresponding author upon request.

Declaration of competing interest

The author reports no conflict of interest in this work.

Author contribution statement

Lingfei Su: Conceptualization, Methodology, Writing (original draft and editing). Yongzhao Hua: Methodology, Writing (review

and editing), Supervision. Xiwang Dong: Supervision, Writing (review and editing), Funding acquisition. Jinhu Lü: Writing (review and editing), Academic advice. Danwei Wang: Writing (review and editing), Research direction guidance.

Use of AI statement

None.

References

- [1] Dong X W, Yu B C, Shi Z Y, et al. Time-varying formation control for unmanned aerial vehicles: theories and applications[J]. *IEEE Transactions on Control Systems Technology*, 2015, 23(1): 340-348.
- [2] Zhang G Y, He X D, Mu Y, et al. Serial distributed detection in multi-hop multirelay wireless sensor networks with end-edge-cloud orchestration under graph-powered computing[J]. *IEEE Internet of Things Journal*, 2025, 12(4): 3720-3733.
- [3] Yuan L W, Ishii H. Resilient distributed economic dispatch in smart grids[J]. *IEEE Transactions on Automatic Control*, 2026, 71(4): 2289-2304.
- [4] Başar T, Olsder G J. Dynamic noncooperative game theory[M]. Philadelphia, PA: SIAM, 1999: 23.
- [5] Geng M J, Ding H F, Yao X Y, et al. Noncooperative game of distributed quadrotor UAVs with multiple constraints[J]. *IEEE Transactions on Aerospace and Electronic Systems*, 2024, 60(4): 4728-4739.
- [6] Huang Y, Meng Z Y, Sun J. Distributed Nash equilibrium seeking for multicenter aggregative game of Euler-Lagrange systems with coupled constraints[J]. *IEEE Transactions on Cybernetics*, 2024, 54(10): 5672-5683.
- [7] Deng Z H, Luo J. Distributed strategy design for multicoalition games with autonomous high-order players and its application in smart grids[J]. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 2024, 54(8): 4726-4735.
- [8] Liu S, Wang D, Wu Z G. Distributed multi-coalition games with general linear systems over Markovian switching networks[J]. *IEEE Transactions on Network Science and Engineering*, 2024, 11(6): 6157-6168.
- [9] Chen Y Y, Li X D, Feng Z, et al. Prescribed-time Nash equilibrium seeking for multicoalition games with heterogeneous general linear dynamics over unbalanced digraphs[J]. *IEEE Transactions on Cybernetics*, 2025, 55(12): 5803-5815.
- [10] Rui J Y, Ding L, Ye M J, et al. Prescribed-time distributed integral sliding-mode-based least-norm Nash equilibrium seeking in monotone games under disturbances[J]. *IEEE Transactions on Cybernetics*, 2026, 56(3): 1638-1648.
- [11] Liu P, Xiao F, Wei B, et al. Generalized Nash equilibrium seeking for noncooperative games with heterogeneous individual dynamics[J]. *IEEE Transactions on Automatic Control*, 2024, 69(4): 2492-2499.
- [12] You J, Wu W C. Online passive identifier for spatially distributed systems using mobile sensor networks[J]. *IEEE Transactions on Control Systems Technology*, 2017, 25(6): 2151-2159.
- [13] Lu K H, Li G Q, Wang L. Online distributed algorithms for seeking generalized Nash equilibria in dynamic environments[J]. *IEEE Transactions on Automatic Control*, 2021, 66(5): 2289-2296.
- [14] Deng Z H, Zuo X L. Distributed online learning algorithm for non-cooperative games over unbalanced digraphs[J]. *IEEE Transactions on Neural Networks and Learning Systems*, 2024, 35(11): 15846-15856.
- [15] Lin Y M, Liu K, Han D Y, et al. Statistical privacy-preserving online distributed Nash equilibrium tracking in aggregative games[J]. *IEEE Transactions on Automatic Control*, 2024, 69(1): 323-330.
- [16] Yu R, Meng M, Li L. Distributed online learning for leaderless multicluster games in dynamic environments[J]. *IEEE Transactions on Control of Network Systems*, 2024, 11(3): 1548-1561.
- [17] Xu H, Lu K H, Wang T B, et al. Online distributed seeking for first-order Nash equilibria of nonconvex noncooperative games with multiple clusters[J]. *IEEE Transactions on Circuits and Systems II: Express Briefs*, 2023, 70(2): 621-625.
- [18] Liu P, Wang L T, Chen Y. Sublinear dynamic regrets for aggregative games with feedback delays and communication delays[J]. *IEEE Transactions on Control of Network Systems*, 2025, 12(1): 1071-1079.
- [19] Meng M, Li X X, Chen J. Decentralized Nash equilibria learning for online game with bandit feedback[J]. *IEEE Transactions on Automatic Control*, 2024, 69(6): 4050-4057.
- [20] Wu P, Huang H Y, Lu H L, et al. Distributed online convex optimization with unknown delays[J/OL]. *IEEE Transactions on Automatic Control*. DOI: 10.1109/TAC.2026.3676242. <https://ieeexplore.ieee.org/document/11449414>.
- [21] Facchinei F, Pang J S. Nash equilibria: the variational approach[M]//Palomar D P, Eldar Y C. Convex Optimization in Signal Processing and Communications. New York: Cambridge University Press, 2009: 443-493.
- [22] Yao Y T, Lu K H, Zhu Q X. Online distributed algorithms for aggregative games with stochastic costs over time-varying networks[J]. *International Journal of Control*, 2026, 99(1): 90-97.
- [23] Sun C, Chen B, Wang J Z, et al. Distributed stochastic Nash equilibrium seeking under heavy-tailed noises[J]. *Automatica*, 2025, 173: 112081.
- [24] Feng Z, Shu P X, Dong X W, et al. Adaptively distributed Nash equilibrium seeking of noncooperative games for uncertain nonlinear heterogeneous multiagent systems[J]. *IEEE Transactions on Automatic Control*, 2026, 71(5): 3263-3278.
- [25] Okamoto K, Hayashi N, Takai S. Distributed online adaptive gradient descent with event-triggered communication[J]. *IEEE Transactions on Control of Network Systems*, 2024, 11(2): 610-622.
- [26] Yao S Q, Xie S Y, Li T. Online distributed optimization algorithm with dynamic regret analysis under unbalanced graphs[J]. *Automatica*, 2025, 174: 112116.
- [27] Su L F, Li X D, Hua Y Z, et al. Optimal tracking of time-varying formation in open nonlinear multiagent systems[J/OL]. *IEEE Transactions on Automatic Control*. DOI: 10.1109/TAC.2026.3705990. <https://ieeexplore.ieee.org/document/11573660>.
- [28] Hobbs B F, Pang J S. Nash-Cournot equilibria in electric power markets with piecewise linear demand functions and joint constraints[J]. *Operations Research*, 2007, 55(1): 113-127.