Received: 10 June 2024 | Revised: 30 September 2024 | Accepted: 21 October 2024
DOI: 10.1049/cth2.12759
IET Control Theory & Applications
The Institution of Engineering and Technology WILEY

ORIGINAL RESEARCH

Distributed zeroth-order online optimization with communication delays

Keito Inoue[1] | Naoki Hayashi[2] | Shigemasa Takai[1]

[1]Graduate School of Engineering, Osaka University, Osaka, Japan

[2]Graduate School of Engineering Science, Osaka University, Osaka, Japan

Correspondence
Naoki Hayashi, Graduate School of Engineering Science, Osaka University, 1-3 Machikaneyama-cho, Toyonaka, Osaka 560-8531, Japan.
Email: n.hayashi@sys.es.osaka-u.ac.jp

Funding information
Japan Society for the Promotion of Science, Grant/Award Number: JP21K04121

Abstract
This paper investigates distributed online optimization in a networked multiagent system, where each agent has its own private objective and constraint functions that vary over time. In many real-world scenarios, computing the gradient of the cost function can be challenging, especially when agents have limited computational capabilities. Moreover, communication delays are common in practical networked systems due to various factors. This paper considers a unified framework for distributed online optimization that can handle bandit feedback and communication delays feedback simultaneously. A distributed primal-dual algorithm is proposed that utilizes bandit feedback, in which the agents estimate the gradients of their objective and constraint functions by sampling the function values. An enlarged network model that incorporates the delayed information exchanged among the agents is introduced. Through theoretical analysis, it is shown that the proposed algorithm achieves sublinear upper bounds on both the dynamic regret and the constraint violation despite communication delays.

1 | INTRODUCTION

Distributed optimization has gained significant attention in recent years as a framework for controlling and optimizing large-scale networked systems [1]. The agents' goal is to collectively find a solution that optimizes the overall system performance rather than just optimizing their individual objectives. One of the most widely studied approaches is consensus-based optimization [2]. In this approach, agents iteratively share information with their neighbours in the communication network to reach a consensus on the decision variables that optimize the global objective function.

While much research has been done on developing distributed algorithms for off-line optimization, where a local cost function does not change over time, practical situations often involve dynamic environments with a time-varying local cost function. Online optimization is a framework that addresses such dynamic settings, aiming to minimize a performance metric called a regret [3]. Many research articles have explored various directions to tackle the challenges of distributed online optimization. One line of research focuses on developing distributed methods that can handle directed communication networks [4, 5]. The influence of noisy gradient information has also been studied [6, 7], where a high-probability bound on the regret was analysed. Several research articles have investigated distributed approaches that incorporate shared constraints such as primal-dual methods [8] and augmented Lagrangian frameworks [9, 10]. Furthermore, in real-world networks, communication bandwidth can impact the performance of distributed algorithms. Researchers have considered distributed approaches that consider communication constraints such as quantization of exchanged information [11] and event-triggered communication schemes [12, 13].

In distributed online optimization, many algorithms assume that the gradient of the cost function can be easily computed. However, this assumption may not always be practical, particularly when agents have limited computational resources. To address scenarios with limited gradient information, one approach is the use of bandit feedback algorithms, also known as zeroth-order algorithms. Bandit feedback is crucial in distributed optimization for several reasons. First, it enables the optimization of black-box functions. This is particularly important in real-world applications, where the exact form of the objective function may not be known or may be too complex to model explicitly. Second, bandit feedback optimization can be more computationally efficient than methods requiring exact gradient computations, as it relies only on the values of the cost and constraint functions at each iteration. This makes it well-suited for resource-constrained environments, such as wireless sensor networks or mobile devices. One class of these algorithms approximates gradients using a single function evaluation [14], while multi-point bandit feedback algorithms estimate gradients through multiple function evaluations [15, 16]. The latter approach allows for more accurate gradient estimates compared to single-point methods.

Communication delays are another significant challenge in distributed optimization. In real-world networked systems, the exchange of information is often subject to delays arising from various factors. Nedić and Ozdaglar analysed the convergence properties of these algorithms in the presence of delays [17]. The authors have shown that delays can slow down the convergence rate and require additional assumptions and modifications to the algorithms to ensure convergence. Recent studies have made significant progress in consensus control with delayed feedback. Lin and Ren proposed a consensus algorithm in multiagent systems with communication delays [18]. Zhang et al. studied continuous-time multiagent systems with both communication noise and delays [19]. Wei et al. considered a model predictive control approach under the conditions of input constraints and bounded time-varying communication delays [20]. The distributed optimization problem with communication delays have also been studied in [21–24]. However, most existing methods assume that agents have access to full gradient information, which may not always hold in practical applications.

To address these challenges, we propose a distributed online optimization algorithm incorporating bandit feedback and communication delays. The proposed algorithm is based on a distributed primal-dual approach and utilizes two-point bandit feedback to estimate gradients without perfect information. Through theoretical analysis, we prove that the proposed algorithm achieves sublinear dynamic regret and constraint violation despite uncertainties from bandit feedback and communication delays. Previous research on distributed bandit feedback algorithms has studied scenarios where gradient information is unavailable [4, 25, 26]. However, the impact of communication delays is not considered in these research articles. While the distributed online optimization methods with the delayed gradient information have been investigated [27], these approaches do not consider delays in communication between agents. Several research articles have addressed the effects of communication delays on the convergence of online distributed algorithms [28, 29]. However, these methods assume perfect gradient information and cannot be applied in the bandit settings.

Compared to the existing methods, the proposed method offers several advantages.

- The proposed method can address both the bandit feedback and the communication delay feedback in a distributed multiagent system. The bandit feedback is particularly advantageous in scenarios where the cost and constraint functions are complex or their gradients are difficult to compute. Communication delays are also a critical issue in many real-world networks, often arising from factors such as network congestion, limited bandwidth, and physical distance between agents. While existing methods usually tackle these issues separately, the proposed algorithm integrates them into a unified framework.
- The proposed algorithm incorporates a distributed primal-dual method with two-point bandit feedback. This approach enables each agent to appropriately estimate gradients without relying on exact gradient information. To address communication delays, we consider an enlarged graph approach, where virtual agents are introduced to the original graph to handle the delayed information. These virtual agents store and forward the delayed information to the appropriate agents. By incorporating these virtual agents, the proposed method can deal with the situation where each agent receives the information with communication delays.

The paper is organized as follows. Section 2 introduces the problem formulation of the online optimization and the bandit feedback. Section 3 proposes a distributed online primal-dual algorithm in the presence of communication delays. Section 4 shows the analysis of the regret bound of the proposed algorithm. Section 5 illustrates the numerical examples of the proposed algorithm. Finally, Section 6 gives concluding remarks.

2 | PROBLEM FORMULATION

2.1 | Notations

The sets of n-dimensional real vectors and real vectors with non-negative components are represented by $\mathbb{R}^n$ and $\mathbb{R}^n_+$, respectively. The set of non-negative integers is represented by $\mathbb{N}$. The symbols $\mathbf{0}$ and $\mathbf{1}$ stand for vectors of appropriate dimensions with all elements equal to 0 and 1, respectively. We use I_n to represent the $n \times n$ identity matrix. Vector inequalities are interpreted componentwise. The Euclidean norm and the ℓ_1-norm of a vector are represented by $\|\cdot\|$ and $\|\cdot\|_1$, respectively. For a matrix A, $[A]_{ij}$ represents its (i,j)th element. Similarly, for a vector x, $[x]_i$ represents its ith element. For functions $h_1 : \mathbb{R} \to \mathbb{R}$ and $h_2 : \mathbb{R} \to \mathbb{R}_+$, $h_1 = O_+(h_2)$ if $|h_1(x)| \leq C_1 h_2(x)$ for some positive constant C_1 and sufficiently large $x \in \mathbb{R}$. Similarly, $h_1 = O(h_2)$ means that $h_1(x) \leq C_2 h_2(x)$ for some positive constant C_2 and sufficiently large $x \in \mathbb{R}$. The n-dimensional closed unit ball and the unit sphere are represented by $\mathbb{B}^n = \{x \in \mathbb{R}^n \mid \|x\| \leq 1\}$ and $\mathbb{S}^n = \{x \in \mathbb{R}^n \mid \|x\| = 1\}$, respectively. For a closed convex set $\mathcal{S} \subset \mathbb{R}^n$, the projection of a vector $x \in \mathbb{R}^n$ onto $\mathcal{S}$ is defined as $\Pi_{\mathcal{S}}(x) = \arg\min_{y \in \mathcal{S}} \|x - y\|$. As a property of the projection operator, we have

$$\|\Pi_{\mathcal{S}}(x) - \Pi_{\mathcal{S}}(y)\| \leq \|x - y\|, \quad \forall x, y \in \mathbb{R}^n. \tag{1}$$

We use $[z]_+$ to represent the componentwise projection of a vector $z \in \mathbb{R}^n$ onto the non-negative orthant $\mathbb{R}^n_+$. Finally, $\mathbb{E}_u[\cdot]$ represents the expected value with respect to the random variable u. However, we omit the random variable u if it is clear from the context.

2.2 | Multiagent system

We consider a multiagent system with N agents whose network is represented by a time-varying directed graph $\mathcal{G}_k = (\mathcal{I}, \mathcal{E}_k)$, where $\mathcal{I} = \{1, 2, \dots, N\}$ is the set of the agents, $\mathcal{E}_k$ is the set of edges at iteration $k \in \mathcal{K}$, and $\mathcal{K} = \{1, 2, \dots, T\}$. An edge $(j, i) \in \mathcal{E}_k$ represents that agent j transmits information to agent i at iteration k. If any agent can be reached by a directed path from any other agent in the graph, the graph is said to be strongly connected. A sequence of communication graphs $\mathcal{G}_1, \mathcal{G}_2, \dots$ is uniformly strongly connected if there exists a positive integer B such that $\mathcal{E}_B(s) = \bigcup_{\ell=(s-1)B+1}^{sB} \mathcal{E}_\ell$ is strongly connected for every $s \in \mathbb{N} \setminus \{0\}$. The weight matrix $Q^{(k)} = [q_{ij}^{(k)}] \in \mathbb{R}^{N \times N}$ consists of the edge weights $q_{ij}^{(k)}$ that has a positive value if $(j, i) \in \mathcal{E}_k$, and $q_{ij}^{(k)} = 0$ for $(j, i) \notin \mathcal{E}_k$. In this paper, the following assumption about the connectivity of the communication graph is made:

Assumption 1. The sequence of the communication graphs $\{\mathcal{G}_k\}_{k \in \mathcal{K}}$ is uniformly strongly connected. Moreover, the weight matrix $Q^{(k)}$ corresponding to $\mathcal{G}_k$ is column stochastic for all $k \in \mathcal{K}$ such that $q_{ij}^{(k)} \geq p$ for all $(j, i) \in \mathcal{E}_k$ and $k \in \mathcal{K}$, and $q_{ii}^{(k)} > p$ for all $i \in \mathcal{I}$ and $k \in \mathcal{K}$, where p is a positive constant.

This assumption ensures that information can propagate through the network over time, which is crucial for convergence in distributed optimization [2].

Each agent i in the system has an associated local cost function $f_{i,k} : \mathcal{X}_i \to \mathbb{R}$ and a local constraint function $g_{i,k} : \mathcal{X}_i \to \mathbb{R}^m$ at iteration k, where $\mathcal{X}_i \subset \mathbb{R}^{n_i}$ is the feasible region of the decision variable for agent i.

Assumption 2. For any $i \in \mathcal{I}$ and $k \in \mathcal{K}$, $f_{i,k}$ and $g_{i,k}$ are convex. Moreover, the local constraint set $\mathcal{X}_i$ is bounded, closed, and convex.

Assumption 2 on the convexity of the cost and constraint functions, and the boundedness of the constraint sets is a fundamental assumption in convex optimization literature [8]. From Assumption 2, several important properties can be derived. First, due to the boundedness of the local constraint set, $\|x_i\| \leq B_x$ holds for all $x_i \in \mathcal{X}_i$ and $i \in \mathcal{I}$, where B_x is a positive constant. Moreover, we have $\|g_{i,k}(x_i)\| \leq B_g$ and $\|D_{g_{i,k}}(x_i)\| \leq C_g$ for all $x_i \in \mathcal{X}_i$, $i \in \mathcal{I}$, and $k \in \mathcal{K}$, where B_g and C_g are positive constants, and $D_{g_{i,k}}(x_i) \in \mathbb{R}^{n_i \times n_i}$ is the Jacobian of $g_{i,k}$ at $x_i \in \mathcal{X}_i$.

2.3 | Constrained online optimization

We consider the global cost function $F_k(x) = \sum_{i \in \mathcal{I}} f_{i,k}(x_i)$ and the global constraint function $G_k(x) = \sum_{i \in \mathcal{I}} g_{i,k}(x_i)$, where $x_i \in \mathcal{X}_i$ is the decision variable of agent i, $x = [x_1^\top, x_2^\top, \dots, x_N^\top]^\top \in \mathcal{X}$, $\mathcal{X} \subset \mathbb{R}^n$ is the concatenation of the local constraint sets of N agents, and $n = \sum_{i \in \mathcal{I}} n_i$.

The multiagent system collaboratively solves the following online optimization problem with inequality constraints at each iteration:

$$\underset{x \in \mathcal{X}}{\text{minimize}} \quad \sum_{k \in \mathcal{K}} F_k(x) \tag{2a}$$

$$\text{subject to} \quad \sum_{k \in \mathcal{K}} G_k(x) \leq \mathbf{0}. \tag{2b}$$

The optimization problem (2) is a general formulation that covers a wide range of constrained distributed online optimization problems. For example, the optimization problem (2) can be applied to resource allocation problems as discussed in [8], where the objective is to optimally allocate limited resources among multiple agents in a dynamic environment. Other examples include online linear regression [30] and online model predictive control problems [31].

The Lagrangian function for the optimization problem (2) is defined as $\sum_{k \in \mathcal{K}} H_k(x, \rho)$, where $\rho \in \mathbb{R}^m_+$ is a dual variable,

$$H_k(x, \rho) = \sum_{i \in \mathcal{I}} H_{i,k}(x_i, \rho) = F_k(x) + \rho^\top G_k(x), \tag{3}$$

and $H_{i,k}(x_i, \rho) = f_{i,k}(x_i) + \rho^\top g_{i,k}(x_i)$. Then, the dual problem of Equation (2) is given by

$$\max_{\rho \in \mathbb{R}^m_+} \sum_{k \in \mathcal{K}} r_k(\rho), \tag{4}$$

where $r_k(\rho) = \min_{x \in \mathcal{Y}_k} H_k(x, \rho)$ and $\mathcal{Y}_k = \{x \in \mathcal{X} \mid G_k(x) \leq \mathbf{0}\}$.

We consider the following Slater's condition:

Assumption 3. For any $k \in \mathcal{K}$, there exists $\check{x}$ in the relative interior of $\mathcal{X}$ such that $G_k(\check{x}) < \mathbf{0}$.

Assumption 3 is a classical assumption in constrained optimization that guarantees strong duality [32].

2.4 | Bandit feedback

In the bandit setting, the true gradient values of the cost and constraint functions are unavailable. Instead, each agent i can only observe the cost function value and constraint function value. Specifically, each agent samples the cost and constraint functions at two points close to its current decision to approximate the local gradient. To this end, we utilize a smoothing approximation and two-point bandit feedback.

We here consider a function $\varphi : \mathbb{K} \to \mathbb{R}^m$, where $\mathbb{K} \subset \mathbb{R}^n$. We consider the following assumption:

Assumption 4.

1. There exist positive constants $\zeta_{\min}$ and $\zeta_{\max}$ such that $\zeta_{\min}\mathbb{B}^n \subset \mathbb{K} \subset \zeta_{\max}\mathbb{B}^n$.
2. φ is convex and Lipschitz continuous on $\mathbb{K}$.
3. For any $x \in \mathbb{K}$, there exists $B_\varphi > 0$ such that $\|\varphi(x)\| \le B_\varphi$.

Assumption 4 regarding the convexity and Lipschitz continuity of the smoothed functions ensures that the gradient estimates obtained through bandit feedback are well-behaved, which is critical for the convergence analysis of zeroth-order optimization algorithms [14].

In this paper, we consider a smoothed function $\tilde{\varphi}$ of a function φ as follows:

$$\tilde{\varphi}(x) = \mathbb{E}_{v\in\mathbb{B}^n}[\varphi(x + \delta_\varphi v)], \tag{5}$$

where $\delta_\varphi \in \left(0, \zeta_{\min}\xi\right]$ is a smoothing parameter and $\xi \in (0, 1)$ is a shrink rate. The expectation in the smoothed function (5) is taken with respect to a random vector v uniformly distributed over the unit ball $\mathbb{B}^n$ in $\mathbb{R}^n$. The random term $\delta_\varphi v$ introduces a small perturbation around the point x, allowing us to explore the local behaviour of the function φ. By averaging over these random perturbations, we obtain a smoothed approximation that captures the essential features of the original function. The shrink rate ξ controls the size of the neighbourhood. A smaller value of ξ results in a tighter approximation to the original function, while a larger value leads to a smoother approximation with a larger neighbourhood of exploration.

Under Assumption 4, we have the following results of convexity and Lipschitz continuity for the smoothed function $\tilde{\varphi}$ (Lemma 2 in [25]).

Lemma 1. *Under Assumption 4, we have the followings:*

1. *$\tilde{\varphi}$ is convex on $(1-\xi)\mathbb{K}$.*
2. *$\tilde{\varphi}$ is Lipschitz continuous, that is, for all $x, y \in \mathbb{K}$, there exists a positive constant L_φ such that $\|\tilde{\varphi}(x) - \tilde{\varphi}(y)\| \le L_\varphi\|x - y\|$.*
3. *For any $x \in (1-\xi)\mathbb{K}$, we have $\varphi(x) \le \tilde{\varphi}(x)$ and $\|\tilde{\varphi}(x) - \varphi(x)\| \le \delta_\varphi L_\varphi$.*

We consider the two-point approximation of the gradient of φ defined as follows:

$$\tilde{d}_\varphi(x) = \frac{n}{2\delta_\varphi}(\varphi(x + \delta_\varphi u_\varphi) - \varphi(x - \delta_\varphi u_\varphi))u_\varphi, \tag{6}$$

where $u_\varphi \in \mathbb{S}^n$ is a unit vector. The intuition behind this approximation is to estimate the gradient by evaluating the function at two points that are slightly perturbed from the original point x.

The properties of $\tilde{d}_\varphi$ are summarized in the following lemma (Lemma 2 in [25]):

Lemma 2. *Under Assumption 4, we have the followings:*

1. *The expected value of $\tilde{d}_\varphi(x)$ is given by the smoothed gradient, that is,*

$$\mathbb{E}_{u_\varphi\in\mathbb{S}^n}[\tilde{d}_\varphi(x)] = \nabla\tilde{\varphi}(x), \quad \forall x \in (1-\xi)\mathbb{K}. \tag{7}$$

2. *There exists a positive constant L_φ such that*

$$\|\nabla\tilde{\varphi}(x) - \nabla\tilde{\varphi}(y)\| \le \frac{nL_\varphi}{\delta_\varphi}\|x - y\|, \quad \forall x, y \in (1-\xi)\mathbb{K}. \tag{8}$$

3. *The norm of $\tilde{d}_\varphi(x)$ is bounded as*

$$\|\tilde{d}_\varphi(x)\| \le nL_\varphi, \quad \forall x \in (1-\xi)\mathbb{K}. \tag{9}$$

The properties in Lemma 2 are fundamental for the analysis of the proposed algorithm because they address the behaviour of the gradient estimates and its smoothed function. These properties are used in the convergence analysis in Section 4 to bound the error terms.

3 | DISTRIBUTED BANDIT FEEDBACK OPTIMIZATION WITH COMMUNICATION DELAYS

In this section, we propose a distributed online algorithm with bandit feedback in the presence of communication delays. Let $x_*^{(k)} = [(x_{1,*}^{(k)})^\top, (x_{2,*}^{(k)})^\top, \ldots, (x_{N,*}^{(k)})^\top]^\top$ be an optimal strategy at iteration k such that $x_*^{(k)} \in \arg\min_{x\in\mathcal{Y}_k} F_k(x)$. Each agent i has the estimations $x_i^{(k)} \in \mathcal{X}_i$ and $\rho_i^{(k)} \in \mathbb{R}_+^m$ corresponding to the optimal dynamic strategy $x_{i,*}^{(k)}$ and the dual optimal solution $\rho_* \in \mathbb{R}_+^m$ for the dual problem (4).

Let $\kappa_{ij}^{(k)} \in \mathbb{N}$ be the time-varying communication delay for the directed edge $(j, i) \in \mathcal{E}_k$ at iteration k. We introduce the following assumption [17]:

Assumption 5. The communication delays are characterized by the following conditions:

- The self-delay is zero, that is, $\kappa_{ii}^{(k)} = 0$ for all $i \in \mathcal{I}$ and $k \in \mathcal{K}$.
- If $(j, i) \in \mathcal{E}_k$ and agent i does not receive the estimate from agent j at iteration $k+1$, then $\kappa_{ij}^{(k)} = 0$.
- There exists a constant $\mu \in \mathbb{N}\setminus\{0\}$ such that $0 \le \kappa_{ij}^{(k)} \le \mu - 1$ for all $i, j \in \mathcal{I}$ and $k \in \mathcal{K}$.

Assumption 5 is used for analysing and ensuring the convergence of the distributed optimization algorithm. By bounding the delays, the algorithm can be effectively designed to manage delays within a specific range. This assumption is reasonable in many practical systems where communication delays are limited due to physical constraints or the properties of the network [17].

Algorithm 1 shows the proposed distributed online algorithm with bandit feedback. The algorithm utilizes the step-sizes

ALGORITHM 1 Distributed online primal-dual push-sum algorithm with Bandit feedback.

Require: For each agent $i \in \mathcal{J}$ and for each iteration $\tau \in \{-(\mu-1), -(\mu-2), \dots, 0\}$, the variables $x_i^{(\tau)}$, $\rho_i^{(\tau)}$, and $y_i^{(\tau)}$ are initialized to $\mathbf{0}$. Moreover, for all $i \in \mathcal{J}$, the variables $x_i^{(1)}$, $\rho_i^{(1)}$, and $y_i^{(1)}$ are initialized as $x_i^{(1)} \in \mathcal{X}_i$, $\rho_i^{(1)} = \mathbf{0}$, and $y_i^{(1)} = g_{i,k}(x_i^{(1)})$.

for $k \in \mathcal{K}$ **do**

Update the estimations by

$$\psi_i^{(k+1)} = \sum_{j \in \mathcal{J}} q_{ij}^{(k)} \psi_j^{(k-\kappa_{ij}^{(k)})}, \tag{13}$$

$$\xi_i^{(k)} = \sum_{j \in \mathcal{J}} q_{ij}^{(k)} \rho_j^{(k-\kappa_{ij}^{(k)})}, \tag{14}$$

$$\zeta_i^{(k)} = \sum_{j \in \mathcal{J}} q_{ij}^{(k)} y_j^{(k-\kappa_{ij}^{(k)})}, \tag{15}$$

$$x_i^{(k+1)} = \Pi_{(1-\xi)\mathcal{X}_i}\left(x_i^{(k)} - a^{(k)}\tilde{s}_i^{(k+1)}\right), \tag{16}$$

$$\rho_i^{(k+1)} = \left[\xi_i^{(k)} + a^{(k)}\left(\frac{\zeta_i^{(k)}}{\psi_i^{(k+1)}} - b^{(k)}\xi_i^{(k)}\right)\right]_+, \tag{17}$$

$$y_i^{(k+1)} = \zeta_i^{(k)} + g_{i,k}(x_i^{(k)}) - g_{i,k-1}(x_i^{(k-1)}). \tag{18}$$

Set to $k := k+1$.

end for

$a^{(k)} > 0$ and $b^{(k)} > 0$, and the gradient-related vector defined as follows:

$$\tilde{s}_i^{(k+1)} = \tilde{d}_{f_{i,k}}(x_i^{(k)}) + \tilde{D}_{g_{i,k}}(x_i^{(k)})^\top \frac{\xi_i^{(k)}}{\psi_i^{(k+1)}}, \tag{10}$$

where $\tilde{d}_{f_{i,k}}(x_i^{(k)})$ and $\tilde{D}_{g_{i,k}}(x_i^{(k)})$ are the two-point approximations defined as follows:

$$\tilde{d}_{f_{i,k}}(x_i^{(k)}) = \frac{n_i}{2\delta_{f_{i,k}}}(f_{i,k}(x_i^{(k)} + \delta_{f_{i,k}}u_{f_{i,k}}) - f_{i,k}(x_i^{(k)} - \delta_{f_{i,k}}u_{f_{i,k}}))u_{f_{i,k}}, \tag{11}$$

$$\tilde{D}_{g_{i,k}}(x_i^{(k)}) = \frac{n_i}{2\delta_{g_{i,k}}}(g_{i,k}(x_i^{(k)} + \delta_{g_{i,k}}u_{g_{i,k}}) - g_{i,k}(x_i^{(k)} - \delta_{g_{i,k}}u_{g_{i,k}}))u_{g_{i,k}}. \tag{12}$$

The gradient estimates are obtained using smoothing parameters $\delta_{f_{i,k}} > 0$ and $\delta_{g_{i,k}} > 0$, and unit vectors $u_{f_{i,k}} \in \mathbb{S}^{n_i}$ and $u_{g_{i,k}} \in \mathbb{S}^m$ that are uniformly chosen from the unit sphere.

In the proposed algorithm, each agent $i \in \mathcal{J}$ initializes the variables $x_i^{(\tau)}$, $\rho_i^{(\tau)}$, and $y_i^{(\tau)}$ to zero for $\tau \in \{-(\mu-1), -(\mu-2), \dots, 0\}$. Moreover, $x_i^{(1)}$ is set to an initial point in the feasible set $\mathcal{X}_i$, $\rho_i^{(1)}$ is set to zero, and $y_i^{(1)}$ is initialized with the value of the constraint function $g_{i,k}(x_i^{(1)})$.

At each iteration k, agent i updates the variables $\psi_i^{(k+1)}$, $\xi_i^{(k)}$, and $\zeta_i^{(k)}$ by combining the information received from neighbouring agents considering the communication delay $\kappa_{ij}^{(k)}$ in Equations (13)–(15). In Algorithm 1, the variable $\psi_i^{(k)}$ is introduced in Equation (13) to address the imbalance of information that arises due to the directed nature of the communication among agents [8].

The primal variable $x_i^{(k+1)}$ is updated using a projected gradient descent step in Equation (16). The dual variable $\rho_i^{(k+1)}$ is updated using a dual ascent step, followed by a projection onto the non-negative orthant in Equation (17). Finally, the variable $y_i^{(k+1)}$ is updated by incorporating the change in the constraint function values between consecutive iterations in Equation (18). The variable $y_i^{(k+1)}$ provides a running total of how far the current estimates $x_i^{(k)}$ are from satisfying the constraints up to the current iteration k. By maintaining this historical information, $y_i^{(k)}$ enables agent i to adapt the estimate towards feasibility in subsequent iterations.

4 | CONVERGENCE ANALYSIS

Since the gradient information $\tilde{d}_{f_{i,k}}(x_i^{(k)})$ and $\tilde{D}_{g_{i,k}}(x_i^{(k)})$ in Equation (10) is estimated based on sampled function values rather than being directly observed, the gradient descent step (16) involves randomness in the decision-making at each iteration. As a result, the regret and constraint violation are not deterministic but involve random variables [25, 26]. In this section, we analyse the convergence of the proposed algorithm using the expected dynamic regret and the expected constraint violation as follows:

$$\widetilde{\mathrm{Reg}} = \mathbb{E}\left[\sum_{k \in \mathcal{K}} \sum_{i \in \mathcal{J}} f_{i,k}\left(x_i^{(k)}\right)\right] - \sum_{k \in \mathcal{K}} \sum_{i \in \mathcal{J}} f_{i,k}\left(x_{i,*}^{(k)}\right), \tag{19}$$

$$\widetilde{\mathrm{Reg}}^{c} = \mathbb{E}\left[\left\|\left[\sum_{k \in \mathcal{K}} \sum_{i \in \mathcal{J}} g_{i,k}\left(x_i^{(k)}\right)\right]_+\right\|\right]. \tag{20}$$

To analyse the impact of communication delays on the convergence of the distributed optimization algorithm, we introduce an enlarged graph $\breve{\mathcal{G}}_k = (\breve{\mathcal{J}}, \breve{\mathcal{E}}_k)$. In the enlarged graph $\breve{\mathcal{G}}_k$, the vertex set $\breve{\mathcal{J}}$ consists of μN nodes. The first N nodes in $\breve{\mathcal{J}}$, labelled as $1, 2, \dots, N$, correspond to the original agents in $\mathcal{G}_k$. In addition to the original agents, the enlarged graph introduces $(\mu - 1)N$ new nodes, referred to as delayed agents. These delayed agents are labelled as $N+1, N+2, \dots, \mu N$. Each delayed agent handles a copy of an original agent's state from a previous iteration. By including these delayed agents in the enlarged graph, the outdated information can be properly tracked.

The connectivity between the original agents in $\mathcal{G}_k$ remains unchanged in the enlarged graph $\breve{\mathcal{G}}_k$. However, when an original agent i in $\mathcal{J}$ sends its estimate $x_i^{(k)}$ to another agent j with a τ-step communication delay, the estimate is sent to the delayed agent labelled as $j + \tau N$ in the enlarged graph. To capture the communication patterns and delays in the enlarged graph, an enlarged weight matrix $\breve{Q}^{(k)} = [\breve{q}_{hi}^{(k)}] \in \mathbb{R}^{\mu N \times \mu N}$ is introduced. The elements of this matrix $\breve{q}_{hi}^{(k)}$ are defined as follows:

- If agent j receives an estimate from agent i with a τ-step communication delay, that is, $\tau = \kappa_{ji}^{(k)}$, then $\breve{q}_{hi}^{(k)} = q_{ji}^{(k)}$, where $h = j + \tau N$. This means that the weight assigned to the delayed agent h receiving the estimate from agent i is the same as the weight assigned to the original agent j receiving the estimate from agent i in the original graph.
- If $h = i - N$, then $\breve{q}_{hi}^{(k)} = 1$. This condition ensures that each delayed agent retains its own information from the previous iteration.
- In all other cases, $\breve{q}_{hi}^{(k)} = 0$, indicating that there is no communication link between the irrelevant agents.

Agent i's state in the enlarged graph $\breve{\mathcal{G}}_k$ is represented by $\breve{x}_i^{(k)} \in \mathbb{R}^{n_i}$. For all $k \in \mathcal{K}$, the state of the original agent i is set as $\breve{x}_i^{(k)} = x_i^{(k)}$, while the state of the delayed agent $i \in \mathcal{I}_{\mathrm{d}} = \{N+1, N+2, \dots, \mu N\}$ is given by $\breve{x}_i^{(k+1)} = \breve{x}_{i-N}^{(k)}$. The delayed agent's cost and constraint values are set as $f_{i,k}(\breve{x}_i^{(k)}) = 0$ and $g_{i,k}(\breve{x}_i^{(k)}) = \mathbf{0}$ for all $i \in \mathcal{I}$ and $k \in \mathcal{K}$.

The proposed algorithm for agent i on the enlarged graph $\breve{\mathcal{G}}_k$ can be expressed as follows:

$$\breve{\psi}_i^{(k+1)} = \sum_{j \in \breve{\mathcal{I}}} \breve{q}_{ij}^{(k)} \breve{\psi}_j^{(k)}, \tag{21}$$

$$\breve{\xi}_i^{(k)} = \sum_{j \in \breve{\mathcal{I}}} \breve{q}_{ij}^{(k)} \breve{\rho}_j^{(k)}, \tag{22}$$

$$\breve{\zeta}_i^{(k)} = \sum_{j \in \breve{\mathcal{I}}} \breve{q}_{ij}^{(k)} \breve{y}_j^{(k)}, \tag{23}$$

$$\breve{x}_i^{(k+1)} = \Pi_{\mathcal{X}_i}\left(\breve{x}_i^{(k)} - a^{(k)} \breve{s}_i^{(k+1)}\right), \tag{24}$$

$$\breve{\rho}_i^{(k+1)} = \left[\breve{\xi}_i^{(k)} + a^{(k)}\left(\frac{\breve{\zeta}_i^{(k)}}{\breve{\psi}_i^{(k+1)}} - b^{(k)} \breve{\xi}_i^{(k)}\right)\right]_+, \tag{25}$$

$$\breve{y}_i^{(k+1)} = \breve{\zeta}_i^{(k)} + g_{i,k}\left(\breve{x}_i^{(k)}\right) - g_{i,k-1}\left(\breve{x}_i^{(k-1)}\right), \tag{26}$$

where $\breve{\psi}_i^{(0)} = 1$, $\breve{\rho}_i^{(0)} = \mathbf{0}$, $g_{i,k}(\breve{x}_i^{(-1)}) = g_{i,k}(\breve{x}_i^{(0)})$, $\breve{y}_i^{(0)} = g_{i,k}(\breve{x}_i^{(0)})$, and $\breve{s}_i^{(k+1)} = \tilde{d}_{f_{i,k}}(\breve{x}_i^{(k)}) + \tilde{D}_{g_{i,k}}(\breve{x}_i^{(k)})^\top \breve{\xi}_i^{(k)} / \breve{\psi}_i^{(k+1)}$. The initial states are set as $\breve{x}_i^{(0)} = x_i^{(0)}$ for $i \in \mathcal{I}$ and $\breve{x}_i^{(0)} = \mathbf{0}$ for $i \in \mathcal{I}_{\mathrm{d}}$.

From Assumption 1, the enlarged weight matrix $\breve{Q}^{(k)}$ is also column stochastic. From the column stochasticity of $\breve{Q}^{(k)}$, we can make use of the exponential decay property as follows [17]:

Lemma 3. *Under Assumptions 1–3, and 5, for every pair of indices $i, j \in \mathcal{I}$ and for any $k \geq s \geq 0$, the following inequality holds: $|[\breve{Q}(k,s)]_{ij} - \breve{\phi}_i^{(s)}| \leq C\sigma^{k-s}$, where $C > 0$, $0 < \sigma < 1$, $\breve{Q}(k,s) = \breve{Q}^{(k)}\breve{Q}^{(k-1)} \cdots \breve{Q}^{(s)}$ is the product sequence of the weight matrices, and $\breve{\phi}^{(s)}$ is a stochastic vector in $\mathbb{R}^{\mu N}$.*

Here, we summarize several preliminary lemmas that play important roles in the analysis of the regret.

Lemma 4. *Under Assumptions 1–5, we have*

$$\left\| \frac{\breve{\zeta}_i^{(k)}}{\breve{\psi}_i^{(k+1)}} - \bar{\breve{y}}^{(k)} \right\| \leq \frac{2C}{r}\Big(\sigma^k \|\breve{y}^{(0)}\|_1 + \sum_{s=1}^{k} \sigma^{k-s} \|g_{s-1}(\breve{x}^{(s-1)}) - g_{s-2}(\breve{x}^{(s-2)})\|_1\Big),$$

$$\left\| \frac{\breve{\xi}_i^{(k)}}{\breve{\psi}_i^{(k+1)}} - \bar{\breve{\rho}}^{(k)} \right\| \leq \frac{2C}{r}\left(\sigma^k \|\breve{\rho}^{(0)}\|_1 + \sum_{s=1}^{k} \sigma^{k-s} \left\|\varepsilon_{\breve{\rho}^{(s)}}\right\|_1\right),$$

where $g_k(x) = [g_{1,k}^\top(x), g_{2,k}^\top(x), \dots, g_{N,k}^\top(x)]^\top$, $r = \inf_{k \in \mathcal{K}} (\min_{i \in \mathcal{I}} [\breve{Q}(k,0)\mathbf{1}]_i)$, $\bar{\breve{y}}^{(k)} = (1/(\mu N)) \sum_{i \in \breve{\mathcal{I}}} \breve{y}_i^{(k)}$, $\bar{\breve{\rho}}^{(k)} = (1/(\mu N)) \sum_{i \in \breve{\mathcal{I}}} \breve{\rho}_i^{(k)}$, $\varepsilon_{\breve{\rho}}^{(k)} = [(\varepsilon_{\breve{\rho}_1}^{(k)})^\top, (\varepsilon_{\breve{\rho}_2}^{(k)})^\top, \dots, (\varepsilon_{\breve{\rho}_N}^{(k)})^\top]^\top$, *and* $\varepsilon_{\breve{\rho}_i}^{(k)} = \left[\breve{\xi}_i^{(k-1)} + a^{(k-1)}((\breve{\zeta}_i^{(k-1)}/\breve{\psi}_i^{(k)}) - b^{(k-1)}\breve{\xi}_i^{(k-1)})\right]_+ - \breve{\xi}_i^{(k-1)}$.

Lemma 5. *Under Assumptions 1–5, for any $i \in \breve{\mathcal{I}}$ and $k \in \mathcal{K}$, we have $r \leq \breve{\psi}_i^{(k)} \leq \mu N$ and $\|\bar{\breve{y}}^{(k)}\| \leq B_g$, where $0 < r \leq 1$.*

Lemma 6. *Under Assumptions 1–5, for all $i \in \breve{\mathcal{I}}$ and $k \in \mathcal{K}$, there exists a positive constant B_y such that $\|\breve{y}_i^{(k)}\| \leq B_y$ and $\|\breve{\zeta}_i^{(k)}\| \leq B_y$. Moreover, we have $\|\breve{\xi}_i^{(k)}\| \leq \breve{\psi}_i^{(k+1)} B_y/(b^{(k)} r^2)$ and $\|\breve{\rho}_i^{(k)}\| \leq \breve{\psi}_i^{(k+1)} B_y/(b^{(k)} r^2 p)$ for all $i \in \breve{\mathcal{I}}$ and $k \in \mathcal{K}$.*

Lemmas 4–6 provide upper-bounds for norms of the intermediate variables in the enlarged system, which helps in ensuring that these variables do not grow unboundedly.

Lemma 7. *Under Assumptions 1–5, we have*

$$\mathbb{E}\left[\sum_{k \in \mathcal{K}} \sum_{i \in \breve{\mathcal{I}}} \left\| \frac{\breve{\xi}_i^{(k)}}{\breve{\psi}_i^{(k+1)}} - \bar{\breve{\rho}}^{(k)} \right\|\right] \leq \frac{4C\sqrt{m}(\mu N)^2 B_y}{r^3(1-\sigma)} \sum_{k=0}^{T-1} a^{(k)},$$

$$\mathbb{E}\left[\sum_{k \in \mathcal{K}} \sum_{i \in \breve{\mathcal{I}}} \left\| \frac{\breve{\zeta}_i^{(k)}}{\breve{\psi}_i^{(k+1)}} - \bar{\breve{y}}^{(k)} \right\|\right] \leq \frac{2C\mu N \|\breve{y}^{(0)}\|_1}{r(1-\sigma)} + \frac{2C(\mu N)^2\sqrt{mn}L_f C_g}{r(1-\sigma)} \sum_{k=0}^{T-1} a^{(k)} + \frac{2C(\mu N)^2\sqrt{mn}L_g C_g B_y}{r^3(1-\sigma)} \sum_{k=0}^{T-1} \frac{a^{(k)}}{b^{(k)}},$$

where $L_f = \max_{k \in \mathcal{K}, i \in \mathcal{I}} L_{f_{i,k}}$ and $L_g = \max_{k \in \mathcal{K}, i \in \mathcal{I}} L_{g_{i,k}}$.

Lemma 7 provides an upper bound for the cumulative deviations of the scaled vector over the entire iterations.

The proofs for Lemmas 4–7 follow the similar arguments in Lemmas 2–5 of the reference [29] and are not included in this paper for brevity.

Next, we present a lemma that evaluates the error introduced by estimating the Lagrangian function. This lemma plays a crucial role in analysing the convergence properties of the proposed algorithm.

Lemma 8. *Under Assumptions 1–5, for any* $\breve{x} = [\breve{x}_1^\top, \breve{x}_2^\top, \dots, \breve{x}_{\mu N}^\top]^\top$ *and* $\breve{\rho} \in \mathbb{R}_+^m$*, we have*

$$\mathbb{E}\big[H_k(\breve{x}^{(k)}, \bar{\rho}^{(k)}) - H_k(\breve{x}, \bar{\rho}^{(k)})\big] \leq \frac{1}{2a^{(k)}} \sum_{i\in\breve{\mathcal{I}}} \left(\left\| \breve{x}_i^{(k)} - \breve{x}_i \right\|^2 - \left\| \breve{x}_i^{(k+1)} - \breve{x}_i \right\|^2 \right) + \frac{a^{(k)} \sum_{i\in\breve{\mathcal{I}}} n_i^2}{2} \left(L_f + \frac{L_g B_y}{b^{(k)} r^2} \right)^2 + 2B_g \sum_{i\in\breve{\mathcal{I}}} \left\| \frac{\breve{\xi}_i^{(k)}}{\breve{\psi}_i^{(k+1)}} - \bar{\rho}^{(k)} \right\| + \delta_f^{(k)} L_f + \delta_g^{(k)} L_g, \quad (27)$$

$$\mathbb{E}\big[H_k(\breve{x}^{(k)}, \breve{\rho}) - H_k(\breve{x}^{(k)}, \bar{\rho}^{(k)})\big] \leq \frac{\mu N}{2a^{(k)}} (\|\bar{\rho}^{(k)} - \breve{\rho}\|^2 - \|\bar{\rho}^{(k+1)} - \breve{\rho}\|^2) + \left(\|\breve{\rho}\| + \frac{B_y}{b^{(k)} r^2} \right) \sum_{i\in\breve{\mathcal{I}}} \left\| \frac{\breve{\zeta}_i^{(k)}}{\breve{\psi}_i^{(k+1)}} - \bar{y}^{(k)} \right\| + \frac{2(\mu N)^3 B_y^2 (r+2)}{r^5} a^{(k)} + \frac{(\mu N)^2 b^{(k)}}{2} \|\breve{\rho}\|^2 + \left(B_g + \frac{2\mu N B_y}{r^2} \right) \sum_{i\in\breve{\mathcal{I}}} \left\| \frac{\breve{\xi}_i^{(k)}}{\breve{\psi}_i^{(k+1)}} - \bar{\rho}^{(k)} \right\|, \quad (28)$$

where $\breve{x}_i \in \mathcal{X}_i$, $\breve{x}^{(k)} = [(\breve{x}_1^{(k)})^\top, (\breve{x}_2^{(k)})^\top, \dots, (\breve{x}_{\mu N}^{(k)})^\top]^\top$, $\delta_f^{(k)} = \sum_{i\in\breve{\mathcal{I}}} \delta_{f_i,k}$*, and* $\delta_g^{(k)} = \sum_{i\in\breve{\mathcal{I}}} \delta_{g_i,k}$.

The proof of Lemma 8 is given in Appendix A. Lemma 8 provides bounds on the error between the Lagrangian function values at the estimated and optimal solutions. This lemma establishes that each agent can manage the errors effectively, ensuring that the deviation remains within acceptable bounds.

Finally, we present a theorem that quantifies the performance of the proposed distributed online algorithm under the constraints of bandit feedback and communication delays.

Theorem 1. *Suppose that the step-sizes* $a^{(k)}$ *and* $b^{(k)}$ *are defined as* $a^{(k)} = 1/\sqrt{k}$ *and* $b^{(k)} = 1/k^c$ *with* $c \in (0, 1/4)$*, where* $a^{(0)} = 1$ *and* $b^{(0)} = 1$*. Suppose also that the smoothing parameters are set as* $\delta_{f_i,k}^{(k)} = h_{f_i}/k$ *and* $\delta_{g_i,k}^{(k)} = h_{g_i}/k$ *with positive constants* h_{f_i} *and* h_{g_i}*. Under Assumptions 1–5, we have*

$$\widetilde{Reg} = O_+\left(T^{\frac{1}{2}+2c}\right) + O_+(V_T), \quad (29)$$

$$\widetilde{Reg}^c = O\left(T^{1-\frac{c}{2}}\right) + O\left(V_T^{\frac{1}{2}} T^{\frac{1-c}{2}}\right), \quad (30)$$

where $V_T = \sum_{k\in\mathcal{K}} (1/a^{(k)}) \sum_{i\in\mathcal{I}} \|x_{i,*}^{(k+1)} - x_{i,*}^{(k)}\|$.

The proof of Theorem 1 is given in Appendix B.

Theorem 1 shows that the proposed distributed online optimization algorithm achieves sublinear dynamic regret (19) and constraint violation (20) under the constraints of bandit feedback and communication delays. The bounds in Equations (29) and (30) consist of two terms each. The first term in both bounds depends on the time horizon T and the choice of the parameter c of the step-size $b^{(k)}$. By setting $c \in (0, 1/4)$, these terms grow sublinearly with respect to T. The second term in both bounds depends on the accumulated variation of the optimal strategies V_T. This term captures the impact of the time-varying nature of the problem on the algorithm's performance. When the variation of the optimal strategies is sufficiently small, that is, when V_T grows slowly with time, the overall regret and constraint violation remain sublinear. This implies that the estimation of each agent approaches the optimal strategy over time.

5 | NUMERICAL EXAMPLE

First, the performance of the proposed algorithm is evaluated through a resource allocation problem in a large-scale power network system that consists of 1000 generator agents ($N = 1000$). The problem objective is to minimize the total quadratic cost of power generation while satisfying time-varying constraints on the total power output as follows [33]:

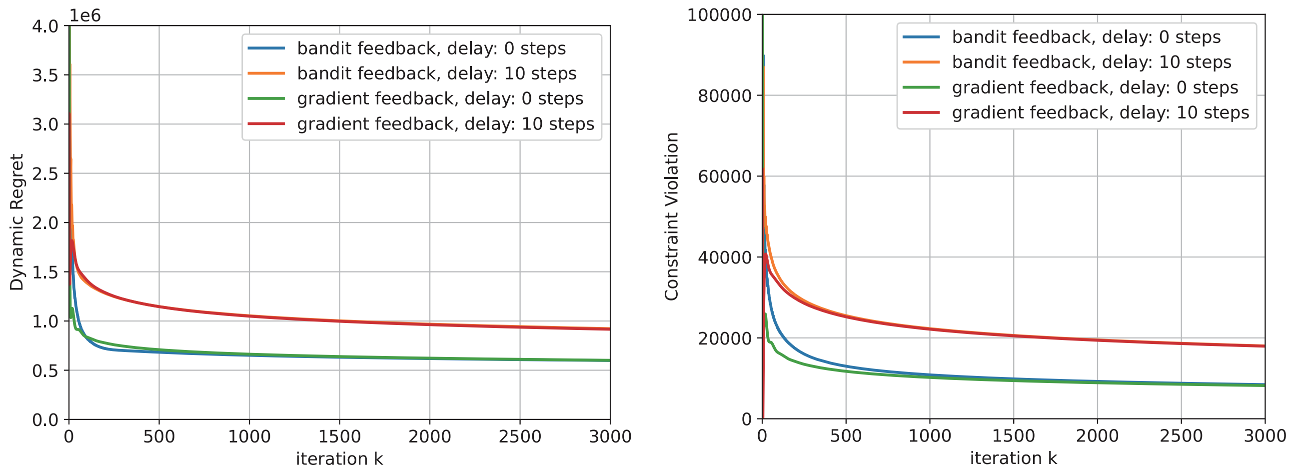
$$\underset{x\in\mathcal{X}}{\text{minimize}} \sum_{k\in\mathcal{K}} \sum_{i\in\mathcal{I}} \left(c_{1i}^{(k)} \left(x_i^{(k)} \right)^2 + c_{2i}^{(k)} x_i^{(k)} + c_{3i}^{(k)} \right) \quad (31a)$$

$$\text{subject to} \sum_{k\in\mathcal{K}} \sum_{i\in\mathcal{I}} \left(x_i^{(k)} - \frac{p_{\max}^{(k)}}{N} \right) \leq 0, \quad (31b)$$

$$\sum_{k\in\mathcal{K}} \sum_{i\in\mathcal{I}} \left(-x_i^{(k)} + \frac{p_{\min}^{(k)}}{N} \right) \leq 0, \quad (31c)$$

where $x_i \in \mathcal{X}_i \subset \mathbb{R}$ is agent i's output power, $p_{\max}^{(k)}$ and $p_{\min}^{(k)}$ are the upper-bound and the lower-bound of power demands at iteration k, respectively. The constraint set is given by $\mathcal{X}_i = [50, 80]$. The bounds of the demand are set as $p_{\max}^{(k)} \in [290N + 0.1, 290N + 20.1]$ and $p_{\min}^{(k)} \in [290N - 20.1, 290N - 0.1]$. The time-varying coefficients of the cost function are randomly generated as $c_{1i}^{(k)} \in [0.025, 0.03]$, $c_{2i}^{(k)} \in [15, 20]$, and $c_{3i}^{(k)} \in [25, 30]$. The step-sizes are given by $a^{(k)} = 1/\sqrt{k}$ and $b^{(k)} = 1/k^{0.2}$, and the smoothing parameters are set as $\delta_{f_i,k}^{(k)} = 0.1/k$ and $\delta_{g_i,k}^{(k)} = 0.1/k$.

Figure 1 illustrates the evolution of the time-averaged dynamic regret and the time-averaged constraint violation for different algorithms. The blue line represents the performance of the bandit feedback algorithm without communication delay [25]. The orange line shows the result of the proposed method for the bandit feedback setting with communication delays. The green line corresponds to the gradient feedback algorithm without communication delays [8]. Lastly, the red line depicts the performance of the gradient feedback algorithm



(a) Dynamic Regret. (b) Constraint Violation.

FIGURE 1 Evolution of the time-averaged dynamic regret and the time-averaged constraint violation for resource allocation.

with communication delays [29]. The communication delays are randomly generated between 0 to 10 iteration steps.

These results show that the proposed algorithm achieves sublinear growth in both the dynamic regret and constraint violation even in the presence of bandit feedback and communication delays. Although the performance is slightly degraded compared to using true gradients, the results show the effectiveness of the bandit gradient estimation in approximating the true gradients. Moreover, the results show that the algorithm is robust to communication delays as the performance remains relatively unaffected. However, we note that it is crucial to avoid constraint violations by configuring with stricter parameters in power control. Furthermore, the algorithm's output should be used with robust control mechanisms or safety constraints to ensure compliance with critical operational limits. Therefore, while the sublinear constraint violation indicates that constraint violations become less frequent, practical implementations would need to incorporate additional safeguards to meet the strict reliability standards required in power systems.

Next, we present a numerical experiment on intersection control with connected automated vehicles (CAVs) [29, 34]. We consider a scenario where 8 vehicles enter an intersection from 4 different directions as illustrated in Figure 2. The intersection control problem is formulated as follows:

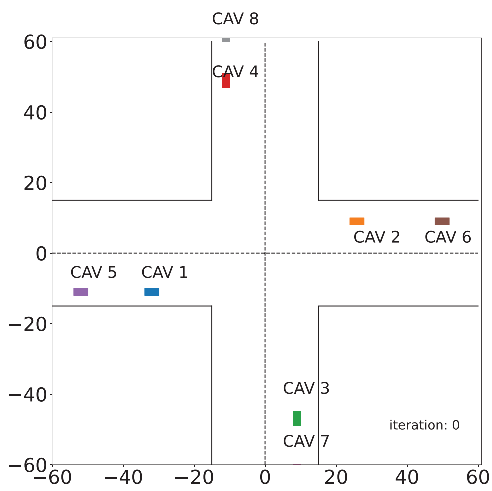
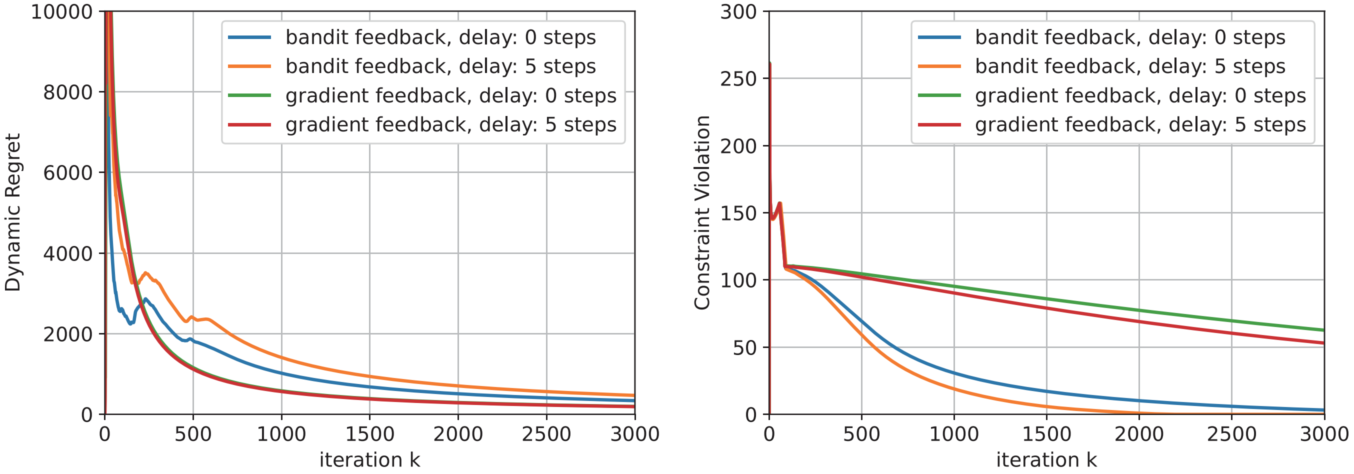


FIGURE 2 Initial positions of 8 CAVs.

where $A_i \in \mathbb{R}^{2\times 2}$ and $B_i \in \mathbb{R}^{2\times 1}$ are the system matrix and input matrix of CAV i, respectively. $\mathcal{X}_i \subset \mathbb{R}^2$ and $\mathcal{Z}_i \subset \mathbb{R}$ are the state and input constraint sets of CAV i. $p_{\mathrm{safe}} \in \mathbb{R}_+$ is the minimum safe distance between any two CAVs. $\chi^{(k)}_{i,k+j} = [r^{(k)}_{i,k+j}, v^{(k)}_{i,k+j}]^\top \in \mathbb{R}^2$ is the state (position $r^{(k)}_{i,k+j}$ and velocity

$$\begin{aligned}
\text{minimize} \quad & \sum_{k=1}^{T}\sum_{i=1}^{N} J_i^{(k)} \\
\text{subject to} \quad & \chi^{(k)}_{i,k+j+1} = A_i \chi^{(k)}_{i,k+j} + B_i u^{(k)}_{i,k+j}, \quad \forall i \in \mathcal{I},\ \forall j \in \{0,1,\dots,K-1\}, \\
& \chi^{(k)}_{i,k+j} \in \mathcal{X}_i, \quad \forall i \in \mathcal{I},\ \forall j \in \{0,1,\dots,K-1\}, \\
& u^{(k)}_{i,k+j} \in \mathcal{Z}_i, \quad \forall i \in \mathcal{I},\ \forall j \in \{0,1,\dots,K-1\}, \\
& p^{(k)}_{i,\ell,k+j} + p^{(k)}_{\ell,i,k+j} \geq p_{\mathrm{safe}}, \quad \forall i \in \mathcal{I},\ \forall \ell \in \check{\mathcal{I}},\ \forall j \in \{0,1,\dots,K-1\},
\end{aligned}$$



(a) $\mathrm{Reg}(T)/T$ (b) $\mathrm{Reg}^{\mathrm{C}}(T)/T$

FIGURE 3 Evolution of the time-averaged dynamic regret and the time-averaged constraint violation for intersection control.

$v_{i,k+j}^{(k)}$) of CAV i at time step $k+j$ based on the information available at time step k. $u_{i,k+j}^{(k)} \in \mathbb{R}$ is the input of CAV i at time step $k+j$. $\check{\mathcal{I}}$ is the set of CAVs that may potentially collide with CAV i. $K \in \mathbb{N}$ is the prediction horizon. The cost function is given by

$$J_i^{(k)} = \sum_{j=0}^{K-1} \left\{ \left(v_{\mathrm{r}} - v_{i,k+j}^{(k)}\right)^{\top} \left(v_{\mathrm{r}} - v_{i,k+j}^{(k)}\right) + \left(u_{i,k+j}^{(k)}\right)^{\top} S_i u_{i,k+j}^{(k)} + \left(u_{i,k+j}^{(k)} - u_{i,k+j-1}^{(k)}\right)^{\top} Q_i \left(u_{i,k+j}^{(k)} - u_{i,k+j-1}^{(k)}\right) \right\} + \left(v_{\mathrm{r}} - v_{i,k+K}^{(k)}\right)^{\top} \left(v_{\mathrm{r}} - v_{i,k+K}^{(k)}\right),$$

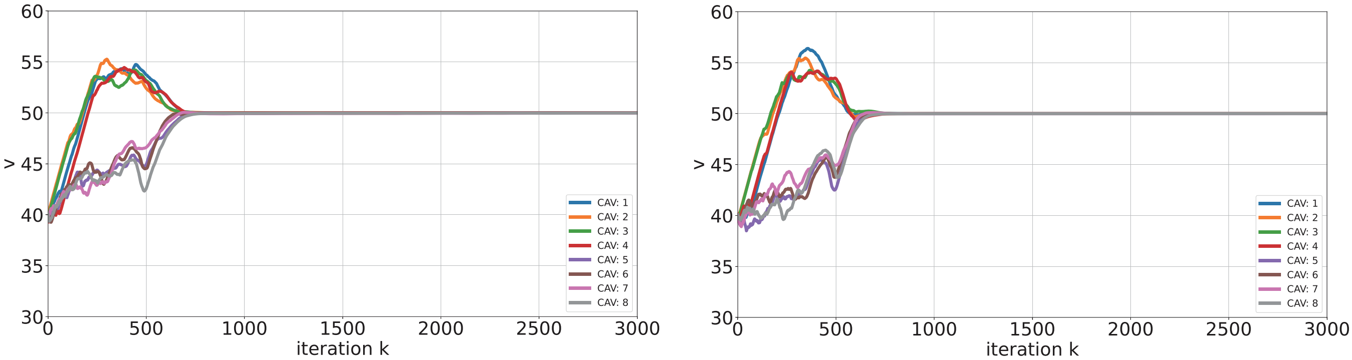
where v_{r} is the reference velocity, and $S_i \in \mathbb{R}^{2\times 2}$ and $Q_i \in \mathbb{R}^{2\times 2}$ are the weight matrices.

The goal is to control the CAVs so that they can pass through the intersection safely with the reference velocity v_{r} (50 [km/h]). For this example, the parameters of the intersection control are set as the same as in the reference [29]. The step-sizes are given by $a^{(k)} = 1/\sqrt{k}$ and $b^{(k)} = 1/k^{0.2}$, and the smoothing parameters are set as $\delta_{f_{i,k}}^{(k)} = 0.1/k$ and $\delta_{g_{i,k}}^{(k)} = 0.1/k$.

Figure 3 shows the evolution of the time-averaged dynamic regret and the time-averaged constraint violation for the intersection control problem. As in the case of the resource allocation problem, we compare the performance of four different algorithms: the bandit feedback algorithm without communication delays [25] (blue line), the proposed method for the bandit feedback setting with communication delays (orange line), the gradient feedback algorithm without communication delays [8] (green line), and the gradient feedback algorithm with communication delays [29] (red line). In this example, the communication delays are randomly generated between 0 to 5 iteration steps.

These results demonstrate that the proposed bandit feedback algorithm with communication delays (orange line) achieves sublinear dynamic regret and constraint violation despite the presence of communication delays and limited feedback information. This indicates that the proposed algorithm is able to effectively control the CAVs and optimize their trajectories while satisfying the constraints. Despite the slight increase in dynamic regret and constraint violation due to communication delays, the proposed method (orange line) demonstrates its resilience. The sublinear growth of both metrics indicates that the algorithm can still converge to the optimal strategy and maintain safety over time. In this example, the gradient feedback algorithms, both without (green line) and with communication delays (red line), achieve lower dynamic regret compared to the bandit feedback algorithms. On the other hand, the gradient feedback algorithms result in higher constraint violations. The gradient feedback algorithms have access to more informative updates as they utilize the gradient of the objective function and constraints. However, the gradient information may not always accurately capture the feasibility of the solutions especially in the presence of complex and time-varying constraints. As a result, the gradient feedback algorithms may violate the constraints more frequently to minimize the objective function in the settings of this numerical example.

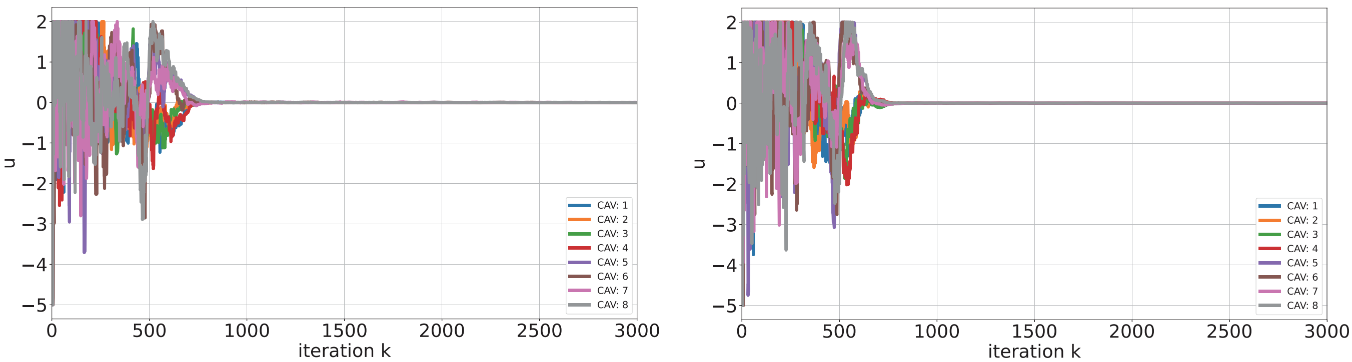
Figures 4 and 5 show the velocities and the inputs of CAVs using the bandit feedback algorithm, comparing scenarios with no communication delays [25] and with 5-step communication delays (proposed algorithm). In the numerical example, the horizontal axis in Figure 4 is labelled by the iteration k to maintain consistency with the horizontal axis in Figure 3. In this example, the sampling period is given by 0.01 s. Therefore, the time t can be directly related to the iteration by $t = 0.01k$. These results indicate that the velocities of all CAVs converge to the reference velocity v_{r} in both cases. However, communication delays lead to more fluctuation in the CAVs' velocities compared to the delay-free scenario. This fluctuation is likely due to the delayed



(a) Without communication delays [25].

(b) With 5-step communication delays.

FIGURE 4 Velocities of 8 CAVs in the bandit feedback algorithm.



(a) Without communication delays [25].

(b) With 5-step communication delays.

FIGURE 5 Inputs of 8 CAVs in the bandit feedback algorithm.

feedback, which causes less timely adjustments. Since the current model does not explicitly account for jerk constraints, the control input exhibits significant variations as shown in Figure 5, leading to relatively large velocity fluctuations in Figure 4. To reduce these fluctuations, the optimization problem needs to be reformulated to include jerk constraints [35]. While reducing jerk would enhance ride comfort for passengers, it could also extend driving time and potentially cause congestion at intersections. Therefore, in practical applications, it is crucial to strike a balance between passenger comfort and minimizing driving time. For future work, we aim to extend the model to include jerk explicitly in the constraint of the optimization problem to provide a more practical control framework.

6 | CONCLUSION

This paper studied the problem of distributed online optimization with bandit feedback and communication delays. We developed a distributed primal-dual algorithm that enables agents to cooperatively optimize a global time-varying objective function while satisfying dynamic constraints using only local bandit feedback and delayed information exchange. The algorithm combines a primal-dual approach with a two-point zeroth-order gradient estimation to handle constrained optimization under bandit feedback. The theoretical analysis was provided to prove the sublinear dynamic regret and constraint violation of the proposed algorithm, demonstrating its tracking performance even with delayed information. The results of this paper open up several interesting directions for future research. One direction is to investigate distributed online optimization with more complex constraint structures.

AUTHOR CONTRIBUTIONS

Keito Inoue: Conceptualization; methodology; writing—original draft; writing—review & editing. **Naoki Hayashi**: Supervision; writing—review & editing. **Shigemasa Takai**: Supervision; writing—review & editing.

CONFLICT OF INTEREST STATEMENT

The authors declare no conflicts of interest.

DATA AVAILABILITY STATEMENT

Author elects to not share data.

ORCID

Naoki Hayashi https://orcid.org/0000-0003-2518-1250

REFERENCES

1. Shorinwa, O., Halsted, T., Yu, J., Schwager, M.: Distributed optimization methods for multi-robot systems: part 2—a survey. IEEE Rob. Autom. Mag. 31(3), 154–169 (2024)
2. Nedić, A., Liu, J.: Distributed optimization for control. Annu. Rev. Control Rob. Auton. Syst. 1(1), 77–103 (2018)
3. Shahrampour, S., Jadbabaie, A.: Distributed online optimization in dynamic environments using mirror descent. IEEE Trans. Autom. Control 63(3), 714–725 (2018)
4. Wang, C., Xu, S., Yuan, D., Zhang, B., Zhang, Z.: Push-sum distributed online optimization with bandit feedback. IEEE Trans. Cybern. 52(4), 2263–2273 (2020)
5. Yamashita, M., Hayashi, N., Hatanaka, T., Takai, S.: Logarithmic regret for distributed online subgradient method over unbalanced directed networks. IEICE Trans. Fundam. Electron., Commun. Comput. Sci. E104-A(8), 1019–1026 (2021)
6. Xu, H., Lu, K., Wang, Y.L.: Online distributed nonconvex optimization with stochastic objective functions: high probability bound analysis of dynamic regrets. Automatica 170, 111863 (2024)
7. Lu, K.: Online distributed algorithms for online noncooperative games with stochastic cost functions: high probability bound of regrets. IEEE Trans. Autom. Control (2024). https://doi.org/10.1109/TAC.2024.3419018
8. Li, X., Yi, X., Xie, L.: Distributed online optimization for multi-agent networks with coupled inequality constraints. IEEE Trans. Autom. Control 66(8), 3575–3591 (2021)
9. Yi, X., Li, X., Xie, L., Johansson, K.H.: Distributed online convex optimization with time-varying coupled inequality constraints. IEEE Trans. Signal Process. 68, 731–746 (2020)
10. Yuan, D., Proutiere, A., Shi, G.: Distributed online optimization with long-term constraints. IEEE Trans. Autom. Control 67(3), 1089–1104 (2022)
11. Zhang, W., Shi, Y., Zhang, B., Lu, K., Yuan, D.: Quantized distributed online projection-free convex optimization. IEEE Control Syst. Lett. 7, 1837–1842 (2023)
12. Cao, X., Başar, T.: Decentralized online convex optimization with event-triggered communications. IEEE Trans. Signal Process. 69, 284–299 (2021)
13. Okamoto, K., Hayashi, N., Takai, S.: Distributed online adaptive gradient descent with event-triggered communication. IEEE Trans. Control Network Syst. 11(2), 610–622 (2024)
14. Wang, C., Xu, S., Yuan, D.: Distributed online stochastic-constrained convex optimization with bandit feedback. IEEE Trans. Cybern. 54(1), 63–75 (2024)
15. Li, J., Gu, C., Wu, Z., Huang, T.: Online learning algorithm for distributed convex optimization with time-varying coupled constraints and bandit feedback. IEEE Trans. Cybern. 52(2), 1009–1020 (2022)
16. Xiong, M., Ho, D.W.C., Zhang, B., Yuan, D., Xu, S.: Distributed online mirror descent with delayed subgradient and event-triggered communications. IEEE Trans. Network Sci. Eng. 11(2), 1702–1715 (2024)
17. Nedić, A., Ozdaglar, A.: Convergence rate for consensus with delays. J. Global Optim. 47, 437–456 (2010)
18. Lin, P., Ren, W.: Constrained consensus in unbalanced networks with communication delays. IEEE Trans. Autom. Control 59(3), 775–781 (2014)
19. Zhang, K., Zhang, Y., Li, R., Chen, G.: Sampled-data-based consensus of multi-agent systems with multiplicative noise and time-delays. Inf. Sci. 609, 1621–1640 (2022)
20. Wei, H., Liu, C., Shi, Y.: A robust distributed MPC framework for multi-agent consensus with communication delays. IEEE Trans. Autom. Control 69(11), 7418–7432 (2024)
21. Wang, D., Liu, J., Lian, J., Liu, Y., Wang, Z., Wang, W.: Distributed delayed dual averaging for distributed optimization over time-varying digraphs. Automatica 150, 110869 (2023)
22. Liu, J., Ho, D.W.C., Li, L.: A generic algorithm framework for distributed optimization over the time-varying Nnetwork with communication delays. IEEE Trans. Autom. Control 69(1), 371–378 (2024)
23. Wei, M., Yu, W., Liu, H., Xu, Q.: Distributed weakly convex optimization under random time-delay interference. IEEE Trans. Network Sci. Eng. 11(1), 212–224 (2024)
24. An, Y., Wang, A., Zhang, X., Xiao, F.: Distributed constrained optimization for multi-agent networks with communication delays under time-varying topologies. Syst. Control Lett. 185, 105733 (2024)
25. Yi, X., Li, X., Yang, T., Xie, L., Chai, T., Johansson, K.H.: Distributed bandit online convex optimization with time-varying coupled inequality constraints. IEEE Trans. Autom. Control 66(10), 4620–4635 (2021)
26. Li, J., Zhu, X., Wu, Z., Huang, T.: Online distributed dual averaging algorithm for multi-agent bandit optimization over time-varying general directed networks. Inf. Sci. 581, 678–693 (2021)
27. Wang, C., Xu, S.: Distributed online constrained optimization with feedback delays. IEEE Trans. Neural Networks Learn. Syst. 35(2), 1708–1720 (2024)
28. Lyu, X., Ni, W., Tian, H., Liu, R.P., Wang, X., Giannakis, G.B.: Distributed online optimization of fog computing for selfish devices with out-of-date information. IEEE Trans. Wireless Commun. 17(11), 7704-7717 (2018)
29. Inoue, K., Hayashi, N., Takai, S.: Distributed online optimization with dynamic coupling constraints under time-varying communication delays. IEEE Access. 11, 87256–87269 (2023)
30. Yi, X., Li, X., Yang, T., Xie, L., Chai, T., Johansson, K.H.: Regret and cumulative constraint violation analysis for distributed online constrained convex optimization. IEEE Trans. Autom. Control 68(5), 2875–2890 (2023)
31. Hayashi, N., Sakurama, K.: Communication-aware distributed rebalancing for cooperative car-sharing service. IET Control Theory Appl. 17(7), 850–867 (2023)
32. Boyd, S., Vandenberghe, L.: Convex Optimization. Cambridge University Press, Cambridge (2004)
33. Yi, P., Hong, Y., Liu, F.: Initialization-free distributed algorithms for optimal resource allocation with feasibility constraints and application to economic dispatch of power systems. Automatica 74, 259–269 (2016)
34. Katriniok, A., Rosarius, B., Mähönen, P.: Fully distributed model predictive control of connected automated vehicles in intersections: theory and vehicle experiments. IEEE Trans. Intell. Transp. Syst. 23(10), 18288–18300 (2022)
35. de Winkel, K.N., Irmak, T., Happee, R., Shyrokau, B.: Standards for passenger comfort in automated vehicles: acceleration and jerk. Appl. Ergon. 106, 103881 (2023)

How to cite this article: Inoue, K., Hayashi, N., Takai, S.: Distributed zeroth-order online optimization with communication delays. IET Control Theory Appl. 19, e12759 (2025). https://doi.org/10.1049/cth2.12759

APPENDIX A: PROOF OF LEMMA 8

We show Equation (27). From Equations (1) and (24), for $\breve{x}_i \in \mathcal{X}_i$, we have

$$\begin{aligned}\left\|\breve{x}_i^{(k+1)} - \breve{x}_i\right\|^2 &\le \left\|\breve{x}_i^{(k)} - \breve{x}_i - a^{(k)}\breve{s}_i^{(k+1)}\right\|^2 \\ &= \left\|\breve{x}_i^{(k)} - \breve{x}_i\right\|^2 + (a^{(k)})^2\left\|\breve{s}_i^{(k+1)}\right\|^2 \\ &\quad - 2a^{(k)}\left(\breve{s}_i^{(k+1)}\right)^\top\left(\breve{x}_i^{(k)} - \breve{x}_i\right). \end{aligned} \tag{A1}$$

From Equation (10), the last term of the right-hand side of Equation (A1) is given as follows:

$$\mathbb{E}\left[-2a^{(k)}\left(\breve{s}_i^{(k+1)}\right)^\top\left(\breve{x}_i^{(k)}-\breve{x}_i\right)\right]$$

$$=-2a^{(k)}\left(\mathbb{E}\left[\tilde{d}_{f_{i,k}}\left(\breve{x}_i^{(k)}\right)\right]+\left(\mathbb{E}\left[\tilde{D}_{g_{i,k}}\left(\breve{x}_i^{(k)}\right)\right]\right)^\top\frac{\xi_i^{(k)}}{\breve{\psi}_i^{(k+1)}}\right)^\top\left(\breve{x}_i^{(k)}-\breve{x}\right)$$

$$\leq -2a^{(k)}\left(\tilde{f}_{i,k}\left(\breve{x}_i^{(k)}\right)-\tilde{f}_{i,t}(\breve{x}_i)+\frac{\xi_i^{(k)}}{\breve{\psi}_i^{(k+1)}}^\top\left(\tilde{g}_{i,k}\left(\breve{x}_i^{(k)}\right)-\tilde{g}_{i,k}(\breve{x}_i)\right)\right)$$

$$=-2a^{(k)}\left(\tilde{H}_{i,k}\left(\breve{x}_i^{(k)},\bar{\rho}^{(k)}\right)-\tilde{H}_{i,k}\left(\breve{x}_i,\bar{\rho}^{(k)}\right)\right)$$

$$-2a^{(k)}\left(\frac{\xi_i^{(k)}}{\breve{\psi}_i^{(k+1)}}-\bar{\rho}^{(k)}\right)^\top\left(\tilde{g}_{i,k}\left(\breve{x}_i^{(k)}\right)-\tilde{g}_{i,k}(\breve{x}_i)\right),$$

where $\tilde{H}_{i,k}(x,\rho)=\tilde{f}_{i,k}(x)+\rho^\top\tilde{g}_{i,k}(x)$. Then, from Lemma 1, we have

$$\tilde{H}_{i,k}\left(\breve{x}_i^{(k)},\bar{\rho}^{(k)}\right)-\tilde{H}_{i,k}\left(\breve{x}_i,\bar{\rho}^{(k)}\right)$$

$$\geq f_{i,k}\left(\breve{x}_i^{(k)}\right)+(\bar{\rho}^{(k)})^\top g_{i,k}\left(\breve{x}_i^{(k)}\right)$$

$$-\left(f_{i,k}(\breve{x}_i)+\delta_{f_{i,k}}^{(k)}L_f+(\bar{\rho}^{(k)})^\top g_{i,k}(\breve{x}_i)+\delta_{g_{i,k}}^{(k)}L_g\right)$$

$$=H_{i,k}\left(\breve{x}_i^{(k)},\bar{\rho}^{(k)}\right)-H_{i,k}\left(\breve{x}_i,\bar{\rho}^{(k)}\right)-\left(\delta_{f_{i,k}}^{(k)}L_f+\delta_{g_{i,k}}^{(k)}L_g\right).$$

Furthermore, from Lemmas 2 and 6, we have

$$\left\|\breve{s}_i^{(k+1)}\right\|=\left\|\tilde{d}_{f_{i,k}}(\breve{x}_i^{(k)})+\tilde{D}_{g_{i,k}}\left(\breve{x}_i^{(k)}\right)^\top\frac{\xi_i^{(k)}}{\breve{\psi}_i^{(k+1)}}\right\|$$

$$\leq\left\|\tilde{d}_{f_{i,k}}(\breve{x}_i^{(k)})\right\|+\frac{\|\xi_i^{(k)}\|}{\breve{\psi}_i^{(k+1)}}\left\|\tilde{D}_{g_{i,k}}\left(\breve{x}_i^{(k)}\right)\right\|$$

$$\leq n_iL_f+\frac{n_iL_gB_y}{b^{(k)}r^2}. \tag{A2}$$

Then, we have

$$\mathbb{E}\left[H_{i,k}\left(\breve{x}_i^{(k)},\bar{\rho}^{(k)}\right)-H_{i,k}\left(\breve{x}_i,\bar{\rho}^{(k)}\right)\right]$$

$$\leq\frac{1}{2a^{(k)}}\left(\left\|\breve{x}_i^{(k)}-\breve{x}_i\right\|^2-\left\|\breve{x}_i^{(k+1)}-\breve{x}_i\right\|^2\right)$$

$$+\left\|\frac{\xi_i^{(k)}}{\breve{\psi}_i^{(k+1)}}-\bar{\rho}^{(k)}\right\|\left\|g_{i,k}\left(\breve{x}_i^{(k)}\right)-g_{i,k}(\breve{x}_i)\right\|$$

$$+\frac{a^{(k)}}{2}\left\|\breve{s}_i^{(k+1)}\right\|^2+\delta_{f_{i,k}}^{(k)}L_f+\delta_{g_{i,k}}^{(k)}L_g$$

$$\leq\frac{1}{2a^{(k)}}\left(\left\|\breve{x}_i^{(k)}-\breve{x}_i\right\|^2-\left\|\breve{x}_i^{(k+1)}-\breve{x}_i\right\|^2\right)$$

$$+\frac{a^{(k)}}{2}\left(n_iL_f+\frac{n_iL_gB_y}{b^{(k)}r^2}\right)^2$$

$$+2B_g\left\|\frac{\xi_i^{(k)}}{\breve{\psi}_i^{(k+1)}}-\bar{\rho}^{(k)}\right\|+\delta_{f_{i,k}}^{(k)}L_f+\delta_{g_{i,k}}^{(k)}L_g. \tag{A3}$$

This yields

$$\mathbb{E}\left[H_k\left(\breve{x}^{(k)},\bar{\rho}^{(k)}\right)-H_k\left(\breve{x},\bar{\rho}^{(k)}\right)\right]$$

$$\leq\frac{1}{2a^{(k)}}\sum_{i\in\breve{\mathcal{J}}}\left(\left\|\breve{x}_i^{(k)}-\breve{x}_i\right\|^2-\left\|\breve{x}_i^{(k+1)}-\breve{x}_i\right\|^2\right)$$

$$+\frac{a^{(k)}\sum_{i\in\breve{\mathcal{J}}}n_i^2}{2}\left(L_f+\frac{L_gB_y}{b^{(k)}r^2}\right)^2$$

$$+2B_g\sum_{i\in\breve{\mathcal{J}}}\left\|\frac{\xi_i^{(k)}}{\breve{\psi}_i^{(k+1)}}-\bar{\rho}^{(k)}\right\|+\delta_f^{(k)}L_f+\delta_g^{(k)}L_g.$$

Equation (28) can be shown in the same way. □

APPENDIX B: PROOF OF THEOREM 1

First, we show Equation (29). From Lemma 8, we have

$$\mathbb{E}\left[H_k(\breve{x}^{(k)},\breve{\rho})-H_k\left(\breve{x},\bar{\rho}^{(k)}\right)\right]$$

$$=\mathbb{E}\left[H_k\left(\breve{x}^{(k)},\bar{\rho}^{(k)}\right)-H_k\left(\breve{x},\bar{\rho}^{(k)}\right)\right]+\mathbb{E}\left[H_k\left(\breve{x}^{(k)},\breve{\rho}\right)-H_k\left(\breve{x}^{(k)},\bar{\rho}^{(k)}\right)\right]$$

$$\leq\frac{1}{2a^{(k)}}\sum_{i\in\breve{\mathcal{J}}}\left(\left\|\breve{x}_i^{(k)}-\breve{x}_i\right\|^2-\left\|\breve{x}_i^{(k+1)}-\breve{x}_i\right\|^2\right)$$

$$+\frac{\mu N}{2a^{(k)}}\left(\|\bar{\rho}^{(k)}-\breve{\rho}\|^2-\|\bar{\rho}^{(k+1)}-\breve{\rho}\|^2\right)$$

$$+\left(3B_g+\frac{2\mu NB_y}{r^2}\right)\sum_{i\in\breve{\mathcal{J}}}\left\|\frac{\xi_i^{(k)}}{\breve{\psi}_i^{(k+1)}}-\bar{\rho}^{(k)}\right\|$$

$$+\left(\|\breve{\rho}\|+\frac{B_y}{b^{(k)}r^2}\right)\sum_{i\in\breve{\mathcal{J}}}\left\|\frac{\xi_i^{(k)}}{\breve{\psi}_i^{(k+1)}}-\bar{y}^{(k)}\right\|$$

$$+\frac{(\mu N)^2b^{(k)}}{2}\|\breve{\rho}\|^2+\frac{2(\mu N)^3B_y^2(r+2)}{r^5}a^{(k)}$$

$$+\frac{\sum_{i\in\breve{\mathcal{J}}}n_i^2}{2}\left(L_f+\frac{L_gB_y}{b^{(k)}r^2}\right)^2a^{(k)}$$

$$+\delta_f^{(k)}L_f+\delta_g^{(k)}L_g. \tag{B1}$$

Let $\breve{x}_*^{(k)}=[(\breve{x}_{1,*}^{(k)})^\top,(\breve{x}_{2,*}^{(k)})^\top,\dots,(\breve{x}_{\mu N,*}^{(k)})^\top]^\top$ be the optimal strategy in the enlarged system at iteration k. From Equation (3), we have

$$\mathbb{E}\left[H_k(\breve{x}^{(k)},\breve{\rho})-H_k\left(\breve{x}_*^{(k)},\bar{\rho}^{(k)}\right)-\frac{(\mu N)^2b^{(k)}}{2}\|\breve{\rho}\|^2\right]$$

$$\geq\mathbb{E}\left[\sum_{i\in\breve{\mathcal{J}}}f_{i,k}\left(\breve{x}_i^{(k)}\right)-\sum_{i\in\breve{\mathcal{J}}}f_{i,k}\left(\breve{x}_{i,*}^{(k)}\right)+\breve{\rho}^\top\sum_{i\in\breve{\mathcal{J}}}g_{i,k}\left(\breve{x}_i^{(k)}\right)\right.$$

$$- \frac{(\mu N)^2 b^{(k)}}{2} \|\breve{\rho}\|^2 \Bigg]$$

$$= \mathbb{E}\left[\sum_{i\in\mathcal{J}} f_{i,k}\left(\breve{x}_i^{(k)}\right) - \sum_{i\in\mathcal{J}} f_{i,k}\left(\breve{x}_{i,*}^{(k)}\right) + \breve{\rho}^\top \sum_{i\in\mathcal{J}} g_{i,k}\left(\breve{x}_i^{(k)}\right) - \frac{(\mu N)^2 b^{(k)}}{2} \|\breve{\rho}\|^2 \right], \tag{B2}$$

where the inequality follows from $\bar{\breve{\rho}}^{(k)} > \mathbf{0}$ and $\sum_{i\in\breve{\mathcal{J}}} g_{i,k}(\breve{x}_{i,*}^{(k)}) < \mathbf{0}$, and the last equality follows from $f_{i,k}(\breve{x}_i^{(k)}) = 0$ and $g_{i,k}(\breve{x}_i^{(k)}) = \mathbf{0}$ for any agent $i \in \mathcal{J}_{\mathrm{d}}$ and $k \in \mathcal{K}$.

Here, we define $\Xi(\breve{\rho})$ as follows:

$$\Xi(\breve{\rho}) = \breve{\rho}^\top \sum_{k\in\mathcal{K}} \sum_{i\in\mathcal{J}} g_{i,k}\left(\breve{x}_i^{(k)}\right) - \frac{(\mu N)^2 \|\breve{\rho}\|^2}{2} \sum_{k\in\mathcal{K}} b^{(k)}. \tag{B3}$$

From Equations (B1)–(B3), we obtain

$$\mathbb{E}\left[\sum_{k\in\mathcal{K}} \sum_{i\in\mathcal{J}} f_{i,k}\left(\breve{x}_i^{(k)}\right) - \sum_{k\in\mathcal{K}} \sum_{i\in\mathcal{J}} f_{i,k}\left(\breve{x}_{i,*}^{(k)}\right) + \Xi(\breve{\rho}) \right]$$

$$\leq Y_0 + Y_1 + Y_2(\breve{\rho}) + Y_3 + Y_4(\breve{\rho}) + Y_5 + Y_6 + Y_7, \tag{B4}$$

where

$$Y_0 = \sum_{k\in\mathcal{K}} \frac{1}{2a^{(k)}} \sum_{i\in\breve{\mathcal{J}}} \left(\left\|\breve{x}_i^{(k+1)} - \breve{x}_*^{(k+1)}\right\|^2 - \left\|\breve{x}_i^{(k+1)} - \breve{x}_{i,*}^{(k)}\right\|^2 \right),$$

$$Y_1 = \sum_{k\in\mathcal{K}} \frac{1}{2a^{(k)}} \sum_{i\in\breve{\mathcal{J}}} \left(\left\|\breve{x}_i^{(k)} - \breve{x}_{i,*}^{(k)}\right\|^2 - \left\|\breve{x}_i^{(k+1)} - \breve{x}_*^{(k+1)}\right\|^2 \right),$$

$$Y_2(\breve{\rho}) = \sum_{k\in\mathcal{K}} \frac{\mu N}{2a^{(k)}} (\|\bar{\breve{\rho}}^{(k)} - \breve{\rho}\|^2 - \|\bar{\breve{\rho}}^{(k+1)} - \breve{\rho}\|^2),$$

$$Y_3 = \left(3B_g + \frac{2\mu N B_y}{r^2}\right) \sum_{k\in\mathcal{K}} \sum_{i\in\breve{\mathcal{J}}} \left\| \frac{\breve{\xi}_i^{(k)}}{\breve{\psi}_i^{(k+1)}} - \bar{\breve{\rho}}^{(k)} \right\|,$$

$$Y_4(\breve{\rho}) = \sum_{k\in\mathcal{K}} \left(\|\breve{\rho}\| + \frac{B_y}{b^{(k)} r^2} \right) \sum_{i\in\breve{\mathcal{J}}} \left\| \frac{\breve{\zeta}_i^{(k)}}{\breve{\psi}_i^{(k+1)}} - \bar{\breve{y}}^{(k)} \right\|,$$

$$Y_5 = \left(L_f^2 \sum_{i\in\breve{\mathcal{J}}} n_i^2 + \frac{2(\mu N)^2 B_y^2 (r+2)}{r^5} \right) \sum_{k\in\mathcal{K}} a^{(k)},$$

$$Y_6 = \frac{B_y^2 L_g^2 \sum_{i\in\breve{\mathcal{J}}} n_i^2}{r^4} \sum_{k\in\mathcal{K}} \frac{a^{(k)}}{(b^{(k)})^2},$$

$$Y_7 = \sum_{k\in\mathcal{K}} \left(\delta_f^{(k)} L_f + \delta_g^{(k)} L_g \right).$$

For Y_0, we have

$$Y_0 = \sum_{k\in\mathcal{K}} \frac{1}{2a^{(k)}} \sum_{i\in\breve{\mathcal{J}}} \left(\breve{x}_{i,*}^{(k+1)} - \breve{x}_{i,*}^{(k)} \right)^\top \left(\breve{x}_{i,*}^{(k+1)} + \breve{x}_{i,*}^{(k)} - 2\breve{x}_i^{(k+1)} \right)$$

$$\leq 4B_x V_T. \tag{B5}$$

From the monotonically increasing property of the sequence $\{1/a^{(k)}\}$, we have

$$Y_1 = \frac{1}{2a^{(1)}} \sum_{i\in\breve{\mathcal{J}}} \left\|\breve{x}_i^{(1)} - \breve{x}_{i,1}^*\right\|^2 - \frac{1}{2a^{(T)}} \sum_{i\in\breve{\mathcal{J}}} \left\|\breve{x}_i(T+1) - \breve{x}_i^*(T+1)\right\|^2 + \frac{1}{2} \sum_{k=2}^{T} \left(\frac{1}{a^{(k)}} - \frac{1}{a^{(k-1)}} \right) \sum_{i\in\breve{\mathcal{J}}} \left\|\breve{x}_i^{(k)} - \breve{x}_{i,*}^{(k)}\right\|^2$$

$$\leq \frac{2\mu N B_x^2}{a^{(T)}}. \tag{B6}$$

For Y_2, we have

$$Y_2(\mathbf{0}) = \frac{\mu N}{2a^{(1)}} \|\bar{\breve{\rho}}^{(1)}\|^2 - \frac{\mu N}{2a^{(T)}} \|\bar{\breve{\rho}}^{(T+1)}\|^2 + \frac{\mu N}{2} \sum_{k=2}^{T} \left(\frac{1}{a^{(k)}} - \frac{1}{a^{(k-1)}} \right) \|\bar{\breve{\rho}}^{(k)}\|^2. \tag{B7}$$

From Lemmas 5 and 6, we have

$$\|\bar{\breve{\rho}}^{(k)}\| = \sum_{i\in\breve{\mathcal{J}}} \frac{\|\breve{\rho}_i^{(k)}\|}{\mu N} \leq \sum_{i\in\breve{\mathcal{J}}} \frac{\breve{\psi}_i^{(k+1)} B_y}{\mu N p r^2 b^{(k)}} \leq \frac{\mu N B_y}{p r^2 b^{(k)}}. \tag{B8}$$

From Equations (B7) and (B8), we obtain

$$Y_2(\mathbf{0}) \leq \frac{(\mu N)^3 B_y^2}{2p^2 r^4 a^{(T)} (b^{(T)})^2}. \tag{B9}$$

For Y_3 and Y_4, from Lemma 7, we have

$$Y_3 \leq \left(3B_g + \frac{2\mu N B_y}{r^2}\right) \frac{4C(\mu N)^2 \sqrt{m} B_y}{r^3(1-\sigma)} \sum_{k=0}^{T-1} a^{(k)}, \tag{B10}$$

$$Y_4(\mathbf{0}) \leq \frac{B_y}{r^2 b^{(T)}} \sum_{k\in\mathcal{K}} \sum_{i\in\breve{\mathcal{J}}} \left\| \frac{\breve{\zeta}_i^{(k)}}{\breve{\psi}_i^{(k+1)}} - \bar{\breve{y}}^{(k)} \right\|$$

$$\leq \frac{2C\mu N B_y}{r^3(1-\sigma) b^{(T)}} \left(\|\breve{y}^{(0)}\|_1 + \mu N \sqrt{mn} L_f C_g \sum_{k=0}^{T-1} a^{(k)} + \frac{\mu N \sqrt{mn} L_g C_g B_y}{r^2 b^{(T)}} \sum_{k=0}^{T-1} a^{(k)} \right). \tag{B11}$$

Y_6 is given as follows:

$$Y_6 = \frac{B_y^2 L_g^2 \sum_{i\in\breve{\mathcal{I}}} n_i^2}{r^4} \sum_{k\in\mathcal{K}} t^{2c-\frac{1}{2}} \leq \frac{B_y^2 L_g^2 \sum_{i\in\breve{\mathcal{I}}} n_i^2}{r^4} \int_1^{T+1} t^{2c-\frac{1}{2}} \mathrm{d}t$$

$$\leq \frac{2B_y^2 L_g^2 (T+1)^{2c+\frac{1}{2}} \sum_{i\in\breve{\mathcal{I}}} n_i^2}{r^4(1+4c)}. \tag{B12}$$

Since $\delta_f^{(k)} = \sum_{i\in\mathcal{J}} h_{f_i}/k$ and $\delta_g^{(k)} = \sum_{i\in\mathcal{J}} h_{g_i}/k$, Y_7 is given as follows:

$$\begin{aligned} Y_7 &= L_f\left(\sum_{i\in\breve{\mathcal{J}}} h_{f_i}\right)\sum_{k\in\mathcal{K}}\frac{1}{k} + L_g\left(\sum_{i\in\breve{\mathcal{J}}} h_{g_i}\right)\sum_{k\in\mathcal{K}}\frac{1}{k} \\ &\le \left(L_f\left(\sum_{i\in\breve{\mathcal{J}}} h_{f_i}\right) + L_g\left(\sum_{i\in\breve{\mathcal{J}}} h_{g_i}\right)\right)\int_1^{T+1}\frac{1}{t}\mathrm{d}t \\ &\le \left(L_f\left(\sum_{i\in\breve{\mathcal{J}}} h_{f_i}\right) + L_g\left(\sum_{i\in\breve{\mathcal{J}}} h_{g_i}\right)\right)\log(T+1). \end{aligned} \tag{B13}$$

Since $a^{(k)} = 1/\sqrt{k}$, we have

$$\sum_{k=0}^{T-1} a^{(k)} \le 2 + \int_1^{T+1} t^{-\frac{1}{2}}\mathrm{d}t = 2(T-1)^{\frac{1}{2}} = O(T^{\frac{1}{2}}). \tag{B14}$$

Then, from Equations (B4)–(B6) and (B9)–(B13), we obtain $\widetilde{\mathrm{Reg}}(T) = O(T^{2c+\frac{1}{2}}) + O(V_T)$.

Next, we show Equation (30). We note that $\Xi(\breve{\rho})$ is maximized at $\breve{\rho} = \breve{\rho}_0$, where

$$\breve{\rho}_0 = \frac{\left[\sum_{k\in\mathcal{K}}\sum_{i\in\mathcal{J}} g_{i,k}(x_i^{(k)})\right]_+}{(\mu N)^2\sum_{k\in\mathcal{K}} b^{(k)}}.$$

Then, Equation (B4) is given as follows:

$$\begin{aligned} &\mathbb{E}\left[f_{i,k}\left(\breve{x}_i^{(k)}\right) - f_{i,k}\left(\breve{x}_{i,*}^{(k)}\right) + \frac{\left\|\left[\sum_{k\in\mathcal{K}}\sum_{i\in\mathcal{J}} g_{i,k}\left(x_i^{(k)}\right)\right]_+\right\|^2}{2(\mu N)^2\sum_{k\in\mathcal{K}} b^{(k)}}\right] \\ &\le Y_0 + Y_1 + Y_2(\breve{\rho}) + Y_3 + Y_4(\breve{\rho}) + Y_5 + Y_6 + Y_7. \end{aligned} \tag{B15}$$

Since $c \in (0, 1/4)$, we have $\sum_{k\in\mathcal{K}} b^{(k)} \le 1 + \int_1^T \mathrm{d}t/t^c = (T^{1-c} - c)/(1-c) \le T^{1-c}/(1-c)$. Moreover, for $T \ge 3$, we have $\sum_{k\in\mathcal{K}} b^{(k)} \ge \int_1^T \mathrm{d}t/t^c = (T^{1-c}-1)/(1-c) \ge T^{1-c}/(2(1-c))$. Thus, for $T \ge 3$, we have

$$\|\breve{\rho}_0\| \le \frac{TB_g}{\mu N^2\sum_{k\in\mathcal{K}} b^{(k)}} \le \frac{2B_g(1-c)}{\mu N^2}T^c. \tag{B16}$$

Therefore, from Equations (B8) and (B16), and Lemma 7, for $T \ge 3$, we obtain

$$\begin{aligned} Y_2(\breve{\rho}) &= \sum_{k\in\mathcal{K}}\frac{1}{2a^{(k)}}(\|\bar{\breve{\rho}}^{(k)} - \breve{\rho}_0\|^2 - \|\bar{\breve{\rho}}^{(k+1)} - \breve{\rho}_0\|^2) \\ &= \frac{\mu N}{2a^{(1)}}\|\bar{\breve{\rho}}^{(1)} - \breve{\rho}_0\|^2 - \frac{\mu N}{2a^{(T)}}\|\bar{\breve{\rho}}^{(T)} - \breve{\rho}_0\|^2 \\ &\quad + \frac{\mu N}{2}\sum_{k=2}^{T}\left(\frac{1}{2a^{(k)}} - \frac{1}{2a^{(k-1)}}\right)\|\bar{\breve{\rho}}^{(k)} - \breve{\rho}_0\|^2 \\ &\le \frac{\mu N}{a^{(T)}}\left(\left(\frac{\mu N B_y}{pr^2 b^{(T)}}\right)^2 + \left(\frac{2B_g(1-c)}{\mu N^2}T^c\right)^2\right) \\ &= O(T^{2c+\frac{1}{2}}), \end{aligned} \tag{B17}$$

$$\begin{aligned} \mathbb{E}[Y_4(\breve{\rho}_0)] &= \mathbb{E}\left[\sum_{k\in\mathcal{K}}\left(\|\breve{\rho}_0\| + \frac{B_y}{b^{(k)}r^2}\right)\sum_{i\in\mathcal{J}}\left(\frac{\breve{\zeta}_i^{(k)}}{\breve{\psi}_i^{(k+1)}} - \bar{\breve{y}}^{(k)}\right)\right] \\ &\le \left(\|\breve{\rho}_0\| + \frac{B_y}{b^{(T)}r^2}\right)\mathbb{E}\left[\sum_{k\in\mathcal{K}}\sum_{i\in\mathcal{J}}\left\|\frac{\breve{\zeta}_i^{(k)}}{\breve{\psi}_i^{(k+1)}} - \bar{\breve{y}}^{(k)}\right\|\right] \\ &\le \frac{2C\mu N}{r(1-\sigma)}\left(\frac{2B_g(1-c)}{\mu N^2} + \frac{B_y}{r^2}\right)T^c \\ &\quad \times\left(\|\breve{y}^{(0)}\|_1 + \mu N\sqrt{mn}L_f C_g\sum_{k=0}^{T-1} a^{(k)} + \frac{\mu N\sqrt{mn}L_g C_g B_y}{r^2 b^{(T)}}\sum_{k=0}^{T-1} a^{(k)}\right) \\ &= O\left(T^{2c+\frac{1}{2}}\right). \end{aligned} \tag{B18}$$

From Equations (B5), (B6), (B10), (B12)–(B15), (B17), and (B18), we have

$$\begin{aligned} &\mathbb{E}\left[f_{i,k}\left(\breve{x}_i^{(k)}\right) - f_{i,k}\left(\breve{x}_{i,*}^{(k)}\right) + \frac{\left\|\left[\sum_{k\in\mathcal{K}}\sum_{i\in\mathcal{J}} g_{i,k}\left(x_i^{(k)}\right)\right]_+\right\|^2}{2(\mu N)^2\sum_{k\in\mathcal{K}} b^{(k)}}\right] \\ &= O(T^{2c+\frac{1}{2}}) + O(V_T). \end{aligned}$$

From the Lipschitz continuity of $f_{i,k}$, we have

$$\begin{aligned} &\sum_{k\in\mathcal{K}}\sum_{i\in\mathcal{J}} f_{i,k}\left(\breve{x}_i^{(k)}\right) - \sum_{k\in\mathcal{K}}\sum_{i\in\mathcal{J}} f_{i,k}\left(\breve{x}_{i,*}^{(k)}\right) \\ &= -\sum_{k\in\mathcal{K}}\sum_{i\in\mathcal{J}}\left(f_{i,k}\left(\breve{x}_{i,*}^{(k)}\right) - f_{i,k}\left(\breve{x}_i^{(k)}\right)\right) \\ &\ge -\sum_{k\in\mathcal{K}}\sum_{i\in\mathcal{J}} L_f\left\|\breve{x}_{i,*}^{(k)} - \breve{x}_i^{(k)}\right\| \\ &\ge -2NTL_f B_x. \end{aligned}$$

It follows that

$$\left(\mathbb{E}\left[\left\|\left[\sum_{k\in\mathcal{K}}\sum_{i\in\mathcal{I}} g_{i,k}\left(x_i^{(k)}\right)\right]_+\right\|\right]\right)^2$$

$$\leq \mathbb{E}\left[\left\|\left[\sum_{k\in\mathcal{K}}\sum_{i\in\mathcal{I}} g_{i,k}\left(x_i^{(k)}\right)\right]_+\right\|^2\right]$$

$$\leq O\left(T^{2c+\frac{1}{2}}\right)\sum_{k\in\mathcal{K}} b^{(k)} + O(V_T)\sum_{k\in\mathcal{K}} b^{(k)} + 4\mu^2 N^3 L_f B_x T \sum_{k\in\mathcal{K}} b^{(k)}$$

$$\leq \frac{T^{1-c}}{1-c} O\left(T^{2c+\frac{1}{2}}\right) + \frac{T^{1-c}}{1-c} O(T) + \frac{T^{1-c}}{1-c} O(V_T)$$

$$\leq O(T^{2-c}) + O(V_T T^{1-c}).$$

Therefore, we obtain $\widetilde{\mathrm{Reg}}^c = O(T^{1-\frac{c}{2}}) + O(V_T^{\frac{1}{2}} T^{\frac{1-c}{2}})$. □