Distributed Gradient Tracking for Unbalanced Optimization With Different Constraint Sets

Songsong Cheng, *Member, IEEE*, Shu Liang, Yuan Fan, and Yiguang Hong, *Fellow, IEEE*

***Abstract*—Gradient tracking methods have become popular for distributed optimization in recent years, partially because they achieve linear convergence using only a constant step-size for strongly convex optimization. In this article, we construct a counterexample on constrained optimization to show that direct extension of gradient tracking by using projections cannot guarantee the correctness. Then, we propose projected gradient tracking algorithms with diminishing step-sizes rather than a constant one for distributed strongly convex optimization with different constraint sets and unbalanced graphs. Our basic algorithm can achieve $O(\ln T/T)$ convergence rate. Moreover, we design an epoch iteration scheme and improve the convergence rate as $O(1/T)$.**

***Index Terms*—Convergence, different constraint sets, distributed optimization, gradient tracking, unbalanced graphs.**

I. Introduction

Distributed optimization has gained much research attention due to its applications in large-scale multiagent systems, such as power grids, sensor networks, and unmanned systems. Various distributed optimization algorithms have been developed based on local gradients or subgradients, including distributed subgradient method, distributed mirror descent methods, and distributed primal-dual methods [1]–[4].

The distributed gradient tracking method, proposed in [5], is based on the gradient descent scheme and replaces the centralized gradient information by a distributed tracking one. The mechanism of this method can be understood as the interaction of distributed averaging the local decision variables and distributed tracking the aggregation of local gradients. The distributed gradient tracking algorithm uses only a constant step-size and achieves linear convergence for strongly convex optimization, even in the presence of time-varying network graphs. With such promising performances, it receives rapidly further investigation, such as gradient tracking with quantized communication [6], gradient tracking over unbalanced graphs [5], [7]–[10], and gradient tracking for stochastic optimization [11]–[13]. However, none of the existing gradient tracking with a constant step-size can be extended directly with projections for constrained optimization. The reasons are as follows. First, a direct extension of gradient tracking with any constant step-sizes cannot even guarantee the correctness of equilibrium points. Second, the projection operator also brings nonlinearty to the algorithm, which makes the average consensus part difficult to analyze.

In practice, agents with different locations or environments have different local set constraints. However, most works on constrained consensus optimization problems consider identical local sets and cannot be applied to different constraints. There are a few works considering such problems with some restrictive assumptions. For example, the algorithm given in [14] requires a complete graph (i.e., agents can communicate with all others). The method in [15] introduces a primal-dual algorithm and requires undirected graphs. A recent work [16] has also focused on such problems by some averaging technique, while the algorithm requires to share local subgradients and local decision variables simultaneously.

In this article, we consider distributed optimization problems over unbalanced graphs and different constraint sets. We develop projected gradient tracking methods to handle constrained optimization with the following contributions.

1) We reveal a fact that the projected gradient tracking with a constant step-size violates the consensus and optimal condition at the optimal solution, for a distributed optimization problem with different constraint sets.
2) We design a projected gradient tracking with diminishing step-sizes to achieve an accurate convergence. Our method overcomes the difficulties caused by the nonlinearity of the projected dynamics and the unbalanced averaging tracking over different local constraint sets and unbalanced graphs.
3) We improve the convergence rate of the projected gradient tracking from $O(\ln T/T)$ to $O(1/T)$ by designing an epoch iteration scheme.

The rest of this article is organized as follows. Section II presents related preliminaries and formulates the distributed optimization problem. Second III provides a projected gradient tracking method and an improved version. Section IV shows an illustrative example. Finally, Section V concludes this article.

II. Preliminaries and Problem Formulation

A. Notations

The positive integer number set, real number set, n-dimensional real column vector set, and $n\times m$ real matrix set are denoted as $\mathbb{N}_+$, $\mathbb{R}$, $\mathbb{R}^n$, and $\mathbb{R}^{n\times m}$, respectively. $\mathbf{1}_n \in \mathbb{R}^n$ is a vector with each component being one, and the subscript is omitted when the dimension is clear. $I_n \in \mathbb{R}^{n\times n}$ is an identity matrix. $I_m^n = \mathbf{1}_n^\top \otimes I_m \in \mathbb{R}^{m\times nm}$, where $\otimes$ is the Kronecker product operator. $A = [a_{ij}] \in \mathbb{R}^{n\times m}$ denotes an $n\times m$ matrix, where a_{ij} is the element in the ith row and jth column of A. $A^\top$ means the transpose of A. $\mathbf{x} = \mathrm{col}\{\mathbf{x}_i\}_{i=1}^n = [\mathbf{x}_1^\top, \mathbf{x}_2^\top, \dots, \mathbf{x}_n^\top]^\top$.

For a row (column) stochastic matrix $[\phi_{ij}]$ ($[\psi_{ij}]$), there exists a stochastic vector $\boldsymbol{\pi}^\top = [\pi_1, \pi_2, \dots, \pi_n]$ ($\mathbf{u}^\top = [\mathrm{u}_1, \mathrm{u}_2, \dots, \mathrm{u}_n]$), such that $\boldsymbol{\pi}^\top[\phi_{ij}] = \boldsymbol{\pi}^\top$ ($[\psi_{ij}]\mathbf{u} = \mathbf{u}$). For $\boldsymbol{\pi}$ and $\mathbf{u}$, define $\bar{\boldsymbol{\pi}} = \boldsymbol{\pi} \otimes I_m$, $\hat{\boldsymbol{\pi}} = (\mathbf{1}_n\boldsymbol{\pi}^\top) \otimes I_m$, $\bar{\mathbf{u}} = \mathbf{u} \otimes I_m$, and $\hat{\mathbf{u}} = (\mathbf{u}\mathbf{1}_{n^\top}) \otimes I_m$. For $\mathbf{x} = \mathrm{col}\{\mathbf{x}_i\}_{i=1}^n$, define a weighted average vector as $\bar{\mathbf{x}} = \bar{\boldsymbol{\pi}}^\top\mathbf{x}$ and define an augmented vector as $\hat{\mathbf{x}} = \hat{\boldsymbol{\pi}}^\top\mathbf{x} = \mathbf{1}_n \otimes \bar{\mathbf{x}}$. Denote x^t as the tth power of x.

Manuscript received 5 April 2022; accepted 10 July 2022. Date of publication 19 July 2022; date of current version 29 May 2023. This work was supported in part by the National Natural Science Foundation of China under Grant 62103003, Grant 72171171, Grant 61903027, Grant 61973002, and Grant 61733018, in part by the Shanghai Municipal Science and Technology Major Project under Grant 2021SHZDZX0100, and in part by the Anhui Provincial Natural Science Foundation under Grant 2008085J32. Recommended by Associate Editor Zhen Kan. *(Corresponding author: Shu Liang.)*

Songsong Cheng and Yuan Fan are with the School of Electrical Engineering and Automation, Anhui University, Hefei 230601, China (e-mail: sscheng@amss.ac.cn; yuanf@ahu.edu.cn).

Shu Liang and Yiguang Hong are with the Department of Control Science and Engineering, Tongji University, Shanghai 200092, China (e-mail: sliangresearch@163.com; yghong@iss.ac.cn).

Color versions of one or more figures in this article are available at https://doi.org/10.1109/TAC.2022.3192316.

Digital Object Identifier 10.1109/TAC.2022.3192316

A set $\mathcal{X}_{\text{o}} \subseteq \mathbb{R}^m$ is convex if $\theta \mathbf{x}_1 + (1-\theta)\mathbf{x}_2 \in \mathcal{X}_{\text{o}}$ for any $\mathbf{x}_1, \mathbf{x}_2 \in \mathcal{X}_{\text{o}}$ and $\theta \in (0,1)$. For a compact and convex set $\mathcal{X}_{\text{o}}$, define a projection operator $P_{\mathcal{X}_{\text{o}}} : \mathbb{R}^m \to \mathcal{X}_{\text{o}}$ as

$$P_{\mathcal{X}_{\text{o}}}(\mathbf{y}) = \underset{\mathbf{x}_a \in \mathcal{X}_{\text{o}}}{\operatorname{argmin}} \|\mathbf{y} - \mathbf{x}_a\|.$$

Clearly, $\|P_{\mathcal{X}_{\text{o}}}(\mathbf{x}_a) - P_{\mathcal{X}_{\text{o}}}(\mathbf{x}_b)\| \le \|\mathbf{x}_a - \mathbf{x}_b\| \ \forall \mathbf{x}_a, \mathbf{x}_b \in \mathbb{R}^m$.

B. Graph Theory

A directed graph of a multiagent system is denoted by $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, where $\mathcal{V} = \{1, 2, \ldots, n\}$ and $\mathcal{E} \subseteq \mathcal{V} \times \mathcal{V}$ are the sets of nodes and edges. $(i,j) \in \mathcal{E}$ means that node i sends information to node j. $\mathcal{N}_{\text{i}}^i = \{j \in \mathcal{V} | (j,i) \in \mathcal{E}\}$ and $\mathcal{N}_{\text{o}}^i = \{j \in \mathcal{V} | (i,j) \in \mathcal{E}\}$ are the in-neighbor and out-neighbor sets of node i, respectively. The corresponding in-degree and out-degree of node i are defined as $d_{\text{i}}^i = |\mathcal{N}_{\text{i}}^i|$ and $d_{\text{o}}^i = |\mathcal{N}_{\text{o}}^i|$, respectively.

We make the following mild assumption on graph $\mathcal{G}$.

Assumption 1: The directed graph $\mathcal{G}$ is strongly connected.

Given graph $\mathcal{G}$, a row stochastic matrix $[\phi_{ij}] \in \mathbb{R}^{n \times n}$ and column stochastic matrix $[\psi_{ij}] \in \mathbb{R}^{n \times n}$ can be defined by

$$\phi_{ij} = \begin{cases} \frac{1}{d_{\text{i}}^i}, & j \in \mathcal{N}_{\text{i}}^i, \\ 0, & \text{otherwise}, \end{cases} \qquad \psi_{ji} = \begin{cases} \frac{1}{d_{\text{o}}^i}, & j \in \mathcal{N}_{\text{o}}^i, \\ 0, & \text{otherwise}. \end{cases}$$

Let $\Phi = [\phi_{ij}] \otimes I_m \in \mathbb{R}^{nm \times nm}$ and $\Psi = [\psi_{ij}] \otimes I_m \in \mathbb{R}^{nm \times nm}$.

Lemma 1 ([17]): Under Assumption 1,

$$\|\Phi^t - \hat{\boldsymbol{\pi}}\| \le \Gamma \gamma^t, \quad \text{and} \quad \|\Psi^t - \hat{\mathbf{u}}\| \le \Gamma \gamma^t$$

for two constants $\Gamma \in \mathbb{R}_+$ and $\gamma \in (0,1)$.

C. Problem Formulation

Consider the following distributed optimization problem:

$$\begin{aligned} \min_{\mathbf{x}} \quad & f(\mathbf{x}) := \sum_{i=1}^n f_i(\mathbf{x}_i) \\ \text{s.t.} \quad & \mathbf{x}_i = \mathbf{x}_j \in \mathcal{X}_{\text{o}} = \textstyle\bigcap_{k=1}^n \mathcal{X}_k \quad \forall i,j \in \mathcal{V} \end{aligned} \tag{1}$$

where the ith agent has local objective function $f_i(\mathbf{x}_i)$, local decision variable $\mathbf{x}_i$, and local constraint set $\mathcal{X}_i$. Here,

$$\mathbf{x} = \text{col}\{\mathbf{x}_i\}_{i=1}^n \in \mathcal{X} = \Pi_{i=1}^n \mathcal{X}_i \subseteq \mathbb{R}^{nm}.$$

For (1), the following mild assumptions are adopted.

Assumption 2: The set $\mathcal{X}$ is compact, convex, and bounded as $\|\mathbf{x}\| \le B_{\mathbf{x}} \ \forall \mathbf{x} \in \mathcal{X}$. Also, $\mathcal{X}_o$ has a nonempty interior.

Assumption 3: Each gradient $\nabla f_i(\mathbf{x}_i)$ is G_{fi}-Lipschitz continuous for some constant $G_{fi} > 0$, i.e., for any $\mathbf{x}_a$, $\mathbf{x}_b \in \mathcal{X}_i$,

$$\|\nabla f_i(\mathbf{x}_a) - \nabla f_i(\mathbf{x}_b)\| \le G_{fi} \|\mathbf{x}_a - \mathbf{x}_b\|.$$

Assumption 4: Each $f_i(\mathbf{x}_i)$ is σ_i-strong convex, namely, for any $\mathbf{x}_a$, $\mathbf{x}_b \in \mathcal{X}_i$,

$$f_i(\mathbf{x}_a) \ge f_i(\mathbf{x}_b) + \langle \nabla f_i(\mathbf{x}_b), \mathbf{x}_a - \mathbf{x}_b \rangle + \tfrac{\sigma_i}{2} \|\mathbf{x}_a - \mathbf{x}_b\|^2.$$

If $\mathcal{X} = \mathbb{R}^{nm}$ and graph $\mathcal{G}$ is balanced, (1) can be solved by the following gradient tracking method [5]:

$$\begin{cases} \mathbf{x}_i^{t+1} = \sum_{j=1}^n w_{ij} \mathbf{x}_j(t) + \alpha \mathbf{g}_i(t) \\ \mathbf{g}_i(t+1) = \sum_{j=1}^n w_{ij} \mathbf{g}_j(t) - \nabla f_i(\mathbf{x}_i(t+1)) + \nabla f_i(\mathbf{x}_i(t)) \end{cases} \tag{2}$$

where $W = [w_{ij}]$ is a doubly stochastic matrix.

III. Algorithm Design and Convergence

In this section, we show the necessity of diminishing step-sizes for the projected gradient tracking, and then provide a valid algorithm design for unbalanced graphs and different constraint sets. Moreover, we improve our gradient tracking method with an epoch scheme to accelerate the convergence.

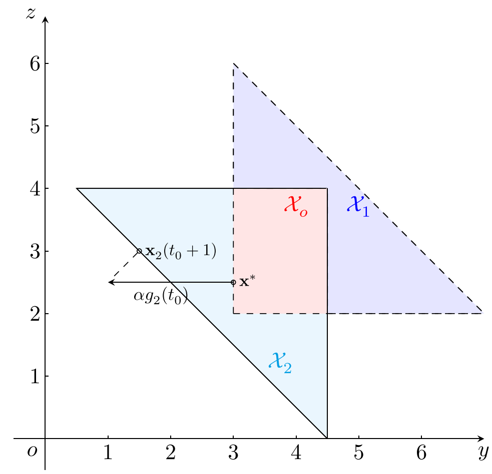

Fig. 1. Inaccuracy convergence of projected gradient tracking with a constant step-size.

A. Counterexample

Consider a two-agent system to solve the problem (1) with

$$f_i(\mathbf{x}_i) = \tfrac{1}{2}\|\mathbf{x}_i - \mathbf{a}_i\|^2, i = 1, 2 \tag{3}$$

where $\mathbf{a}_1 = [1, 2.5]^\top$, $\mathbf{a}_2 = [2, 2.5]^\top$, $\mathcal{X}_1 = \{[y,z]^\top | z \le 9 - y, y \ge 3, z \ge 2\}$ (the light blue region surrounded by dashed line in Fig. 1), $\mathcal{X}_2 = \{[y,z]^\top | z \ge 4.5 - y, y \le 4.5, z \le 4\}$ (the cyan region surrounded by solid line in Fig. 1), and $\mathcal{X}_o = \mathcal{X}_1 \cap \mathcal{X}_2 = \{[y,z]^\top | 3 \le y \le 4.5, 2 \le z \le 4\}$ (the red region in Fig. 1). Apparently, $\mathbf{x}^* = [3, 2.5]^\top$.

To solve (1) with objective functions in (3), we test a direct extension of gradient tracking with projections as

$$\begin{cases} \mathbf{x}_i(t+1) = P_{\mathcal{X}_i}\{\sum_{j=1}^n w_{ij}\mathbf{x}_j(t) + \alpha \mathbf{g}_i(t)\} \\ \mathbf{g}_i(t+1) = \sum_{j=1}^n w_{ij}\mathbf{g}_j(t) - \nabla f_i(\mathbf{x}_i(t+1)) + \nabla f_i(\mathbf{x}_i(t)) \end{cases} \tag{4}$$

where $W = [w_{ij}] = \frac{1}{2}\mathbf{1}_2\mathbf{1}_2^\top$. To show the incorrectness of (4), it suffices to verify that the optimal solution $\mathbf{x}^*$ is not a fixed-point of (4). Suppose that (4) achieves the optimal solution $\mathbf{x}^* = [3, 2.5]^\top$ at the t_0th iteration, namely $\forall i = 1, 2$,

$$\mathbf{x}_i(t_0) = [3, 2.5]^\top, ~~\mathbf{g}_i(t_0) = -\sum_{j=1}^2 \nabla f_j(\mathbf{x}^*) = [-3, 0]^\top.$$

Then, for $\mathbf{x}_1$,

$$\begin{aligned} \mathbf{x}_1(t_0+1 &= P_{\mathcal{X}_1}\left\{\frac{1}{2}[\mathbf{x}_1(t_0) + \mathbf{x}_2(t_0)] + \alpha \mathbf{g}_1(t_0)\right\} \\ &= P_{\mathcal{X}_1}\left\{[3, 2.5]^\top - \alpha[3, 0]^\top\right\} \\ &= \mathbf{x}^* \quad \forall \alpha > 0. \end{aligned}$$

However, for $\mathbf{x}_2$,

$$\begin{aligned} \mathbf{x}_2(t_0+1) &= P_{\mathcal{X}_2}\left\{\frac{1}{2}[\mathbf{x}_1(t_0) + \mathbf{x}_2(t_0)] + \alpha \mathbf{g}_2(t_0)\right\} \\ &= P_{\mathcal{X}_2}\left\{[3, 2.5]^\top - \alpha[3, 0]^\top\right\}. \end{aligned}$$

If $0 < \alpha \le \frac{1}{3}$, then $[3, 2.5]^\top - \alpha[3, 0]^\top \in \mathcal{X}_2$, which implies $\mathbf{x}_2(t_0+1) = [3, 2.5]^\top - \alpha[3, 0]^\top \ne \mathbf{x}^*$ since $\alpha > 0$. If $\alpha \ge \frac{1}{3}$, then $[3, 2.5]^\top - \alpha[3, 0]^\top \notin \mathcal{X}_2$, which implies that $\mathbf{x}_2(t_0+1)$ is projected on the lower left boundary of $\mathcal{X}_2$ rather than $\mathbf{x}^*$. As a result, even the optimal solution is achieved at some certain iteration step, the projected dynamics (4) will violate the consensus at the next step.

Therefore, the gradient tracking with a constant step-size cannot be extended directly to the projected version for the constrained distributed

Algorithm 1: Projected Gradient Tracking.

Input: $\mathbf{x}_i(1)\in\mathcal{X}_i$, $\mathbf{g}_i(1)=-\nabla f_i(\mathbf{x}_i(1))$, $\forall i\in\mathcal{V}$, $T\in\mathbb{N}_+$;

for $t=1,2,\cdots,T-1$ **do**

$\quad \mathbf{x}_i(t+1)=P_{\mathcal{X}_i}\left\{\sum_{j=1}^{n}\phi_{ij}\mathbf{x}_j(t)+\alpha(t)\mathbf{g}_i(t)\right\}$,

$\quad \mathbf{g}_i(t+1)=\sum_{j=1}^{n}\psi_{ij}\mathbf{g}_j(t)-\nabla f_i(\mathbf{x}_i(t+1))+\nabla f_i(\mathbf{x}_i(t))$;

end

return $\mathbf{x}_{i,\mathrm{av}}(T)=\frac{1}{T}\sum_{t=1}^{T}\mathbf{x}_i(t)$, $\forall i\in\mathcal{V}$;

optimization problem. One remedy is to design projected gradient tracking with diminishing step-sizes so that it is possible to ensure correctness and convergence.

Remark 1: For a distributed optimization problem with different constraint sets, the work in [16] showed that if local decision variables are initialized at the optimal value of the corresponding local objective functions, the dual averaging method cannot achieve the global optimal value. Moreover, under a different local regularization terms setting, the work in [18] demonstrated that proximal gradient with constant step-sizes also cannot converge to the global optimal value. Thus, it remains challenging to solve distributed constrained optimization with different local constraint sets.

B. Distributed Gradient Tracking With Projection

For problem (1), we propose Algorithm 1 by gradient tracking. Its convergence result is presented as follows.

Theorem 1: Under Assumptions 1–4, let $(\mathbf{x}_i(t),\mathbf{g}_i(t))$ be generated by Algorithm 1. If $\alpha(t)=\frac{1}{\sigma n\boldsymbol{\pi}^\top\mathbf{u}t}$, then

$$\|\mathbf{x}_{\mathrm{av}}(T)-\mathbf{1}_n\otimes\mathbf{x}^*\|^2\leq\frac{C_1+C_2\ln(T+1)}{T}+\frac{C_3^2\ln^2(T+1)}{T^2}=O\left(\frac{\ln T}{T}\right)$$

where $\mathbf{x}_{\mathrm{av}}(T)=\mathrm{col}\{\mathbf{x}_{i,\mathrm{av}}(T)\}_{i=1}^n$ and

$$C_0=\frac{C_4}{1-\gamma}+\frac{4B_{\mathbf{x}}G_f\Gamma^2}{\gamma(1-\gamma)^2\boldsymbol{\pi}^\top\mathbf{u}},\ C_1=\frac{4C_4B_{\mathbf{x}}}{\sigma(1-\gamma)}+\frac{16B_{\mathbf{x}}^2G_f\Gamma^2}{\sigma\boldsymbol{\pi}^\top\mathbf{u}(1-\gamma)^2}$$

$$C_2=\frac{4}{n\sigma^2\boldsymbol{\pi}^\top\mathbf{u}}\left[\frac{L_g^2}{\boldsymbol{\pi}^\top\mathbf{u}}+2C_0\left(L_g+4C_0\boldsymbol{\pi}^\top\mathbf{u}\right)\right]$$

$$C_3=\frac{2\Gamma\left(2n^2B_{\mathbf{x}}+d_o\right)}{(1-\gamma)d_o}\left[B_{\mathbf{x}}+\frac{2L_g}{n\sigma\boldsymbol{\pi}^\top\mathbf{u}}+\frac{8C_0}{n\sigma}+\frac{\sigma\left(C_1+C_2\right)}{4C_0}\right]$$

with $C_4=\Gamma\left[2nB_{\mathbf{x}}\left(n\sigma+G_f\left(1+\frac{2nB_{\mathbf{x}}}{d_o}\right)\right)+\frac{(n+2)L_g}{\boldsymbol{\pi}^\top\mathbf{u}}\right]$, $G_f=\max\{G_{f1},\ldots,G_{fn}\}$, and $\sigma=\min\{\sigma_1,\ldots,\sigma_n\}$.

The design of Algorithm 1 is inspired by the surplus-variable-based method in [19], push-DIGing method in [5], Aug-DGM method in [20], and EXTRA method in [21]. As pointed out in [19], the row stochastic matrix brings the consensus of the decision variables and the column stochastic matrix brings an "optimal average value," In [5], [20], and [21], the authors have introduced additional dynamics to learn the gradients of global objective functions. Here, we formulate a row stochastic matrix $[\phi_{ij}]$ for the local decision variables $\mathbf{x}_i$ to achieve the consensus and formulate a column stochastic matrix $[\psi_{ij}]$ for the gradient tracking variables $\mathbf{g}_i$ to learn the gradient of the global objective function.

Remark 2: The proposed projected gradient tracking can handle unbalanced graphs, which is more general than balanced graphs in [14] and [22]–[24]. Moreover, our projected gradient tracking has just two dynamics based on in-degree and out-degree information, which is simpler than those gradient-push methods with four dynamics [7], [25]. Furthermore, in comparison with the unconstrained distributed optimization in [5]–[8] and [11], we consider the problem with different constraint sets and overcome the inexact averaging caused by the update flow with different local projections.

Algorithm 2: Projected Gradient Tracking with Epoch.

Input: $\mathbf{x}_i(1,1)\in\mathcal{X}_i$, $\mathbf{g}_i(1,1)=-\nabla f_i(\mathbf{x}_i(1,1))$, $\forall i\in\mathcal{V}$, $k=1$, $\theta_1\in\mathbb{N}_+\backslash\{1\}$, $\theta_2=\frac{1}{\theta_1}$, and $\bar{k}$, $T(1)\in\mathbb{N}_+$;

while $k\leq\bar{k}$ **do**

$\quad$ **for** $\tau=1,2,\cdots,T(k)-1$ **do**

$\quad\quad \mathbf{x}_i(k,\tau+1)=P_{\mathcal{X}_i}\left\{\sum_{j=1}^{n}\phi_{ij}\mathbf{x}_j(k,\tau)+\alpha(k)\mathbf{g}_i(k,\tau)\right\}$,

$\quad\quad \mathbf{g}_i(k,\tau+1)=\sum_{j=1}^{n}\psi_{ij}\mathbf{g}_j(k,\tau)-\nabla f_i(\mathbf{x}_i(k,\tau+1))+\nabla f_i(\mathbf{x}_i(k,\tau))$;

$\quad$ **end**

$\quad \mathbf{x}_i(k+1,1)=\frac{1}{T(k)}\sum_{\tau=1}^{T(k)}\mathbf{x}_i(k,\tau)$,

$\quad \mathbf{g}_i(k+1,1)=-\nabla f_i(\mathbf{x}_i(k+1,1))$,

$\quad T(k+1)=\theta_1T(k)$, $\alpha(k+1)=\theta_2\alpha(k)$, $k=k+1$;

end

return $\mathbf{x}_i(\bar{k}+1,1)$, $\forall i\in\mathcal{V}$;

C. Projected Gradient Tracking With Epoch Scheme

In view of Theorem 1, we are intuitively wondering whether the convergence rate can be improved from $O(\ln T/T)$ to $O(1/T)$. To this end, we propose Algorithm 2.

Remark 3: In Algorithm 2, the total number of iteration $T=\sum_{k=1}^{\bar{k}}T(k)=\frac{(\theta_1^{\bar{k}}-1)T(1)}{\theta_1-1}$ is divided into $\bar{k}$ epochs. At the kth epoch, the inner loop of Algorithm 2 can be viewed as an algorithm with a fixed step-size $\alpha(k)$. For the time indexes t and (k,τ), we have $t=\sum_{\kappa=1}^{k-1}T(\kappa)+\tau$ for $k\geq2$ or $t=\tau$ for $k=1$.

Then, we provide the convergence result of Algorithm 2.

Theorem 2: Under Assumptions 1–4, let $\mathbf{x}_i(k,\tau)$ and $\mathbf{g}_i(k,\tau)$ be generated by Algorithm 2. If $\alpha(k+1)=\frac{1}{2}\alpha(k)$, $T(k+1)=2T(k)$, $\alpha(1)=\frac{1}{\boldsymbol{\pi}^\top\mathbf{u}\sigma T_0}$, and $T(1)=4T_0$, then

$$\|\mathbf{x}\left(\bar{k}+1,1\right)-\mathbf{1}_n\otimes\mathbf{x}^*\|^2\leq\frac{2C_5}{T^2}+\frac{1}{T}\left(8C_6+\frac{2C_7}{\boldsymbol{\pi}^\top\mathbf{u}\sigma}\right)=O\left(\frac{1}{T}\right)$$

where C_5, C_6, and C_7 are defined hereinafter.

Remark 4: Theorems 1 and 2 also reveal a fact that a stronger connectivity of graph (smaller γ), a smaller range of the feasible sets (smaller $B_{\mathbf{x}}$), a larger step-size, and smaller iteration rounds in each epoch (a smaller T_0 in Theorem 2) yield a faster convergence rate.

IV. Numerical Simulation

In this section, we consider the following distributed optimization problem to demonstrate the effectiveness of the proposed algorithms.

$$\begin{cases}\min\limits_{\mathbf{x}_i\in\mathcal{X}_i} & \sum_{i=1}^{10}\frac{1}{2}\left\|A_i\mathbf{x}_i-\mathbf{b}_i\right\|^2\\ \text{s.t.} & \mathbf{x}_i=\mathbf{x}_j\quad\forall i,j\in\mathcal{V}\end{cases}\tag{5}$$

where $A_i\in\mathbb{R}^{2\times6}$, $\mathbf{b}_i\in\mathbb{R}^2$, and $\mathcal{X}_i=\{[x_{i1},\ x_{i,2},\ \ldots,\ x_{i6}]^\top\,|\,\underline{x}_{ij}\ \leq x_{ij}\ \leq\bar{x}_{ij}\quad\forall i\in\mathcal{V},\quad j=1,2,\ldots,6\}$ with $\underline{x}_{ij}=x_j^\star+10+\sin(ij\pi/120)$, $\bar{x}_{ij}=x_j^\star+20+10\sin(ij\pi/120)$, and $\mathbf{x}^\star=\mathrm{col}\{x_i^\star\}_{i=1}^6:=\left(\sum_{i=1}^{10}A_i^\top A_i\right)^{-1}\left(\sum_{i=1}^{10}A_i^\top\mathbf{b}_i\right)$.

The graph of the multiagent system is shown in Fig. 2 and the ith agent has access to $(A_i,\mathbf{b}_i)$ whose entries are randomly chosen by using

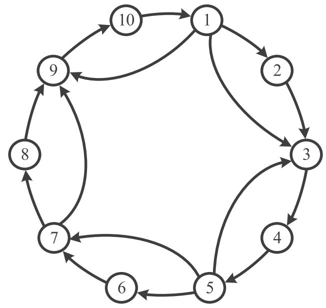

Fig. 2. Unbalanced graph of numerical simulation.

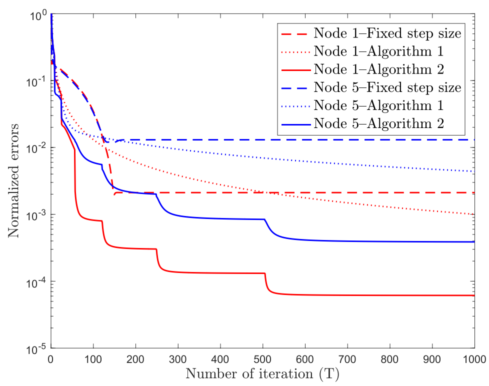

Fig. 3. Results of numerical simulation.

a uniform distribution with supports [0,0.1] and $[0,5]$, respectively. We utilize the following three methods to solve this optimization problem:

1) the method in (4) with a fixed step-size $\alpha = 0.1$;
2) Algorithm 1 with $\alpha(t) = \frac{10}{t}$;
3) Algorithm 2 with $\alpha(k) = \frac{8}{T_0 2^k}$ and $T_0 = 4$.

We show normalized errors $\|\mathbf{x}_{i,\mathrm{av}}(t) - \mathbf{x}^*\|/\|\mathbf{x}_i(1) - \mathbf{x}^*\|$ of agents 1 and 5 in Fig. 3, where the dashed, dotted, and solid lines are the errors of the method in (4) with a fixed step-size, Algorithms 1 and 2, respectively. From Fig. 3, compared with the inaccurate solution of the gradient tracking with a fixed step-size, the proposed methods achieve an accurate solution. Moreover, the proposed projected gradient tracking with the epoch scheme can accelerate the convergence.

V. Conclusion

In this article, we proposed a projected gradient tracking method for distributed optimization problem over unbalanced graphs with different constraint sets. First, we provided a counterexample to illustrate the necessity of diminishing step-size design for the projected gradient tracking. Then, we analyzed the convergence performance of the projected gradient tracking with diminishing step-sizes and obtained $O(\ln T/T)$ convergence rate. Moreover, we improved the convergence rate as $O(1/T)$ by designing an epoch scheme. Finally, we showed the effectiveness of the proposed algorithms with an example.

Appendix

Here, we mainly prove Theorems 1 and 2. The sketch is as follows. We first define two virtual variables $\bar{\mathbf{v}}(t)$ and $\bar{\mathbf{v}}(k,\tau) \in \mathcal{X}_o$ and prove that the corresponding averaging $\bar{\mathbf{v}}_{\mathrm{av}}(T) = \frac{1}{T}\sum_{t=1}^{T}\bar{\mathbf{v}}(t)$ and $\bar{\mathbf{v}}(\bar{k}+1,1) = \frac{1}{T(\bar{k})}\sum_{\tau=1}^{T(\bar{k})}\bar{\mathbf{v}}(k,\tau) \in \mathcal{X}_o$ converge to the optimal solution $\mathbf{x}^*$. Moreover, we show that $\mathbf{x}_{i,\mathrm{av}}(T)$ and $\mathbf{x}_i(\bar{k}+1,1)$ reach consensus and track $\bar{\mathbf{v}}_{\mathrm{av}}(T)$ and $\bar{\mathbf{v}}(\bar{k}+1,1)$, respectively.

A. Technical Lemmas

Lemma 2: Let Assumptions 1–3 hold. If $(\mathbf{x}_i(t), \mathbf{g}_i(t))$ is generated by Algorithm 1, then

$$\begin{cases} \|\mathbf{g}(t)\| \le 2nG_f B_{\mathbf{x}} \frac{2\Gamma+1-\gamma}{1-\gamma} =: L_g & \text{(6a)} \\ \|\mathbf{x}(t) - \hat{\mathbf{x}}(t)\| \le \Gamma B_{\mathbf{x}}\gamma^{t-1} + \Gamma\sum_{\tau=1}^{t-1}\gamma^{\tau-1}\|\boldsymbol{\epsilon}(t-\tau)\| & \text{(6b)} \\ \|\mathbf{x}(\tau) - \mathbf{x}(\tau-1)\| \le 2\Gamma\left[B_{\mathbf{x}}\gamma^{\tau-2} + \sum_{\varsigma=1}^{\tau-1}\gamma^{\varsigma-2}\|\boldsymbol{\epsilon}(\tau-\varsigma)\|\right] & \text{(6c)} \end{cases}$$

where $\mathbf{x}(t) = \mathrm{col}\{\mathbf{x}_i(t)\}_{i=1}^n$, $\mathbf{g}(t) = \mathrm{col}\{\mathbf{g}_i(t)\}_{i=1}^n$, $\boldsymbol{\epsilon}(t) = \boldsymbol{\xi}(t) + \alpha(t)\mathbf{g}(t)$, $\boldsymbol{\xi}(t) = \mathrm{col}\{\boldsymbol{\xi}_i(t)\}_{i=1}^n$, $\boldsymbol{\xi}_i(t) = P_{\mathcal{X}_i}\{\sum_{i=1}^n \phi_{ij}\mathbf{x}_j(t) + \alpha(t)\mathbf{g}_i(t)\} - \sum_{i=1}^n \phi_{ij}\mathbf{x}_j(t) - \alpha(t)\mathbf{g}_i(t)$, and $\hat{\mathbf{x}}(t) = \mathbf{1}_n \otimes \bar{\mathbf{x}}(t)$ with $\bar{\mathbf{x}}(t) = \sum_{i=1}^n \pi_i\mathbf{x}_i(t)$.

Proof: Rewrite Algorithm 1 in the following compact form:

$$\begin{cases} \mathbf{x}(t+1) = P_{\mathcal{X}}\{\Phi\mathbf{x}(t) + \alpha(t)\mathbf{g}(t)\} & \text{(7a)} \\ \mathbf{g}(t+1) = \Psi\mathbf{g}(t) - \nabla f(\mathbf{x}(t+1)) + \nabla f(\mathbf{x}(t)) & \text{(7b)} \end{cases}$$

where $\nabla f(\mathbf{x}(t)) = \mathrm{col}\{\nabla f_i(\mathbf{x}_i(t))\}_{i=1}^n$. Defining $\mathbf{s}(t+1) = \nabla f(\mathbf{x}(t+1)) - \nabla f(\mathbf{x}(t))$, we rewrite (7b) as follows:

$$\mathbf{g}(t+1) = \Psi\mathbf{g}(t) - \mathbf{s}(t+1).$$

Then, it is not hard to obtain

$$\mathbf{g}(t+1) = \Psi^t\mathbf{g}(1) - \sum_{\iota=2}^{t+1}\Psi^{t+1-\iota}\mathbf{s}(\iota). \tag{8}$$

Substituting $\mathbf{g}_i(1) = -\nabla f_i(\mathbf{x}_i(1))$ and $\mathbf{s}(\iota) = \nabla f(\mathbf{x}(\iota)) - \nabla f(\mathbf{x}(\iota-1))$ with $\iota \ge 2$ into (8), yields

$$\begin{aligned} &\|\mathbf{g}(t+1)\| \\ &\le \sum_{\iota=1}^{t}\left\|\Psi^{t-\iota} - \Psi^{t-\iota+1}\right\|\|\nabla f(\mathbf{x}(\iota))\| + \|\nabla f(\mathbf{x}(t+1))\| \end{aligned}$$

where $\Psi^0 := I$. According to the Lipschitz continuous of $\nabla f_i(\mathbf{x}_i)$ and the boundness of feasible set, $\|\nabla f_i(\mathbf{x}_i(t))\| \le 2G_f B_{\mathbf{x}}$. Then, by Lemma 1, we obtain (6a). For (7a)

$$\begin{aligned} \mathbf{x(t+1)} &= \Phi\mathbf{x}(t) + \boldsymbol{\epsilon}(t) \\ &= \Phi^t\mathbf{x}(1) + \sum_{\tau=1}^{t}\Phi^{\tau-1}\boldsymbol{\epsilon}(t+1-\tau). \end{aligned} \tag{9}$$

Multiplying $\hat{\boldsymbol{\pi}}^\top$ on both sides of (7a) yields

$$\hat{\mathbf{x}}(t+1) = \hat{\boldsymbol{\pi}}^\top\mathbf{x}(1) + \sum_{\tau=1}^{t}\hat{\boldsymbol{\pi}}^\top\boldsymbol{\epsilon}(t+1-\tau). \tag{10}$$

By (9) and (10)

$$\begin{aligned} &\|\mathbf{x}(t+1) - \hat{\mathbf{x}}(t+1)\| \\ &\le \|\Phi^t - \hat{\boldsymbol{\pi}}^\top\|\|\mathbf{x}(1)\| + \sum_{\tau=1}^{t}\boldsymbol{\epsilon}(t+1-\tau)\|\Phi^{\tau-1} - \hat{\boldsymbol{\pi}}^\top\| \\ &\le \Gamma B_{\mathbf{x}}\gamma^t + \Gamma\sum_{\tau=1}^{t}\gamma^{\tau-1}\boldsymbol{\epsilon}(t+1-\tau) \end{aligned} \tag{11}$$

where the last inequality follows from Lemma 1 and Assumption 2. For (11) at the tth step, we obtain (6b).

Moreover, by substituting $t = \tau - 1$ and $t = \tau - 2$ into (9) and taking Lemma 1, it yields (6c). □

Lemma 3: Let Assumptions 1–3 hold. If $(\mathbf{x}_i(t), \mathbf{g}_i(t))$ is generated by Algorithm 1, then

$$\begin{aligned} &\left\|\boldsymbol{\pi}^\top\mathbf{u}\sum_{i=1}^n\nabla f_i(\mathbf{x}_i(t)) + \sum_{i=1}^n\pi_i\mathbf{g}_i(t)\right\| \\ &\le \Gamma L_g\gamma^{t-1} + 2\Gamma^2 B_{\mathbf{x}}G_f(t-1)\gamma^{t-2} \\ &\quad + 2\Gamma^2 G_f\sum_{\tau=2}^{t}\sum_{\varsigma=1}^{\tau-1}\gamma^{t-\tau+\varsigma-2}\|\boldsymbol{\epsilon}(\tau-\varsigma)\|. \end{aligned}$$

Proof: For (7b), we get

$$\begin{aligned}&\hat{\mathbf{u}}\left[\mathbf{g}(t)+\nabla f(\mathbf{x}(t))\right]\\&=(\mathbf{u}\otimes I_m)\mathbf{1}_n^{m\top}\left[\mathbf{g}(t-1)+\nabla f(\mathbf{x}(t-1))\right]\\&=(\mathbf{u}\otimes I_m)\mathbf{1}_n^{m\top}\left[\mathbf{g}(1)+\nabla f(\mathbf{x}(1))\right]\\&=0\end{aligned}$$

where $\mathbf{1}_n^m=\mathbf{1}_n\otimes I_m$. Therefore,

$$\hat{\mathbf{u}}\nabla f(\mathbf{x}(t))+\mathbf{g}(t)=-\hat{\mathbf{u}}\mathbf{g}(t)+\mathbf{g}(t). \tag{12}$$

According to (7b),

$$\hat{\mathbf{u}}\mathbf{g}(t)=\hat{\mathbf{u}}\mathbf{g}(1)-\sum_{\iota=2}^{t}\hat{\mathbf{u}}\mathbf{s}(\iota). \tag{13}$$

Left multiplying $\bar{\boldsymbol{\pi}}^\top$ on the both two sides of (12), and according to (8) and (13), we obtain

$$\begin{aligned}&\bar{\boldsymbol{\pi}}^\top\hat{\mathbf{u}}\nabla f(\mathbf{x}(t))+\bar{\boldsymbol{\pi}}^\top\mathbf{g}(t)\\&=\bar{\boldsymbol{\pi}}^\top(\Psi^{t-1}-\hat{\mathbf{u}})\mathbf{g}(1)+\bar{\boldsymbol{\pi}}^\top\sum_{\iota=2}^{t}(\Psi^{t-\iota}-\hat{\mathbf{u}})\mathbf{s}(\iota).\end{aligned}$$

Consequently,

$$\begin{aligned}&\left\|\boldsymbol{\pi}^\top\mathbf{u}\sum_{i=1}^{n}\nabla f_i(\mathbf{x}_i(t))+\sum_{i=1}^{n}\pi_i\mathbf{g}_i(t)\right\|\\&\le L_g\|\Psi^{t-1}-\hat{\mathbf{u}}\|+G_f\sum_{\tau=2}^{t}\|\Psi^{t-\tau}-\hat{\mathbf{u}}\|\|\mathbf{x}(\tau)-\mathbf{x}(\tau-1)\|. \end{aligned}\tag{14}$$

By Lemmas 1 and 2, the conclusion follows. □

Recalling Lemmas 2 and 3 and the corresponding proofs, it is not difficult to obtain the following lemma for the analysis of Algorithm 2.

Lemma 4: If Assumptions 1–3 hold, then for Algorithm 2, $\|\mathbf{g}(k,\tau)\|\le L_g$ and

$$\begin{cases}\|\mathbf{x}(k,\tau)-\hat{\mathbf{x}}(k,\tau)\|\le B_{\mathbf{x}}\Gamma\gamma^{\tau-1}+\Gamma\sum_{\iota=1}^{\tau-1}\gamma^{\iota-1}\|\boldsymbol{\epsilon}(k,\tau-\iota)\|\\ \|\mathbf{x}(k,\tau)-\mathbf{x}(k,\tau-1)\|\le 2\Gamma\Big[B_{\mathbf{x}}\gamma^{\tau-2}+\sum_{\varsigma=1}^{\tau-1}\gamma^{\varsigma-2}\|\boldsymbol{\epsilon}(k,\tau-\varsigma)\|\Big].\end{cases}$$

Moreover,

$$\begin{aligned}&\left\|\boldsymbol{\pi}^\top\mathbf{u}\sum_{i=1}^{n}\nabla f_i(\mathbf{x}_i(k,\tau))+\sum_{i=1}^{n}\pi_i\mathbf{g}_i(k,\tau)\right\|\le\Gamma L_g\gamma^{\tau-1}\\&+2\Gamma^2G_f\left[B_{\mathbf{x}}(\tau-1)\gamma^{\tau-2}+\sum_{\iota=2}^{\tau}\sum_{\varsigma=1}^{\iota-1}\gamma^{\tau-\iota+\varsigma-2}\|\boldsymbol{\epsilon}(k,\iota-\varsigma)\|\right]\end{aligned}$$

where $\mathbf{x}(k,\tau)=\mathrm{col}\{\mathbf{x}_i(k,\tau)\}_{i=1}^n$, $\mathbf{g}(k,\tau)=\mathrm{col}\{\mathbf{g}_i(k,\tau)\}_{i=1}^n$, and $\boldsymbol{\epsilon}(k,\tau)=\boldsymbol{\xi}(k,\tau)+\alpha(k)\mathbf{g}(k,\tau)$.

B. Proof of Theorem 1

By the dynamic of $\mathbf{x}_i(t)$ in Algorithm 1,

$$\begin{aligned}&\sum_{i=1}^{n}\pi_i\|\mathbf{x}_i(t+1)-\mathbf{x}^*\|^2\\&\le\sum_{i=1}^{n}\pi_i\Big\|\sum_{j=1}^{n}\phi_{ij}\mathbf{x}_j(t)-\mathbf{x}^*+\alpha(t)\mathbf{g}_i(t)\Big\|^2-\|\boldsymbol{\xi}(t)\|^2\\&\le\sum_{i=1}^{n}\pi_i\|\mathbf{x}_i(t)-\mathbf{x}^*\|^2+L_g^2(\alpha(t))^2\\&\quad+2\alpha(t)\sum_{i=1}^{n}\pi_i\left\langle\sum_{j=1}^{n}\phi_{ij}\mathbf{x}_j(t)-\mathbf{x}^*,\mathbf{g}_i(t)\right\rangle-\|\boldsymbol{\xi}(t)\|^2\\&=\sum_{i=1}^{n}\pi_i\|\mathbf{x}_i(t)-\mathbf{x}^*\|^2+L_g^2(\alpha(t))^2\\&\quad+2\alpha(t)\left[\sum_{i=1}^{n}\pi_i\left\langle\sum_{j=1}^{n}\phi_{ij}\mathbf{x}_j(t)-\bar{\mathbf{x}}(t),\mathbf{g}_i(t)\right\rangle\right.\\&\quad+\left\langle\bar{\mathbf{x}}(t)-\mathbf{x}^*,\sum_{i=1}^{n}\pi_i\mathbf{g}_i(t)+\boldsymbol{\pi}^\top\mathbf{u}\sum_{i=1}^{n}\nabla f_i(\mathbf{x}_i(t))\right\rangle\\&\quad-\boldsymbol{\pi}^\top\mathbf{u}\left\langle\bar{\mathbf{x}}(t)-\mathbf{x}^*,\sum_{i=1}^{n}\left(\nabla f_i(\mathbf{x}_i(t))-\nabla f_i(\bar{\mathbf{x}}(t))\right)\right\rangle\\&\quad\left.-\boldsymbol{\pi}^\top\mathbf{u}\left\langle\bar{\mathbf{x}}(t)-\mathbf{x}^*,\sum_{i=1}^{n}\nabla f_i\left(\bar{\mathbf{x}}(t)\right)\right\rangle\right]-\|\boldsymbol{\xi}(t)\|^2\end{aligned}\tag{15}$$

where the first inequality follows from $\|P_{\mathcal{X}_i}\{\mathbf{a}\}-\mathbf{b}\|^2\le\|\mathbf{a}-\mathbf{b}\|^2-\|P_{\mathcal{X}_i}\{\mathbf{a}\}-\mathbf{a}\|^2\ \ \forall\mathbf{a}\in\mathbb{R}^m,\mathbf{b}\in\mathcal{X}_i\subseteq\mathbb{R}^m$, the second one follows from $\|\mathbf{a}+\mathbf{b}\|^2=\|\mathbf{a}\|^2+\|\mathbf{b}\|^2+2\langle\mathbf{a},\mathbf{b}\rangle$, $\|\mathbf{g}_i(t)\|\le L_g$, and the convexity of quadratic function, and the last equality follows from adding and subtracting $2\alpha(t)\sum_{i=1}^n\pi_i\left\langle\bar{\mathbf{x}}(t),\mathbf{g}_i(t)\right\rangle$ and $2\alpha(t)\boldsymbol{\pi}^\top\mathbf{u}\left\langle\bar{\mathbf{x}}(t)-\mathbf{x}^*,\sum_{i=1}^n\left[\nabla f_i(\mathbf{x}_i(t))-\nabla f_i(\bar{\mathbf{x}}(t))\right]\right\rangle$. Recalling Assumption 4,

$$\langle\bar{\mathbf{x}}(t)-\mathbf{x}^*,\nabla f_i(\bar{\mathbf{x}}(t))\rangle\ge f_i(\bar{\mathbf{x}}(t))-f_i(\mathbf{x}^*)+\frac{\sigma}{2}\|\bar{\mathbf{x}}(t)-\mathbf{x}^*\|. \tag{16}$$

Substituting (16) into (15), yields

$$\begin{aligned}&\boldsymbol{\pi}^\top\mathbf{u}\sum_{i=1}^{n}\left[f_i(\bar{\mathbf{x}}(t))-f_i(\mathbf{x}^*)\right]\\&\le\frac{1}{2\alpha(t)}\sum_{i=1}^{n}\pi_i\left[\|\mathbf{x}_i(t)-\mathbf{x}^*\|^2-\|\mathbf{x}_i(t+1)-\mathbf{x}^*\|^2\right]\\&\quad+\frac{L_g^2\alpha(t)}{2}+\sum_{i=1}^{n}\pi_i\left\langle\sum_{j=1}^{n}\phi_{ij}\mathbf{x}_j(t)-\bar{\mathbf{x}}(t),\mathbf{g}_i(t)\right\rangle\\&\quad+\left\langle\bar{\mathbf{x}}(t)-\mathbf{x}^*,\sum_{i=1}^{n}\pi_i\mathbf{g}_i(t)+\boldsymbol{\pi}^\top\mathbf{u}\sum_{i=1}^{n}\nabla f_i(\mathbf{x}_i(t))\right\rangle\\&\quad-\boldsymbol{\pi}^\top\mathbf{u}\left\langle\bar{\mathbf{x}}(t)-\mathbf{x}^*,\sum_{i=1}^{n}\left[\nabla f_i(\mathbf{x}_i(t))-\nabla f_i(\bar{\mathbf{x}}(t))\right]\right\rangle\\&\quad-\frac{\sigma\boldsymbol{\pi}^\top\mathbf{u}}{2}\sum_{i=1}^{n}\|\bar{\mathbf{x}}(t)-\mathbf{x}^*\|^2-\frac{\|\boldsymbol{\xi}(t)\|^2}{2\alpha(t)}.\end{aligned}\tag{17}$$

Define $\bar{\mathbf{v}}(t)=\frac{d_o}{d(t)+d_o}\bar{\mathbf{x}}(t)+\frac{d(t)}{d(t)+d_o}\mathbf{x}_o$, where $d(t)=\sum_{i=1}^n\mathrm{dist}\{\bar{\mathbf{x}}(t),\mathcal{X}_i\}$, $\mathbf{x}_o\in\mathcal{X}_o$, and $d_o>0$ with $\{\mathbf{z}\in\mathbb{R}^m|\|\mathbf{z}-\mathbf{x}_o\le d_o\|\}\subset\mathcal{X}_o$. From [14, Lemma 2], $\bar{\mathbf{v}}(t)\in\mathcal{X}_o$. Then

$$\begin{aligned}&f(\bar{\mathbf{v}}(t))-f(\mathbf{x}^*)\\&=\sum_{i=1}^{n}\left[f_i(\bar{\mathbf{v}}(t))-f_i(\bar{\mathbf{x}}(t))+f_i(\bar{\mathbf{x}}(t))-f_i(\mathbf{x}^*)\right]\\&\le 2nB_{\mathbf{x}}G_f\|\bar{\mathbf{v}}(t)-\bar{\mathbf{x}}(t)\|+\sum_{i=1}^{n}\left[f_i(\bar{\mathbf{x}}(t))-f_i(\mathbf{x}^*)\right]\\&=\frac{2nB_{\mathbf{x}}G_fd(t)}{d(t)+d_o}\|\bar{\mathbf{x}}(t)-\mathbf{x}_o\|+\sum_{i=1}^{n}\left[f_i(\bar{\mathbf{x}}(t))-f_i(\mathbf{x}^*)\right]\\&\le\frac{4n^2B_{\mathbf{x}}^2G_f}{d_o}\|\mathbf{x}(t)-\hat{\mathbf{x}}(t)\|+\sum_{i=1}^{n}\left[f_i(\bar{\mathbf{x}}(t))-f_i(\mathbf{x}^*)\right]\end{aligned}\tag{18}$$

where the first inequality follows from $f_i(\bar{\mathbf{v}}(t))-f_i(\bar{\mathbf{x}}(t))\le 2B_{\mathbf{x}}G_f\|\bar{\mathbf{v}}(t)-\bar{\mathbf{x}}(t)\|$, the second inequality holds by $\|\bar{\mathbf{x}}(t)-\mathbf{x}_o\|\le 2B_{\mathbf{x}}$, and $d(t)\le\sum_{i=1}^n\|\bar{\mathbf{x}}(t)-\mathbf{x}_i(t)\|\le n\|\mathbf{x}(t)-\hat{\mathbf{x}}(t)\|$. Since $\alpha(t)=\frac{1}{\sigma n\boldsymbol{\pi}^\top\mathbf{u}t}$, $\|\mathbf{g}_i\|\le L_g$, $\|\bar{\mathbf{x}}(t)-\mathbf{x}^*\|\le 2B_{\mathbf{x}}$, $\|\nabla f_i(\mathbf{x}_i(t))-$

$\nabla f_i(\bar{\mathbf{x}}(t))\| \leq G_f\|\mathbf{x}_i(t)-\bar{\mathbf{x}}\|$, it follows from (17) and (18) that,

$$\begin{aligned} &f(\bar{\mathbf{v}}(t)) - f(\mathbf{x}^*) + \frac{\|\boldsymbol{\xi}(t)\|^2}{2\boldsymbol{\pi}^\top\mathbf{u}\alpha(t)} \\ &\quad\leq \frac{n\sigma t}{2}\sum_{i=1}^n \pi_i\left[\|\mathbf{x}_i(t)-\mathbf{x}^*\|^2 - \|\mathbf{x}_i(t+1)-\mathbf{x}^*\|^2\right] \\ &\qquad - \frac{n\sigma}{2}\sum_{i=1}^n \pi_i\|\bar{\mathbf{x}}(t)-\mathbf{x}^*\|^2 + \frac{L_g^2\alpha(t)}{2\boldsymbol{\pi}^\top\mathbf{u}} + \frac{nL_g}{\boldsymbol{\pi}^\top\mathbf{u}}\|\Phi\mathbf{x}(t)-\hat{\mathbf{x}}(t)\| \\ &\qquad + \frac{2B_{\mathbf{x}}}{\boldsymbol{\pi}^\top\mathbf{u}}\Big\|\sum_{i=1}^n\pi_i\mathbf{g}_i(t) + \boldsymbol{\pi}^\top\mathbf{u}\sum_{i=1}^n\nabla f_i(\mathbf{x}_i(t))\Big\| \\ &\qquad + 2nB_{\mathbf{x}}G_f\left(1+\frac{2nB_{\mathbf{x}}}{d_o}\right)\|\mathbf{x}(t)-\hat{\mathbf{x}}(t)\|. \end{aligned} \tag{19}$$

Taking the summation of (19) from $t=1$ to T, we get

$$\begin{aligned} &\sum_{t=1}^T\left[f(\bar{\mathbf{v}}(t)) - f(\mathbf{x}^*) + \frac{\|\boldsymbol{\xi}(t)\|^2}{2\boldsymbol{\pi}^\top\mathbf{u}\alpha(t)}\right] \\ &\quad\leq 2nB_{\mathbf{x}}\left[n\sigma + G_f\left(1+\frac{2nB_{\mathbf{x}}}{d_o}\right)\right]\sum_{t=1}^T\|\mathbf{x}(t)-\hat{\mathbf{x}}(t)\| \\ &\qquad + \frac{L_g^2}{2\boldsymbol{\pi}^\top\mathbf{u}}\sum_{t=1}^T\alpha(t) + \frac{nL_g}{\boldsymbol{\pi}^\top\mathbf{u}}\sum_{t=1}^T\|\Phi\mathbf{x}(t)-\hat{\mathbf{x}}(t)\| \\ &\qquad + \frac{2B_{\mathbf{x}}}{\boldsymbol{\pi}^\top\mathbf{u}}\sum_{t=1}^T\Big\|\sum_{i=1}^n\pi_i\mathbf{g}_i(t) + \boldsymbol{\pi}^\top\mathbf{u}\sum_{i=1}^n\nabla f_i(\mathbf{x}_i(t))\Big\| \end{aligned} \tag{20}$$

where the inequality follows from $\sum_{t=1}^T t\sum_{i=1}^n\pi_i[\|\mathbf{x}_i(t)-\mathbf{x}^*\|^2 - \|\mathbf{x}_i(t+1)-\mathbf{x}^*\|^2] \leq \sum_{t=1}^T\sum_{i=1}^n\pi_i\|\mathbf{x}_i(t)-\mathbf{x}^*\|^2$ and $\sum_{t=1}^T\sum_{i=1}^n\pi_i[\|\mathbf{x}_i(t)-\mathbf{x}^*\|^2 - \|\bar{\mathbf{x}}(t)-\mathbf{x}^*\|^2] \leq 4nB_{\mathbf{x}}\sum_{t=1}^T\|\mathbf{x}(t)-\hat{\mathbf{x}}(t)\|$. Based on $\|\boldsymbol{\epsilon}(t)\| \leq L_g\alpha(t) + \|\boldsymbol{\xi}(t)\| \leq (L_g + 4C_0\boldsymbol{\pi}^\top\mathbf{u})\alpha(t) + \frac{\|\boldsymbol{\xi}(t)\|^2}{4C_0\boldsymbol{\pi}^\top\mathbf{u}\alpha(t)}$, Lemmas 2 and 3, and the strong convexity of $f_i(\cdot)$

$$\begin{aligned} &\frac{n\sigma}{2}\sum_{t=1}^T\|\bar{\mathbf{v}}(t)-\mathbf{x}^*\|^2 + \sum_{t=1}^T\frac{\|\boldsymbol{\xi}(t)\|^2}{2\boldsymbol{\pi}^\top\mathbf{u}\alpha(t)} \\ &\leq \sum_{t=1}^T f(\bar{\mathbf{v}}(t)) - f(\mathbf{x}^*) + \sum_{t=1}^T\frac{\|\boldsymbol{\xi}(t)\|^2}{2\boldsymbol{\pi}^\top\mathbf{u}\alpha(t)} \\ &\leq \left[2nB_{\mathbf{x}}\left(n\sigma + G_f\left(1+\frac{2nB_{\mathbf{x}}}{d_o}\right)\right) + \frac{(n+2)L_g}{\boldsymbol{\pi}^\top\mathbf{u}}\right]\Gamma B_{\mathbf{x}}\sum_{t=1}^T\gamma^{t-1} \\ &\quad + \left[2nB_{\mathbf{x}}\left(n\sigma + G_f\left(1+\frac{2nB_{\mathbf{x}}}{d_o}\right)\right) + \frac{nL_g}{\boldsymbol{\pi}^\top\mathbf{u}}\right] \\ &\quad \times\Gamma\sum_{t=2}^T\sum_{\tau=1}^{t-1}\gamma^{\tau-1}\|\boldsymbol{\epsilon}(t-\tau)\| \\ &\quad + \frac{L_g^2}{2\boldsymbol{\pi}^\top\mathbf{u}}\sum_{t=1}^T\alpha(t) + \frac{4B_{\mathbf{x}}^2G_f\Gamma^2}{\boldsymbol{\pi}^\top\mathbf{u}}\sum_{t=2}^T(t-1)\gamma^{t-2} \\ &\quad + \frac{4B_{\mathbf{x}}G_f\Gamma^2}{\boldsymbol{\pi}^\top\mathbf{u}}\sum_{t=2}^T\sum_{\tau=2}^t\sum_{\varsigma=1}^{\tau-1}\gamma^{t-\tau+\varsigma-2}\|\boldsymbol{\epsilon}(\tau-\varsigma)\| \\ &\leq C_4B_{\mathbf{x}}M_1(T) + \frac{L_g^2}{2\left(\boldsymbol{\pi}^\top\mathbf{u}\right)^2n\sigma}M_2(T) + \frac{4B_{\mathbf{x}}^2G_f\Gamma^2}{\boldsymbol{\pi}^\top\mathbf{u}}M_3(T) \\ &\quad + C_4M_4(T) + \frac{4B_{\mathbf{x}}G_f\Gamma^2}{\boldsymbol{\pi}^\top\mathbf{u}}M_5(T) \end{aligned}$$

where

$$M_1(T) := \sum_{t=1}^T\gamma^{t-1} \leq \frac{1}{1-\gamma}, M_2(T) := \sum_{t=1}^T\frac{1}{t} \leq 2\ln T$$

$$M_3(T) := \sum_{t=2}^T(t-1)\gamma^{t-2} \leq \frac{1}{(1-\gamma)^2}$$

$$\begin{aligned} M_4(T) &:= \sum_{t=2}^T\sum_{\tau=1}^{t-1}\gamma^{\tau-1}\|\boldsymbol{\epsilon}(t-\tau)\| \leq \sum_{t=1}^{T-1}\frac{\|\boldsymbol{\epsilon}(t)\|}{1-\gamma} \\ &\leq \frac{2(L_g+4C_0\boldsymbol{\pi}^\top\mathbf{u})\ln T}{(1-\gamma)\sigma n\boldsymbol{\pi}^\top\mathbf{u}} + \sum_{t=1}^T\frac{\|\boldsymbol{\xi}(t)\|^2}{4(1-\gamma)C_0\boldsymbol{\pi}^\top\mathbf{u}\alpha(t)} \\ M_5(T) &:= \sum_{t=2}^T\sum_{\tau=2}^t\sum_{\varsigma=1}^{\tau-1}\gamma^{t-\tau+\varsigma-2}\|\boldsymbol{\epsilon}(\tau-\varsigma)\| \leq \sum_{t=1}^{T-1}\frac{\|\boldsymbol{\epsilon}(t)\|}{\gamma(1-\gamma)^2} \\ &\leq \frac{2(L_g+4C_0\boldsymbol{\pi}^\top\mathbf{u})\ln T}{\gamma(1-\gamma)^2\sigma n\boldsymbol{\pi}^\top\mathbf{u}} + \sum_{t=1}^T\frac{\|\boldsymbol{\xi}(t)\|^2}{4\gamma(1-\gamma)^2C_0\boldsymbol{\pi}^\top\mathbf{u}\alpha(t)}. \end{aligned}$$

As a result,

$$\frac{n\sigma}{2}\sum_{t=1}^T\|\bar{\mathbf{v}}(t)-\mathbf{x}^*\|^2 + \sum_{t=1}^T\frac{\|\boldsymbol{\xi}(t)\|^2}{4\boldsymbol{\pi}^\top\mathbf{u}\alpha(t)} \leq \frac{\sigma}{4}\left(C_1 + C_2\ln T\right). \tag{21}$$

Define $\bar{\mathbf{v}}_{\rm av}(T) = \frac{\sum_{t=1}^T\bar{\mathbf{v}}(t)}{T}$ and $\mathbf{x}_{\rm av}(T) = \frac{\sum_{t=1}^T\mathbf{x}(t)}{T}$,

$$\begin{aligned} &\|\mathbf{x}_{\rm av}(T) - \mathbf{1}_n\otimes\bar{\mathbf{v}}_{\rm av}(T)\|^2 \\ &\leq \left[\frac{1}{T}\sum_{t=1}^T\left(\|\mathbf{x}(t)-\hat{\mathbf{x}}(t)\| + n\|\bar{\mathbf{x}}(t)-\bar{\mathbf{v}}(t)\|\right)\right]^2 \\ &\leq \left[\frac{2n^2B_{\mathbf{x}}+d_o}{d_oT}\sum_{t=1}^T\|\mathbf{x}(t)-\hat{\mathbf{x}}(t)\|\right]^2 \\ &\leq \left[\frac{(2n^2B_{\mathbf{x}}+d_o)\Gamma}{d_oT}\left(B_{\mathbf{x}}M_1(T) + M_4(T)\right)\right]^2 \leq \frac{C_3^2\ln^2(T+1)}{2T^2} \end{aligned} \tag{22}$$

where $C_3 = \frac{2\Gamma(2n^2B_{\mathbf{x}}+d_o)}{(1-\gamma)d_o}\left[B_{\mathbf{x}} + \frac{2L_g}{n\sigma\boldsymbol{\pi}^\top\mathbf{u}} + \frac{8C_0}{n\sigma} + \frac{\sigma(C_1+C_2)}{4C_0}\right]$, the second inequality is deduced by $\|\bar{\mathbf{x}}(t)-\bar{\mathbf{v}}(t)\| \leq \frac{2nB_{\mathbf{x}}}{d_o}\|\mathbf{x}(t)-\hat{\mathbf{x}}(t)\|$ in (18), the third one follows from Lemma 2, and the last one infers from $\sum_{t=1}^T\frac{\|\boldsymbol{\xi}(t)\|^2}{\sigma\boldsymbol{\pi}^\top\mathbf{u}\alpha(t)} \leq (C_1+C_2\ln T) \leq (C_1+C_2)\ln T$. Combining (21) and (22), yields

$$\begin{aligned} &\|\mathbf{x}_{\rm av}(T) - \mathbf{1}_n\otimes\mathbf{x}^*\|^2 \\ &\quad\leq 2n\|\bar{\mathbf{v}}_{\rm av}(T)-\mathbf{x}^*\|^2 + 2\|\mathbf{x}_{\rm av}(T) - \mathbf{1}_n\otimes\bar{\mathbf{v}}_{\rm av}(T)\|^2 \\ &\quad\leq \frac{C_1 + C_2\ln(T+1)}{T} + \frac{C_3^2\ln^2(T+1)}{T^2}. \end{aligned} \tag{23}$$

Therefore, we completed the proof. □

A3 Proof of Theorem 2

Defining $\bar{\mathbf{v}}(k,\tau)$ and $d(k,\tau)$ as the counterparts in the proof of Theorem 1, and according to (17) with a fixed step-size $\alpha(k)$ at the kth epoch, and with similar deductions of (18) and (19), we have

$$\begin{aligned} &f(\bar{\mathbf{v}}(k,\tau)) - f(\mathbf{x}^*) \\ &\quad\leq \frac{1}{2\boldsymbol{\pi}^\top\mathbf{u}\alpha(k)}\sum_{i=1}^n\pi_i\left[\|\mathbf{x}_i(k,\tau)-\mathbf{x}^*\|^2 - \|\mathbf{x}_i(k,\tau+1)\right. \\ &\qquad\left.-\mathbf{x}^*\|^2\right] + \frac{L_g^2\alpha(k)}{2\boldsymbol{\pi}^\top\mathbf{u}} + \frac{nL_g}{\boldsymbol{\pi}^\top\mathbf{u}}\|\Phi\mathbf{x}(k,\tau)-\hat{\mathbf{x}}(k,\tau)\| \\ &\qquad + \frac{2B_{\mathbf{x}}}{\boldsymbol{\pi}^\top\mathbf{u}}\Big\|\sum_{i=1}^n\pi_i\mathbf{g}_i(k,\tau) + \boldsymbol{\pi}^\top\mathbf{u}\sum_{i=1}^n\nabla f_i(\mathbf{x}_i(k,\tau))\Big\| \\ &\qquad + 2nB_{\mathbf{x}}G_f\left(1+\frac{2nB_{\mathbf{x}}}{d_o}\right)\|\mathbf{x}(k,\tau)-\hat{\mathbf{x}}(k,\tau)\| - \frac{\|\boldsymbol{\xi}(k,\tau)\|^2}{2\boldsymbol{\pi}^\top\mathbf{u}\alpha(k)}. \end{aligned} \tag{24}$$

According to Assumption 4 and Lemma 4,

$$
\begin{aligned}
n\|\bar{\mathbf{v}}(\bar{k}+1,1)-\mathbf{x}^*\|^2 &\le \frac{2}{\sigma}\left[f(\bar{\mathbf{v}}(\bar{k}+1),1)-f(\mathbf{x}^*)\right]\\
&\le \frac{2}{\sigma T(\bar{k})}\sum_{\tau=1}^{T(\bar{k})}\left[f(\bar{\mathbf{v}}(\bar{k},\tau))-f(\mathbf{x}^*)\right]\\
&\le \frac{\|\mathbf{x}(\bar{k},1)-\hat{\mathbf{x}}^*\|^2-\eta(\bar{k})}{\boldsymbol{\pi}^\top\mathbf{u}\sigma\alpha(\bar{k})T(\bar{k})}+\frac{L_g^2\alpha(\bar{k})}{\boldsymbol{\pi}^\top\mathbf{u}\sigma}\\
&\quad+\frac{2n\Gamma}{\sigma T(\bar{k})}\left[\frac{L_g}{\boldsymbol{\pi}^\top\mathbf{u}}+2B_{\mathbf{x}}G_f\left(1+\frac{2nB_{\mathbf{x}}}{d_o}\right)\right]\sum_{\tau=1}^{T(\bar{k})}\Big[B_{\mathbf{x}}\gamma^{\tau-1}\\
&\quad+\sum_{\iota=1}^{\tau-1}\gamma^{\iota-1}\|\boldsymbol{\epsilon}(k,\tau-\iota)\|+\frac{8B_{\mathbf{x}}\Gamma^2G_f}{\boldsymbol{\pi}^\top\mathbf{u}\sigma T(\bar{k})}\sum_{\tau=1}^{T(\bar{k})}\Big[\frac{L_g}{2\Gamma G_f}\gamma^{\tau-1}\\
&\quad+B_{\mathbf{x}}(\tau-1)\gamma^{\tau-2}+\sum_{\iota=2}^{\tau}\sum_{\varsigma=1}^{\iota-1}\gamma^{\tau-\iota+\varsigma-2}\|\boldsymbol{\epsilon}(k,\iota-\varsigma)\|\Big]\\
&\le \frac{\|\mathbf{x}(\bar{k},1)-\mathbf{x}^*\|^2-\eta(\bar{k})}{\boldsymbol{\pi}^\top\mathbf{u}\sigma\alpha(\bar{k})T(\bar{k})}+\frac{L_g^2\alpha(\bar{k})}{\boldsymbol{\pi}^\top\mathbf{u}\sigma}+\frac{C_9}{T(\bar{k})}\Big[B_{\mathbf{x}}\\
&\quad+\sum_{\tau=1}^{T(\bar{k})}\|\boldsymbol{\epsilon}(\bar{k},\tau)\|\Big]+\frac{C_8}{T(\bar{k})}\left[1+\frac{2\Gamma G_f}{L_g}\left(B_{\mathbf{x}}+\sum_{\tau=1}^{T(\bar{k})}\|\boldsymbol{\epsilon}(\bar{k},\tau)\|\right)\right]\\
&\le \frac{C_6}{T(\bar{k})}+C_7\alpha(\bar{k})+\frac{2\|\mathbf{x}(\bar{k},1)-\hat{\mathbf{x}}^*\|^2-\eta(\bar{k})}{2\boldsymbol{\pi}^\top\mathbf{u}\sigma\alpha(\bar{k})T(\bar{k})}
\end{aligned}\tag{25}
$$

where $\eta(k)=\sum_{\tau=1}^{T(k)}\|\boldsymbol{\xi}(k,\tau)\|^2$, the last inequality is deduced by taking $\|\boldsymbol{\epsilon}(k,\tau)\|=L_g\alpha(k)+\|\boldsymbol{\xi}(k,\tau)\|\le(L_g+2\boldsymbol{\pi}^\top\mathbf{u}\sigma C_{10})\alpha(k)+\frac{\|\boldsymbol{\xi}(k,\tau)\|^2}{2\boldsymbol{\pi}^\top\mathbf{u}\sigma C_{10}\alpha(k)}$, and rearranging related terms with the following parameters:

$$
\begin{aligned}
C_6 &= C_8\left(1+\frac{2B_{\mathbf{x}}\Gamma G_f}{L_g}\right)+C_9B_{\mathbf{x}}\\
C_7 &= C_{10}(L_g+2\boldsymbol{\pi}^\top\mathbf{u}\sigma C_{10})+\frac{L_g^2}{\boldsymbol{\pi}^\top\mathbf{u}\sigma}\\
C_8 &= \frac{4B_{\mathbf{x}}\Gamma L_g}{\boldsymbol{\pi}^\top\mathbf{u}\sigma\gamma^2(1-\gamma)^2}\\
C_9 &= \frac{2n\Gamma}{\sigma(1-\gamma)}\left[4B_{\mathbf{x}}G_f\left(1+\frac{nB_{\mathbf{x}}}{d_o}\right)+\frac{L_g}{\boldsymbol{\pi}^\top\mathbf{u}}\right]\\
C_{10} &= \frac{2\Gamma G_fC_8}{L_g}+C_9.
\end{aligned}
$$

Taking the second inequality of (15) with index (k,τ), adding and subtracting $\langle\sum_{i=1}^n\pi_i\mathbf{g}_i(k,\tau),\bar{\mathbf{x}}(k,\tau)\rangle$, we have

$$
\begin{aligned}
&-\left\langle\sum_{i=1}^n\pi_i\mathbf{g}_i(k,\tau),\bar{\mathbf{x}}(k,\tau)-\mathbf{x}^*\right\rangle+\frac{\|\boldsymbol{\xi}(k,\tau)\|^2}{2\alpha(k)}\\
&\le\frac{1}{2\alpha(k)}\sum_{i=1}^n\pi_i\left[\|\mathbf{x}_i(k,\tau)-\mathbf{x}^*\|^2-\|\mathbf{x}_i(k,\tau+1)-\mathbf{x}^*\|^2\right]\\
&\quad+\frac{L_g^2\alpha(k)}{2}+\sum_{i=1}^n\pi_i\left\langle\sum_{j=1}^n\phi_{ij}\mathbf{x}_j(k,\tau)-\bar{\mathbf{x}}(k,\tau),\mathbf{g}_i(k,\tau)\right\rangle.
\end{aligned}\tag{26}
$$

Note that

$$
\begin{aligned}
&\sum_{i=1}^n[f_i(\bar{\mathbf{v}}(k,\tau))-f_i(\mathbf{x}_i(k,\tau))]\\
&\quad\le 2\sqrt{n}B_{\mathbf{x}}G_f\|\mathbf{1}_n\otimes\bar{\mathbf{v}}(k,\tau)-\mathbf{x}(k,\tau)\|\\
&\quad\le 2nB_{\mathbf{x}}G_f\left[\|\bar{\mathbf{v}}(k,\tau)-\bar{\mathbf{x}}(k,\tau)\|+\|\hat{\mathbf{x}}(k,\tau)-\mathbf{x}(k,\tau)\|\right]\\
&\quad\le 2nB_{\mathbf{x}}G_f\left[1+\frac{2nB_{\mathbf{x}}}{d_o}\right]\|\hat{\mathbf{x}}(k,\tau)-\mathbf{x}(k,\tau)\|
\end{aligned}\tag{27}
$$

where the first inequality follows from $\nabla f_i(\cdot)\le 2B_{\mathbf{x}}G_f$, second one follows from adding and subtracting $\hat{\mathbf{x}}(k,\tau):=\mathbf{1}_n\otimes\bar{\mathbf{x}}(k,\tau)$, and the last one is deduced by $\|\bar{\mathbf{v}}(k,\tau)-\bar{\mathbf{x}}(k,\tau)\|\le\frac{2nB_{\mathbf{x}}}{d_o}\|\hat{\mathbf{x}}(k,\tau)-\mathbf{x}(k,\tau)\|$ in (18). Moreover,

$$
\begin{aligned}
&\frac{\sigma}{2}\sum_{i=1}^n\|\mathbf{x}_i(k,\tau)-\mathbf{x}^*\|^2\\
&\le\sum_{i=1}^n\langle\nabla f_i(\mathbf{x}_i(k,\tau)),\mathbf{x}_i(k,\tau)-\mathbf{x}^*\rangle+\sum_{i=1}^n[f_i(\mathbf{x}^*)\\
&\quad-f_i(\bar{\mathbf{v}}(k,\tau))+f_i(\bar{\mathbf{v}}(k,\tau))-f_i(\mathbf{x}_i(k,\tau))]\\
&\le\sum_{i=1}^n\langle\nabla f_i(\mathbf{x}_i(k,\tau)),\mathbf{x}_i(k,\tau)-\bar{\mathbf{x}}(k,\tau)\rangle\\
&\quad+\frac{1}{\boldsymbol{\pi}^\top\mathbf{u}}\Big\langle\boldsymbol{\pi}^\top\mathbf{u}\sum_{i=1}^n\nabla f_i(\mathbf{x}_i(k,\tau))+\sum_{i=1}^n\pi_i\mathbf{g}_i(k,\tau)\\
&\quad\bar{\mathbf{x}}(k,\tau)-\mathbf{x}^*\Big\rangle-\frac{1}{\boldsymbol{\pi}^\top\mathbf{u}}\left\langle\sum_{i=1}^n\pi_i\mathbf{g}_i(k,\tau),\bar{\mathbf{x}}(k,\tau)-\mathbf{x}^*\right\rangle\\
&\quad+2nB_{\mathbf{x}}G_f\left[1+\frac{2nB_{\mathbf{x}}}{d_o}\right]\|\mathbf{x}(k,\tau)-\hat{\mathbf{x}}(k,\tau)\|\\
&\le\sum_{i=1}^n\langle\nabla f_i(\mathbf{x}_i(k,\tau)),\mathbf{x}_i(k,\tau)-\bar{\mathbf{x}}(k,\tau)\rangle+\frac{L_g^2\alpha(k)}{2\boldsymbol{\pi}^\top\mathbf{u}}\\
&\quad+\frac{1}{\boldsymbol{\pi}^\top\mathbf{u}}\sum_{i=1}^n\pi_i\left\langle\sum_{j=1}^n\phi_{ij}\mathbf{x}_j(k,\tau)-\bar{\mathbf{x}}(k,\tau),\mathbf{g}_i(k,\tau)\right\rangle\\
&\quad+\frac{1}{\boldsymbol{\pi}^\top\mathbf{u}}\left\langle\boldsymbol{\pi}^\top\mathbf{u}\sum_{i=1}^n\nabla f_i(\mathbf{x}_i(k,\tau))+\sum_{i=1}^n\pi_i\mathbf{g}_i(k,\tau),\bar{\mathbf{x}}(k,\tau)-\mathbf{x}^*\right\rangle\\
&\quad+\frac{1}{2\boldsymbol{\pi}^\top\mathbf{u}\alpha(k)}\sum_{i=1}^n\pi_i\left[\|\mathbf{x}_i(k,\tau)-\mathbf{x}^*\|^2-\|\mathbf{x}_i(k,\tau+1)-\mathbf{x}^*\|^2\right]\\
&\quad+2nB_{\mathbf{x}}G_f\left[1+\frac{2nB_{\mathbf{x}}}{d_o}\right]\|\hat{\mathbf{x}}(k,\tau)-\mathbf{x}(k,\tau)\|-\frac{\|\boldsymbol{\xi}(k,\tau)\|^2}{2\boldsymbol{\pi}^\top\mathbf{u}\alpha(k)}
\end{aligned}\tag{28}
$$

where the first inequality follows from Assumption 4 and adding and subtracting $\sum_{i=1}^n f_i(\bar{\mathbf{v}}(k,\tau))$, the second inequality follows from $\sum_{i=1}^n[f_i(\mathbf{x}^*)-f_i(\bar{\mathbf{v}}(k,\tau))]\le 0$ and (27), and the last one is deduced by combining (26).

Based on (28), taking the same deduction of (25), we have

$$
\begin{aligned}
&\|\mathbf{x}(k+1,1)-\hat{\mathbf{x}}^*\|^2\le\frac{1}{T(k)}\sum_{\tau=1}^{T(k)}\sum_{i=1}^n\|\mathbf{x}_i(k,\tau)-\mathbf{x}^*\|^2\\
&\le\frac{2n\Gamma}{(1-\gamma)\sigma}\left[4B_{\mathbf{x}}G_f\left(\frac{nB_{\mathbf{x}}}{d_o}+1\right)+\frac{L_g}{\boldsymbol{\pi}^\top\mathbf{u}}\right]\left[\frac{B_{\mathbf{x}}}{T(k)}+\frac{\sum_{\tau=1}^{T(k)}\|\boldsymbol{\epsilon}(k,\tau)\|}{T(k)}\right]\\
&\quad+\frac{C_8}{T(\bar{k})}\left[1+\frac{2\Gamma G_f}{L_g}\left(B_{\mathbf{x}}+\sum_{\tau=1}^{T(k)}\|\boldsymbol{\epsilon}(k,\tau)\|\right)\right]+\frac{L_g^2\alpha(k)}{\boldsymbol{\pi}^\top\mathbf{u}\sigma}\\
&\quad+\frac{\|\mathbf{x}(k,1)-\hat{\mathbf{x}}^*\|^2-\eta(k)}{\boldsymbol{\pi}^\top\mathbf{u}\sigma\alpha(k)T(k)}\\
&=\frac{C_6}{T(k)}+C_7\alpha(k)+\frac{2\|\mathbf{x}(k,1)-\hat{\mathbf{x}}^*\|^2-\eta(k)}{2\boldsymbol{\pi}^\top\mathbf{u}\sigma\alpha(k)T(k)}.
\end{aligned}\tag{29}
$$

According to (25) and (29),

$$\begin{aligned}
&n\|\bar{\mathbf{v}}(\bar{k}+1,1)-\mathbf{x}^*\|^2+\frac{1}{2}\sum_{\kappa=1}^{\bar{k}}\frac{\eta(\kappa)}{\Pi_{\varsigma=0}^{\bar{k}-\kappa}\boldsymbol{\pi}^\top\mathbf{u}\sigma\alpha(\bar{k}-\varsigma)T(\bar{k}-\varsigma)}\\
&\le\frac{C_6}{T(\bar{k})}+C_7\alpha(\bar{k})+\sum_{\kappa=1}^{\bar{k}-1}\left[\frac{C_6}{T(\bar{k}-\kappa)}+C_7\alpha(\bar{k}-\kappa)\right]\\
&\quad\times\frac{1}{\Pi_{\varsigma=0}^{\kappa-1}\boldsymbol{\pi}^\top\mathbf{u}\sigma\alpha(\bar{k}-\varsigma)T(\bar{k}-\varsigma)}+\frac{\|\mathbf{x}(1,1)-\hat{\mathbf{x}}^*\|^2}{\Pi_{\varsigma=0}^{\bar{k}-1}\boldsymbol{\pi}^\top\mathbf{u}\sigma\alpha(\bar{k}-\varsigma)T(\bar{k}-\varsigma)}\\
&\le\left(\frac{C_6}{T_0}+\frac{C_7}{4\boldsymbol{\pi}^\top\mathbf{u}\sigma T_0}\right)\sum_{\kappa=1}^{\bar{k}}2^{-(\bar{k}+\kappa)}+4B_{\mathrm{x}}^2 4^{-\bar{k}}\\
&\le\left(\frac{C_6}{T_0}+\frac{C_7}{4\boldsymbol{\pi}^\top\mathbf{u}\sigma T_0}\right)2^{-\bar{k}}+4B_{\mathrm{x}}^2 4^{-\bar{k}} \qquad (30)
\end{aligned}$$

where the second inequality follows from $\alpha(k+1)=\frac{1}{2}\alpha(k)$ and $T(k+1)=2T(k)$ with $\alpha(1)=\frac{1}{\boldsymbol{\pi}^\top\mathbf{u}\sigma T_0}$ and $T(1)=4T_0$ and the last one follows from $\sum_{\kappa=1}^{\bar{k}}2^{-\kappa}<1$. Note that $T=4T_0\sum_{\kappa=1}^{\bar{k}}2^{\kappa-1}$, which implies $2^{-\bar{k}}=\frac{4T_0}{T+4T_0}\le\frac{4T_0}{T}$. Then,

$$n\|\bar{\mathbf{v}}(\bar{k}+1,1)-\mathbf{x}^*\|^2\le\left(4C_6+\frac{C_7}{\boldsymbol{\pi}^\top\mathbf{u}\sigma}\right)\frac{1}{T}+\frac{64B_{\mathbf{x}}^2T_0^2}{T^2}. \qquad (31)$$

Furthermore, referring to (22) with index (k,τ), we get

$$\begin{aligned}
&\|\mathbf{x}(\bar{k}+1,1)-\mathbf{1}_n\otimes\bar{\mathbf{v}}(\bar{k}+1,1)\|^2\\
&\le\left[\frac{2n^2B_{\mathbf{x}}+d_o}{d_oT(\bar{k})}\sum_{\tau=1}^{T(\bar{k})}\|\mathbf{x}(\bar{k},\tau)-\hat{\mathbf{x}}(\bar{k},\tau)\|\right]^2\\
&\le\left[\frac{(2n^2B_{\mathbf{x}}+d_o)\Gamma}{d_o(1-\gamma)}\left(\frac{B_{\mathbf{x}}}{T(\bar{k})}+\bar{L}_g\alpha(k)+\frac{\eta(\bar{k})}{8C_{10}}\right)\right]^2\\
&\le\left[\frac{2(2n^2B_{\mathbf{x}}+d_o)\Gamma}{d_o(1-\gamma)}\left(B_{\mathbf{x}}+\frac{4\bar{L}_g}{\boldsymbol{\pi}^\top\mathbf{u}\sigma}+C_{11}\right)\right]^2\frac{1}{T^2} \qquad (32)
\end{aligned}$$

where $\bar{L}_g=L_g+2\boldsymbol{\pi}^\top\mathbf{u}\sigma C_{10}$ and $C_{11}=\frac{1}{C_{10}}\left(4C_6+\frac{C_7}{\boldsymbol{\pi}^\top\mathbf{u}\sigma}+64B_{\mathbf{x}}^2T_0^2\right)$, the second inequality follows from lemma 4, and the last one is obtained from $\alpha(k)T(k)\boldsymbol{\pi}^\top\mathbf{u}\sigma=4$ and $T(\bar{k})=\frac{T+T(1)}{2}\ge\frac{T}{2}$ because of $T=\sum_{\kappa=1}^{\bar{k}}T(k)$ and $T(k+1)=2T(k)$. Combining (31) and (32),

$$\begin{aligned}
&\|\mathbf{x}(\bar{k}+1,1)-\mathbf{1}_n\otimes\mathbf{x}^*\|^2\\
&\le 2\|\mathbf{x}(\bar{k}+1,1)-\mathbf{1}_n\otimes\bar{\mathbf{v}}(\bar{k}+1,1)\|^2+2n\|\bar{\mathbf{v}}(\bar{k}+1,1)-\mathbf{x}^*\|^2\\
&\le\frac{2C_5}{T^2}+\frac{1}{T}\left[8C_6+\frac{2C_7}{\boldsymbol{\pi}^\top\mathbf{u}\sigma}\right]
\end{aligned}$$

where $C_5=\left[\frac{2(2n^2B_{\mathbf{x}}+d_o)\Gamma}{d_o(1-\gamma)}(B_{\mathbf{x}}+\frac{4\bar{L}_g}{\boldsymbol{\pi}^\top\mathbf{u}\sigma}+C_{11})\right]^2+64B_{\mathrm{x}}^2T_0^2$. $\square$

References

[1] T. Yang et al., "A survey of distributed optimization," *Annu. Rev. Control*, vol. 47, pp. 278–305, 2019.
[2] H. Hong, X. Yu, W. Yu, D. Zhang, and G. Wen, "Distributed convex optimization on state-dependent undirected graphs: Homogeneity technique," *IEEE Trans. Control Netw. Syst.*, vol. 7, no. 1, pp. 42–52, Mar. 2020.
[3] K. Lu, G. Jing, and L. Wang, "Online distributed optimization with strongly pseudoconvex-sum cost functions," *IEEE Trans. Autom. Control*, vol. 65, no. 1, pp. 426–433, Jan. 2020.
[4] D. Yuan, Y. Hong, D. W. Ho, and G. Jiang, "Optimal distributed stochastic mirror descent for strongly convex optimization," *Automatica*, vol. 90, pp. 196–203, 2018.
[5] A. Nedić, A. Olshevsky, and W. Shi, "Achieving geometric convergence for distributed optimization over time-varying graphs," *SIAM J. Optim.*, vol. 27, no. 4, pp. 2597–2633, 2017.
[6] Y. Kajiyama, N. Hayashi, and S. Takai, "Linear convergence of consensus-based quantized optimization for smooth and strongly convex cost functions," *IEEE Trans. Autom. Control*, vol. 66, no. 3, pp. 1254–1261, Mar. 2021.
[7] Q. Lü, H. Li, and D. Xia, "Geometrical convergence rate for distributed optimization with time-varying directed graphs and uncoordinated step-sizes," *Inf. Sci.*, vol. 422, pp. 516–530, 2018.
[8] R. Xin and U. A. Khan, "Distributed heavy-ball: A generalization and acceleration of first-order methods with gradient tracking," *IEEE Trans. Autom. Control*, vol. 65, no. 6, pp. 2627–2633, Jun. 2020.
[9] S. Pu, W. Shi, J. Xu, and A. Nedić, "Push-pull gradient methods for distributed optimization in networks," *IEEE Trans. Autom. Control*, vol. 66, no. 1, pp. 1–16, Jan. 2021.
[10] S. Cheng and S. Liang, "Distributed optimization for multi-agent system over unbalanced graphs with linear convergence rate," *Kybernetika*, vol. 56, no. 3, pp. 559–577, 2020.
[11] S. Pu and A. Nedić, "Distributed stochastic gradient tracking methods," *Math. Program.*, vol. 187, pp. 409–457, 2021.
[12] B. Li, S. Cen, Y. Chen, and Y. Chi, "Communication-efficient distributed optimization in networks with gradient tracking and variance reduction," in *Proc. Int. Conf. Artif. Intell. Statist.*, 2020, pp. 1662–1672.
[13] R. Xin, U. A. Khan, and S. Kar, "Variance-reduced decentralized stochastic optimization with accelerated convergence," *IEEE Trans. Signal Process.*, vol. 68, pp. 6255–6271, 2020.
[14] A. Nedić, A. Ozdaglar, and P. A. Parrilo, "Constrained consensus and optimization in multi-agent networks," *IEEE Trans. Autom. Control*, vol. 55, no. 4, pp. 922–938, Apr. 2010.
[15] J. Lei, H. Chen, and H. Fang, "Primal-dual algorithm for distributed constrained optimization," *Syst. Control Lett.*, vol. 96, pp. 110–117, 2016.
[16] L. Romao, K. Margellos, G. Notarstefano, and A. Papachristodoulou, "Subgradient averaging for multi-agent optimisation with different constraint sets," *Automatica*, vol. 131, 2021, Art. no. 109738.
[17] A. Nedić and A. Olshevsky, "Distributed optimization over time-varying directed graphs," *IEEE Trans. Autom. Control*, vol. 60, no. 3, pp. 601–615, Mar. 2015.
[18] S. A. Alghunaim, E. K. Ryu, K. Yuan, and A. H. Sayed, "Decentralized proximal gradient algorithms with linear convergence rates," *IEEE Trans. Autom. Control*, vol. 66, no. 6, pp. 2787–2794, Jun. 2021.
[19] C. Xi and U. A. Khan, "Distributed subgradient projection algorithm over directed graphs," *IEEE Trans. Autom. Control*, vol. 62, no. 8, pp. 3986–3992, Aug. 2017.
[20] J. Xu, S. Zhu, Y. C. Soh, and L. Xie, "Augmented distributed gradient methods for multi-agent optimization under uncoordinated constant step-sizes," in *Proc. IEEE Conf. Decis. Control*, 2015, pp. 2055–2060.
[21] W. Shi, Q. Ling, G. Wu, and W. Yin, "EXTRA: An exact first-order algorithm for decentralized consensus optimization," *SIAM J. Optim.*, vol. 25, no. 2, pp. 944–966, 2015.
[22] B. Gharesifard and J. Cortés, "Distributed continuous-time convex optimization on weight-balanced digraphs," *IEEE Trans. Autom. Control*, vol. 59, no. 3, pp. 781–786, Mar. 2014.
[23] X. Zeng, P. Yi, and Y. Hong, "Distributed continuous-time algorithm for constrained convex optimizations via nonsmooth analysis approach," *IEEE Trans. Autom. Control*, vol. 62, no. 10, pp. 5227–5233, Oct. 2017.
[24] S. Liang, L. Wang, and G. Yin, "Distributed smooth convex optimization with coupled constraints," *IEEE Trans. Autom. Control*, vol. 65, no. 1, pp. 347–353, Jan. 2020.
[25] S. Liang, L. Wang, and G. Yin, "Dual averaging push for distributed convex optimization over time-varying directed graph," *IEEE Trans. Autom. Control*, vol. 65, no. 4, pp. 1785–1791, Apr. 2020.