KYBERNETIKA — VOLUME *56* (2020), NUMBER 3, PAGES 559–577

DISTRIBUTED OPTIMIZATION FOR MULTI-AGENT SYSTEM OVER UNBALANCED GRAPHS WITH LINEAR CONVERGENCE RATE

Songsong Cheng and Shu Liang

Distributed optimization over unbalanced graphs is an important problem in multi-agent systems. Most of literatures, by introducing some auxiliary variables, utilize the Push-Sum scheme to handle the widespread unbalance graph with row or column stochastic matrix only. But the introduced auxiliary dynamics bring more calculation and communication tasks. In this paper, based on the in-degree and out-degree information of each agent, we propose an innovative distributed optimization algorithm to reduce the calculation and communication complexity of the conventional Push-Sum scheme. Furthermore, with the aid of small gain theory, we prove the linear convergence rate of the proposed algorithm.

Keywords: multi-agent systems, distributed optimization, unbalanced graph, small gain theory, linear convergence rate

Classification: 90C33, 68W15

1. INTRODUCTION

Owing to many applications in multi-agent network systems, such as smart grids [31], virtualized networks [9] and machine learning [16] etc., distributed optimization has gained much research attentions and many distributed algorithms have been developed [11, 13, 22, 33]. These distributed algorithms solve the optimization problem only using local data and exchanging information with neighbors of agents.

Under the circumstance of undirected communication graphs, many excellent methods were proposed, such as subgradient[18], dual average [5], and ADMM [10] and corresponding extensions to some certain constraints [3, 17, 25]. Besides, some scholars considered the distributed optimization problem with more general directed graphs [2, 7, 19], but these graphs are also limited to balanced graphs. Although, for any strongly connected directed graph, some balanced weights can be formulated [8], it may be impossible in some practice or the corresponding computation burden may be larger than that of solving the optimization problem [4, 6].

Recently, for the purpose of overcoming the constraints of unbalanced communication graphs, some advanced approaches have been achieved. By learning the Perron vector,

DOI: 10.14736/kyb-2020-3-0559

[12] proposed the Push-Sum method to adjust the weight scalars to ensure the exact convergence. By combining the Push-Sum scheme, [24] extended the dual average method into the unbalanced circumstances. In order to accelerate the convergence rate with fixed step sizes, some new techniques have come forth. Xi et. al., combining the EXTRA [23] and the Push-Sum scheme, proposed the DEXTRA method based on the out-degree of each agent [26]. Furthermore, Xin et. al., developed FROST by utilizing the in-degree information [29]. With the aid of DIGing techniques, Nedić, et.al. achieved the Push-DIGing to obtain the linear convergence [21]. But almost all of the Push-Sum based methods need to employ one or more auxiliary dynamics, which increases the burden of computation and communication. Another way to overcome the unbalanced constraints is the Surplus-based method [1], which ensures the average consensus by introducing a surplus variable to eliminate the unbalance. By utilizing the surplus based method, Xi et.al. extended the subgradient method with convergence rate $O(\frac{\ln t}{\sqrt{t}})$ [27, 28] and Liang et.al, extended the quasi-monotone subgradient method with a increasing convergence rate as $O(\frac{1}{\sqrt{t}})$ [14]. But the convergence rate of the surplus based method is limited to sublinear because of the necessary of diminishing step sizes.

Motivated by the above discussions, we devote to integrate the Push-DIGing and Surplus-based method to reduce the auxiliary variables and communication burden of Push-DIGing, achieve a linear convergence rate. The main contributions of this paper are listed as follows. Firstly, we consider distributed optimization over unbalanced graphs, which is not an easy problem because the graph can be directed and it is hard to construct a doubly stochastic matrix. Secondly, we present a novel distributed algorithm to solve the problem and the algorithm inherits all merits of the Surplus-based algorithm and Push-DIGing algorithm. In particular, our algorithm can solve not only the average consensus problem as the Surplus-based algorithm but also distributed optimization problem as the Push-DIGing algorithm. Moreover, our algorithm utilizes less variables than the Push-DIGing algorithm to reduce the computation and communication burden.

The rest of this paper is organized as follows. Section 2 presents related preliminaries on basic notations, graph theory and formulates the distributed optimization problem. Second 3 provides the proposed distributed algorithm and analyzes its convergence performance. Section 4 gives an illustrative example and Section 5 concludes this paper.

2. PRELIMINARIES AND PROBLEM FORMULATION

In this section, we introduce necessary preliminaries and formulate the problem.

2.1. Basic notation and notions

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 all of the entries are one with a proper dimension. $I_n \in \mathbb{R}^{n\times n}$ is an identity matrix and $I_m^n = \mathbf{1}_n^\top \otimes I_m \in \mathbb{R}^{m\times nm}$, where $\otimes$ is the Kronecker product operator. $A^\top$ is the transpose of matrix A, diag$\{\boldsymbol{a}\}$ denotes a diagonal matrix composed by the elements of vector $\mathbf{a}$ in the diagonal position. $\mathbf{a}\cdot\mathbf{b}$ is the component wise multiplication of the two vectors $\mathbf{a}$ and $\mathbf{b}$. $A = [a_{ij}] \in \mathbb{R}^{n\times m}$ denotes a $n\times m$ matrix with a_{ij} is the corresponding i row and j column entry. Considering n vector $\mathbf{x}_1, \cdots, \mathbf{x}_n \in \mathbb{R}^m$, the accumulated vector is $\mathbf{x} = [\mathbf{x}_1^\top, \cdots, \mathbf{x}_n^\top]^\top$, the corresponding average is $\bar{\mathbf{x}} = \frac{1}{n} I_m^n \mathbf{x} = \frac{1}{n}\sum_{i=1}^n \mathbf{x}_i$ and the consensus violation among $\mathbf{x}_1, \cdots, \mathbf{x}_n$ is $\tilde{\mathbf{x}} = \mathbf{x} - \mathbf{1}_n \otimes \bar{\mathbf{x}} = \mathcal{L}\mathbf{x}$ with $\mathcal{L} = I_{nm} - \frac{1}{n}\mathbf{1}_n \otimes I_m^n$.

2.2. Graph theory

An unbalanced graph of a multi-agent system is denoted by $\mathcal{G} = (\mathcal{E}, \mathcal{V})$ where $\mathcal{V} = \{1, 2, \cdots, n\}$ and $\mathcal{E} \subseteq \mathcal{V} \times \mathcal{V}$ are the set of nodes and edges. $(i, j) \in \mathcal{E}$ with $i \neq j$ means that node i can sent information to j. $\mathcal{N}_\mathrm{i}^i = \{j \in \mathcal{V} | (j, i) \in \mathcal{E}, i \neq j\}$ is the in-neighbor set of node i and $\mathcal{N}_\mathrm{o}^i = \{j \in \mathcal{V} | (i, j) \in \mathcal{E}, i \neq j\}$ is the out-neighbor set of node i at time t. The corresponding in-degree and out-degree of node i can be formulated as $d_\mathrm{i}^i = |\mathcal{N}_\mathrm{i}^i|$ and $d_\mathrm{o}^i = |\mathcal{N}_\mathrm{o}^i|$, respectively. For the considered unbalanced communication graph, following mild assumption is necessary.

Assumption 1. The considered unbalanced graph is connected.

2.3. Problem formulations

Consider an optimization problem with following global objective function

$$\min_{\mathbf{x}_\mathrm{o} \in \mathbb{R}^m} \quad f(\mathbf{x}_\mathrm{o}) := \sum_{i=1}^n f_i(\mathbf{x}_\mathrm{o}), \tag{1}$$

where the local objective function $f_i(\mathbf{x}_\mathrm{o})$ is differentiable and convex. A multi-agent system is adopt to solve the optimization problem in distributed manner. Problem (1) can be equivalently transformed as follows

$$\begin{aligned} \min_{\mathbf{x} \in \mathbb{R}^{mn}} \quad & f(\mathbf{x}) := \sum_{i=1}^n f_i(\mathbf{x}_i), \\ \text{s.t.} \quad & \mathbf{x}_i = \mathbf{x}_j. \end{aligned} \tag{2}$$

where $\mathbf{x} = [\mathbf{x}_1^\top, \mathbf{x}_2^\top, \cdots, \mathbf{x}_n^\top]^\top$. For the problem in (2), following two basic assumptions should be given.

Assumption 2. Each local objective function $f_i(\mathbf{x})$ has L_i-Lipschitz continuously gradient, namely, for $\forall\ \mathbf{a}, \mathbf{b} \in \mathbb{R}^m$,

$$\|\nabla f_i(\mathbf{a}) - \nabla f_i(\mathbf{b})\| \leq L_i \|\mathbf{a} - \mathbf{b}\|, \tag{3}$$

where L_i is the Lipschitz constant of $f_i(\mathbf{x})$.

Assumption 3. Each local objective function $f_i(\mathbf{x})$ is μ_i-strong convexity, namely, for $\forall\ \mathbf{a}, \mathbf{b} \in \mathbb{R}^n$,

$$f_i(\mathbf{a}) \geq f_i(\mathbf{b}) + \langle \nabla f_i(\mathbf{a}), \mathbf{a} - \mathbf{b} \rangle + \frac{\mu_i}{2}\|\mathbf{a} - \mathbf{b}\|^2, \tag{4}$$

where μ_i is the strong convexity constant of $f_i(\mathbf{x})$.

For the considered objective function with strong convex constants μ_i and Lipschitz constants L_i with $i = 1, \cdots, n$, following lemmas can be introduced.

Lemma 2.1. (Nedić et al. [21]) Under Assumptions 2−3, we have

$$\begin{aligned} \frac{1}{n}\sum_{i=1}^{n}\langle f_i(\mathbf{x}_i^t), \mathbf{x}^* - \bar{\mathbf{x}}^{t+1}\rangle \le f(\mathbf{x}^*) - f(\bar{\mathbf{x}}^{t+1}) - \tfrac{\beta\bar{\mu}}{2(1+\beta)}\|\bar{\mathbf{x}}^t - \mathbf{x}^*\|^2 \\ + \frac{1}{n}\sum_{i=1}^{n}\left[\tfrac{\beta\mu_i}{2} + \tfrac{(1+\eta)L_i}{2\eta}\right]\|\mathbf{x}_i^t - \bar{\mathbf{x}}^t\|^2 + \tfrac{(1+\eta)\bar{L}}{2}\|\bar{\mathbf{x}}^{t+1} - \bar{\mathbf{x}}^t\|^2, \end{aligned} \tag{5}$$

where $\bar{\mu} = \frac{1}{n}\sum_{i=1}^{n}\mu_i$ and $\bar{L} = \frac{1}{n}\sum_{i=1}^{n}L_i$.

2.4. Small gain theory

Following small gain theory is the basic scheme to analyze the convergence property of the proposed algorithm.

Lemma 2.2. (Nedić et al. [21]) For a given sequence $\mathbf{h}_1, \mathbf{h}_2, \cdots, \mathbf{h}_J$, satisfies following closed circle

$$\|\mathbf{h}_{(j\backslash J)+1}\|_\lambda^T \le \varepsilon_j\|\mathbf{h}_j\|_\lambda^T + \theta_j, \tag{6}$$

where all of gains $\varepsilon_1, \varepsilon_2, \cdots, \varepsilon_J$ are positive constants. If $\varepsilon \triangleq \varepsilon_1\varepsilon_2\cdots\varepsilon_J < 1$ holds, we have

$$\|\mathbf{h}_1\|_\lambda \le \tfrac{1}{1-\varepsilon}\sum_{j=1}^{J}\theta_j\prod_{k=j+1}^{J}\varepsilon_k, \tag{7}$$

with definition $\varepsilon_{J+1} = 1$.

3. MAIN RESULTS

In this section, we present the distributed algorithm design and give the convergence analysis.

3.1. Distributed algorithm

Firstly, we consider the distributed optimization problem over time-unvarying unbalanced communication graph, we propose the distributed algorithm as shown in Algorithm 1.

For the convenience of notations and subsequent convergence analysis, we define $\nabla f(\mathbf{x}^t) = [\nabla f_1^\top(\mathbf{x}^t), \nabla f_2^\top(\mathbf{x}^t), \cdots, \nabla f_n^\top(\mathbf{x}^t)]^\top$ and formulate a row stochastic matrix Φ and a column stochastic matrix Ψ as

$$\begin{cases} \Phi = [\phi_{ij}] \in \mathbb{R}^{n\times n}, \\ \Psi = [\psi_{ij}] \in \mathbb{R}^{n\times n}, \end{cases} \tag{8}$$

Algorithm 1

Initialization: For each $i \in \mathcal{V}$,

$\rho \in (0,1),\ \mathbf{x}_i^0 \in \mathbb{R}^m,\ \mathbf{g}_i^0 = -\nabla f_i(\mathbf{x}_i^0) \in \mathbb{R}^m.$

Update flows: For each $i \in \mathcal{V}$,

$$\mathbf{x}_i^{t+1} = (1-\rho)\mathbf{x}_i^t + \tfrac{\rho}{d_{\mathrm{i}}^i} \sum_{j\in\mathcal{N}_{\mathrm{i}}^i} \mathbf{x}_j^t + \alpha \mathbf{g}_i^t, \tag{S1}$$

$$\mathbf{g}_i^{t+1} = \tfrac{1}{d_{\mathrm{o}}^i+1} \sum_{j\in\mathcal{N}_{\mathrm{o}}^i \bigcup i} \mathbf{g}_j^t - \nabla f_i(\mathbf{x}_i^{t+1}) + \nabla f_i(\mathbf{x}_i^t). \tag{S2}$$

where

$$\phi_{ij} = \begin{cases} \frac{1}{d_{\mathrm{i}}^i}\rho & j = i, \\ 1-\rho & j \in \mathcal{N}_{\mathrm{i}}^i, \\ 0 & \text{otherwise,} \end{cases} \quad \text{and} \quad \psi_{ij} = \begin{cases} \frac{1}{d_{\mathrm{o}}^i+1} & j \in \mathcal{N}_{\mathrm{o}}^i \cup i, \\ 0 & \text{otherwise.} \end{cases} \tag{9}$$

According to above notations in (8)-(9), the proposed Algorithm 1 can be formulated in a compact form

$$\begin{cases} \mathbf{x}^{t+1} = (\Phi \otimes I_m)\mathbf{x}^t + \alpha \mathbf{g}^t, & (10a) \\ \mathbf{g}^{t+1} = (\Psi \otimes I_m)\mathbf{g}^t - \nabla f(\mathbf{x}^{t+1}) + \nabla f(\mathbf{x}^t). & (10b) \end{cases}$$

For the considered column stochastic matrix Ψ, we have following result.

Lemma 3.1. For any column stochastic matrix Ψ, we can obtain a row stochastic matrix sequence as follows

$$\hat{\Psi}^t = (V^t)^{-1}(\Psi \otimes I_m)V^{t-1} \tag{11}$$

where V^t is governed by $V^t = \mathrm{diag}\{\mathbf{v}^t\} \otimes I_m$ and $\mathbf{v}^{t+1} = (\Psi \otimes I_m)\mathbf{v}^t$ with $\mathbf{v}^0 = \mathbf{1}_{nm}$

The proof of Lemma 3.1 can be obtained directly by extending Lemma 4 in [20] with the Kronecker product.

Lemma 3.2. Under Assumptions 1−3 with $Q = 1$, following properties on objective function's gradient hold

$$\begin{cases} \left\| \frac{1}{n} \sum_{i=1}^{n} \nabla f_i(\mathbf{x}_i^t) \right\| \leq \frac{1}{n} L_{\mathrm{m}} \|\mathbf{p}^t\|, & (12a) \\ \mathbf{g}^t = -\sum_{k=0}^{t} (\Psi^{t-k} \otimes I_m)\mathbf{y}^k, & (12b) \\ \|\mathbf{g}^t\| \leq L. & (12c) \end{cases}$$

where $\mathbf{y}^0 = -\mathbf{g}^0$, $L_{\mathrm{m}} = \max\{L_1, \cdots, L_n\}$ and $L = \frac{2mn^2\xi L_{\mathrm{m}}}{1-\zeta} + L_{\mathrm{m}}$.

Lemma 3.3. Under Assumption 1 with $Q = 1$, for the introduced V^t in Lemma 3.5 and Ψ, following properties hold

$$\begin{cases} \|I_m^n[(V^t)^{-1} - (V^*)^{-1}]\| \le c_1\gamma^t, & \text{(13a)} \\ \|\Psi^t \otimes I_m - \Psi^* \otimes I_m\| \le c_2\gamma^t, & \text{(13b)} \\ I_m^n(V^*)^{-1}(\Psi^* \otimes I_m) = I_m^n, & \text{(13c)} \end{cases}$$

where $c_1 = \frac{2(2-\gamma)\sqrt{nm}}{(1-\gamma)v_{\mathrm{m}}}$, $c_2 = \frac{2(2-\gamma)\sqrt{nm}}{1-\gamma}$ and $\gamma = \max\left\{\frac{d_{\mathrm{o}}^1}{d_{\mathrm{o}}^1+1}, \frac{d_{\mathrm{o}}^2}{d_{\mathrm{o}}^2+1}, \cdots, \frac{d_{\mathrm{o}}^n}{d_{\mathrm{o}}^n+1}\right\}$, with $v_{\mathrm{m}} = \min\left\{v_i^t v_i^*, i \in \{1, 2, \cdots, nm\}, t \in \mathbb{N}^+\right\}$.

3.2. Connection with other methods

For the balanced graphs, by combining a gradient tracking and inexact gradient method, [21] proposed the DIGing method with double stochastic weight matrix W as follows,

$$\begin{cases} \mathbf{x}^{t+1} = (W \otimes I_n)\mathbf{x}^t + \alpha\mathbf{g}^t \\ \mathbf{g}^{t+1} = (W \otimes I_n)\mathbf{g}^t - \nabla f(\mathbf{x}^{t+1}) + \nabla f(\mathbf{x}^t). \end{cases} \tag{14}$$

And then, the method was extended to following Push-DIGing method for the unbalanced graphs with the aid of two auxiliary variables $\mathbf{v}^t$ and $\mathbf{u}^t$

$$\begin{cases} \mathbf{v}^{t+1} = (\Psi \otimes I_n)\mathbf{v}^t \\ \mathbf{u}^{t+1} = (\Psi \otimes I_n)(\mathbf{x}^t + \alpha\mathbf{g}^t) \\ \mathbf{x}^{t+1} = (\mathrm{diag}\{\mathbf{v}^{t+1}\})^{-1}\mathbf{u}^{t+1} \\ \mathbf{g}^{t+1} = (\Psi \otimes I_n)\mathbf{g}^t - \nabla f(\mathbf{x}^{t+1}) + \nabla f(\mathbf{x}^t). \end{cases} \tag{15}$$

Apparently, the condition on the double stochastic weight matrix W in (14) is weakened as the column stochastic matrix Ψ in (15). However, the introduced two auxiliary variables $\mathbf{v}^t$ and $\mathbf{u}^t$ increase the computation and communication burdens of the multi-agent system. In the proposed algorithm, we remove the two auxiliary variables by introducing the cheap in-degree and out-degree information of each-agent.

Moreover, consider the following average problem

$$\lim_{t\to\infty} \mathbf{x}_i^t = \mathbf{x}^* = \tfrac{1}{n}\sum_{i=1}^{n} \mathbf{x}_i^0. \tag{16}$$

In order to deal with the problem in (16) under unbalanced graphs, [1] proposed the Surplus-based method as follows

$$\begin{bmatrix} \mathbf{x}^{t+1} \\ \mathbf{s}^{t+1} \end{bmatrix} = \left(\begin{bmatrix} \Phi & \epsilon I_n \\ I_n - \Phi & \Psi - \epsilon I_n \end{bmatrix} \otimes I_m \right) \begin{bmatrix} \mathbf{x}^t \\ \mathbf{s}^t \end{bmatrix} \tag{17}$$

where $\mathbf{s}^t$ serves as the so-called surplus variable.

Transform the average consensus problem (16) to be an optimization problem as follows, whose optimal solution is the average consensus value,

$$\min_{\mathbf{x}_i \in \mathbb{R}^m} \tfrac{1}{2}\sum_{i=1}^{n} \|\mathbf{x}_i - \mathbf{x}_i^0\|. \tag{18}$$

By applying the proposed algorithm in (12) into the problem in (18), and inserting (10a) into (10b), it yields

$$\begin{bmatrix} \mathbf{x}^{t+1} \\ \mathbf{g}^{t+1} \end{bmatrix} = \left(\begin{bmatrix} \Phi & \alpha I_n \\ I_n - \Phi & \Psi - \alpha I_n \end{bmatrix} \otimes I_m \right) \begin{bmatrix} \mathbf{x}^{t} \\ \mathbf{g}^{t} \end{bmatrix} \tag{19}$$

which is identical to the Surplus-based algorithm in (17) with $\epsilon = \alpha$ and $\mathbf{s}^t = \mathbf{g}^t$.

3.3. Convergence analysis

The convergence analysis can be organized based on small gain theory. Some basic notations should be defined as follows

$$\begin{cases} \mathbf{p}^t = \mathbf{x}^t - \mathbf{1} \otimes \mathbf{x}^*, \\ \mathbf{y}^t = \nabla f(\mathbf{x}^t) - \nabla f(\mathbf{x}^{t-1}), \\ \tilde{\mathbf{g}}^t = \left(I_{nm} - \frac{1}{n}\mathbf{1}_n \otimes I_m^n\right)\mathbf{g}^t, \\ \tilde{\mathbf{x}}^t = \left(I_{nm} - \frac{1}{n}\mathbf{1}_n \otimes I_m^n\right)\mathbf{x}^t. \end{cases} \tag{20}$$

1) The first arrow $\mathbf{p} \to \mathbf{y}$.

Lemma 3.4. (Nedić et al. [21]) For $\mathbf{p}$ and $\mathbf{y}$, if Assumption 2 holds, we have

$$\|\mathbf{y}\|_\lambda^T \leq \varepsilon_1 \|\mathbf{p}\|_\lambda^T, \tag{21}$$

where $T \in \mathbb{N}^+, 0 < \lambda < 1$ and $\varepsilon_1 = \frac{(\lambda+1)L_{\mathrm{m}}}{\lambda}$.

2) The second arrow $\mathbf{y} \to \tilde{\mathbf{w}}$.

For the second arrow, because Ψ is column stochastic, we do not deduce the arrow between $\mathbf{y}$ and $\tilde{\mathbf{g}}$ directly, but the $\tilde{\mathbf{w}} = \mathcal{L}\mathbf{w}$ instead, where $\mathbf{w}^t = (V^t)^{-1}\mathbf{g}^t$. The relationship between $\mathbf{y}$ and $\tilde{\mathbf{w}}$ can be achieved as follows.

Lemma 3.5. For $\mathbf{y}$ and $\tilde{\mathbf{w}}$, if $\sup\{\|\hat{\Psi}\|_{\mathcal{L}}\} < \lambda$ and $0 < \sup\{\|\hat{\Psi}\|_{\mathcal{L}}\} < \lambda$ holds, we have

$$\|\tilde{\mathbf{w}}\|_\lambda^T \leq \varepsilon_2 \|\mathbf{y}\|_\lambda^T, \tag{22}$$

where $T \in \mathbb{N}^+, 0 < \lambda < 1$ and $\varepsilon_2 = \frac{\lambda\delta_1}{\lambda-\delta_2}$ with $\delta_1 = \sup\{\|(V^{t+1})^{-1}\|_{\mathcal{L}}\}$ and $\delta_2 = \sup\{\|\hat{\Psi}\|_{\mathcal{L}}\}$.

3) The third arrow $\tilde{\mathbf{w}} \to \tilde{\mathbf{x}}$.

Lemma 3.6. For $\tilde{\mathbf{x}}$, $\tilde{\mathbf{w}}$ and $\mathbf{p}$, if $\sup\{\|\Phi \otimes I_m\|_{\mathcal{L}}\} \leq \lambda$ and $\gamma < \lambda$ hold, we have

$$\|\tilde{\mathbf{x}}\|_\lambda^T \leq \alpha\tau_1 \|\tilde{\mathbf{w}}\|_\lambda^T + \alpha\tau_2 \|\mathbf{p}\|_\lambda^T + \tau_3, \tag{23}$$

where $T \in \mathbb{N}^+, 0 < \lambda < 1$, $\tau_1 = \frac{\delta_4}{\lambda-\delta_3}$, $\tau_2 = \frac{(c_2c_4+1)(\lambda+1)}{\lambda-\gamma}\frac{L_{\mathrm{m}}\tau_1}{\sqrt{n}}$ and $\tau_3 = \frac{\alpha c_1 c_3 \delta_4}{\sqrt{n}(\lambda-\delta_3)}$ with $\delta_3 = \sup\{\|\Phi \otimes I_n\|_{\mathcal{L}}\}$ and $\delta_4 = \sup\{\|\mathbf{v}^t\|_{\mathcal{L}}\}$.

Remark 3.7. Lemma 3.6 shows the relation between $\|\tilde{\mathbf{x}}\|_\lambda^T$ and $\|\tilde{\mathbf{w}}\|_\lambda^T$ but perturbed by the third part $\|\mathbf{p}\|_\lambda^T$. That is to say, it is not the pure arrow $\tilde{\mathbf{w}} \to \tilde{\mathbf{x}}$. Fortunately, we present the relationship between the $\|\tilde{\mathbf{x}}\|_\lambda^T$ and $\|\mathbf{p}\|_\lambda^T$ in Lemma 3.9. Therefore, based on Lemma 3.9, Lemma 3.8 will be exhibited the pure arrow $\tilde{\mathbf{w}} \to \tilde{\mathbf{x}}$.

Lemma 3.8. For $\tilde{\mathbf{x}}$ and $\tilde{\mathbf{w}}$, if $\sup\{\|\Phi \otimes I_n\|_{\mathcal{L}}\} \leq \lambda$ holds, we have

$$\|\tilde{\mathbf{x}}\|_\lambda^T \leq \varepsilon_3 \|\tilde{\mathbf{w}}\|_\lambda^T + \theta_3, \tag{24}$$

where $T \in \mathbb{N}^+, 0 < \lambda < 1$, $\varepsilon_3 = \frac{\alpha\tau_1}{1-\alpha\tau_2\varepsilon_4}$ and $\theta_3 = \frac{\alpha\tau_2\theta_4+\tau_3}{1-\alpha\tau_2\varepsilon_4}$.

4) The last arrow $\tilde{\mathbf{x}} \to \mathbf{p}$.

Lemma 3.9. For $\mathbf{p}$ and $\tilde{\mathbf{x}}$, if $1-\alpha(1+\eta)\bar{L} \geq 0$, $\kappa_2\lambda \leq \sqrt{n}$, and $\kappa_1 \leq \lambda < 1$ hold with certain $\xi \in (0,1)$, $\eta \in \mathbb{R}$, and $\beta \in \mathbb{R}$, we have

$$\|\mathbf{p}\|_\lambda^T \leq \varepsilon_4 \|\tilde{\mathbf{x}}\|_\lambda^T + \theta_4, \tag{25}$$

where $T \in \mathbb{N}^+, 0 < \lambda < 1$, $\varepsilon_4 = \frac{1+\kappa_3\lambda}{1-\kappa_2\lambda\sqrt{n}^{-1}}$, and $\theta_4 = \frac{\sqrt{n}}{1-\kappa_2\lambda\sqrt{n}^{-1}}\|\bar{\mathbf{x}}^0 - \mathbf{x}^*\|$ with $\kappa_1^2 = \frac{1}{1-\xi}\left[1-\frac{\alpha\beta\bar{\mu}}{(1+\beta)}\right]$, $\kappa_2^2 = \frac{1}{\alpha\bar{\mu}\xi}\|\mathbf{k}\|^2$, and $\kappa_3^2 = \frac{L_{\mathrm{m}}(1+\eta)+\beta\mu_{\mathrm{m}}n\eta}{\bar{\mu}\eta}$.

Remark 3.10. As shown in the second and third arrows of the small gain theory, we introduce the variable $\mathbf{w}^t$ to replace the dynamics of $\mathbf{g}^t$ for the convenient of formulating the circle of the small gain. However, the auxiliary variable $\mathbf{w}^t$ just need the existence rather than need to be computed and let alone to be communicated with other agents.

Based on these basic Lemmas 3.4, 3.5, 3.8 and 3.9, we are ready to develop the main result, which establishes the convergence rate and the corresponding valid step size.

Theorem 3.11. Under Assumptions 1−3 with $Q = 1$, if $\forall\ \zeta \in (0,1)$ and $\xi \in \big(0, \max\{1, \frac{(1-\bar{\delta})^2}{\sigma}\}\big)$ hold, $\mathbf{x}^t$ converges to $\mathbf{x}^*$ with linear rate $O(\lambda^t)$, and we can estimate λ as

$$\lambda = \begin{cases} \sqrt{\frac{3-2\alpha\bar{\mu}}{3(1-\xi)}}, & 0 < \alpha \leq \frac{1.5(\lambda^o-\bar{\delta})^2}{\bar{\mu}\sigma}, \\ \bar{\delta} + \sqrt{\frac{2\alpha\bar{\mu}\sigma}{3}}, & \frac{1.5(\lambda^o-\bar{\delta})^2}{\bar{\mu}\sigma} < \alpha \leq \frac{1.5(1-\bar{\delta})^2}{\bar{\mu}\sigma}, \end{cases} \tag{26}$$

where $\bar{\delta} = \max\{\delta_2, \delta_3, \gamma\}$, $\lambda^o = \frac{\bar{\delta}+\sqrt{[1-(1-\xi)\bar{\delta}^2]+(1-\xi)\sigma^2}}{1+(1-\xi)\sigma}$ and $\sigma = \frac{3\delta_4 L_{\mathrm{m}}(1+\kappa_3)[\sqrt{n}\delta_1+(c_2c_4+1)]}{\zeta\bar{\mu}}$.

Proof. Combining Lemmas 3.4, 3.5, 3.6 and 3.9, we have

$$\begin{cases} \|\mathbf{y}\|_\lambda^T \leq \varepsilon_1 \|\mathbf{p}\|_\lambda^T \\ \|\tilde{\mathbf{w}}\|_\lambda^T \leq \varepsilon_2 \|\mathbf{y}\|_\lambda^T \\ \|\tilde{\mathbf{x}}\|_\lambda^T \leq \varepsilon_3 \|\tilde{\mathbf{w}}\|_\lambda^T + \theta_3 \\ \|\mathbf{p}\|_\lambda^T \leq \varepsilon_4 \|\tilde{\mathbf{x}}\|_\lambda^T + \theta_4, \end{cases} \tag{27}$$

where $\varepsilon_1 = \frac{\lambda+1}{\lambda}L_{\mathrm{m}}$, $\varepsilon_2 = \frac{\lambda\delta_1}{\lambda-\delta_2}$, $\varepsilon_3 = \frac{\alpha\tau_1}{1-\alpha\tau_2\varepsilon_4}$, $\varepsilon_4 = \frac{1+\kappa_3\lambda}{1-\kappa_2\lambda\sqrt{n}^{-1}}$, $\theta_3 = \frac{\alpha\tau_2\theta_4+\tau_3}{1-\alpha\tau_2\varepsilon_4}$ and $\theta_4 = \frac{\sqrt{n}}{1-\kappa_2\lambda\sqrt{n}^{-1}}\|\bar{\mathbf{x}}^0 - \mathbf{x}^*\|$ with $\tau_1 = \frac{\delta_4}{\lambda-\delta_3}$, $\tau_2 = \frac{(c_2c_4+1)(\lambda+1)}{\lambda-\gamma}\frac{L_{\mathrm{m}}\tau_1}{\sqrt{n}}$ and $\tau_3 = \frac{\alpha c_1 c_3\delta_4}{\sqrt{n}(\lambda-\delta_3)}$.

For the purpose of ensuring the convergence of the proposed algorithm, according to the small gain in Lemma 2.2, we have

$$\varepsilon_1\varepsilon_2\varepsilon_3\varepsilon_4 \leq 1. \tag{28}$$

Submitting some related parameters into (28), we have

$$\alpha \leq \frac{\sqrt{n}(\lambda-\delta_2)(\lambda-\delta_3)(\lambda-\gamma)}{\varepsilon_4\delta_4 L_{\mathrm{m}}\left[\sqrt{n}\delta_1(\lambda-\gamma)+(c_2c_4+1)(\lambda-\delta_2)\right](\lambda+1)}. \tag{29}$$

Defining $\overline{\delta} = \max\{\delta_2, \delta_3, \gamma\}$ and $\underline{\delta} = \min\{\delta_2, \delta_3, \gamma\}$, it follows

$$[\sqrt{n}\delta_1(\lambda-\gamma) + (c_2c_4+1)(\lambda-\delta_2)](\lambda+1) \leq 2[\sqrt{n}\delta_1 + (c_2c_4+1)](\lambda-\underline{\delta}). \tag{30}$$

Furthermore, considering all of the conditions on step size α in Lemmas 3.4 − 3.9

$$\begin{cases} \delta_2 = \sup\{\|\hat{\Psi}\|_{\mathcal{L}}\} < \lambda, & \text{(31a)}\\ 1-\alpha(1+\eta)\bar{L} \geq 0, & \text{(31b)}\\ \kappa_1^2 = \frac{1}{1-\xi}\left[1 - \frac{\alpha\beta\bar{\mu}}{(1+\beta)}\right] \leq \lambda^2, & \text{(31c)}\\ \kappa_2\lambda = \frac{\lambda\|\mathbf{k}\|}{\sqrt{\alpha\bar{\mu}\xi}} \leq \sqrt{n}. & \text{(31d)} \end{cases}$$

For (31d), if we choose $\|\mathbf{k}\| = \frac{(1-\zeta)\sqrt{\alpha\xi n\bar{\mu}}}{\lambda}$ with $\zeta \in (0,1)$, we have $1-\kappa_2\lambda\sqrt{n}^{-1} = \zeta$. Then

$$\varepsilon_4 = \frac{1+\kappa_3\lambda}{1-\kappa_2\lambda\sqrt{n}^{-1}} \leq \frac{1+\kappa_3}{\zeta}. \tag{32}$$

Therefore, combining (30) and (32) the up bound of α in (29) can be decreased as

$$\alpha \leq \frac{\zeta(\lambda-\overline{\delta})^2}{2\delta_4 L_{\mathrm{m}}(1+\kappa_3)[\sqrt{n}\delta_1+(c_2c_4+1)]} \triangleq \frac{1.5(\lambda-\overline{\delta})^2}{\bar{\mu}\sigma} \triangleq U_\alpha, \tag{33}$$

where $\sigma = \frac{3\delta_4 L_{\mathrm{m}}(1+\kappa_3)[\sqrt{n}\delta_1+(c_2c_4+1)]}{\zeta\bar{\mu}}$ and U_α means the upper bound of α.

By choosing $\eta = 1$ and $\beta = \frac{2L_{\mathrm{m}}}{\mu_{\mathrm{m}}} \geq 2$, standing on (31c), we have

$$\alpha \geq \frac{1.5[1-(1-\xi)\lambda^2]}{\bar{\mu}} \triangleq L_\alpha \tag{34}$$

L_α means the lower bound of α.

Considering λ increasing from $\overline{\delta}$ to 1, U_α increasing from 0 to $\frac{1.5(1-\overline{\delta})^2}{\bar{\mu}\sigma}$ and L_α decreasing from $\frac{1.5[1-(1-\xi)\overline{\delta}^2]}{\bar{\mu}}$ to $\frac{1.5\xi}{\bar{\mu}}$. Therefore, for the purpose of ensuring (33) and (34) are compatible, we need

$$\xi \leq \frac{(1-\overline{\delta})^2}{\sigma}, \tag{35}$$

and

$$\lambda^o = \frac{\overline{\delta}+\sqrt{[1-(1-\xi)\overline{\delta}^2]+(1-\xi)\sigma^2}}{1+(1-\xi)\sigma}. \tag{36}$$

Therefore, if the step size chosen as

$$0 < \alpha \leq \frac{1.5(\lambda^o-\overline{\delta})^2}{\bar{\mu}\sigma}, \tag{37}$$

we have

$$\lambda = \sqrt{\frac{3-2\alpha\bar{\mu}}{3(1-\xi)}}, \tag{38}$$

and if the step size chosen as

$$\frac{1.5(\lambda^o-\overline{\delta})^2}{\bar{\mu}\sigma} < \alpha \leq \frac{1.5(1-\overline{\delta})^2}{\bar{\mu}\sigma}, \tag{39}$$

we have

$$\lambda = \overline{\delta} + \sqrt{\frac{2\alpha\bar{\mu}\sigma}{3}}. \tag{40}$$

Therefore, the proof is completed. □

Remark 3.12. We became aware of a recent work to apply the algorithm to solve linear algebraic function $A\mathbf{x} = \mathbf{b}$ in [30] and gave the up bound of the step size under the double stochastic weight matrix. However, as pointed out in the Remark 7 of [30], it is challenge to obtain the up bound of step size in the unbalanced graphs circumstance because of the mix row stochastic and column stochastic matrices. In this paper, for the general strongly convex objective function, the valid step size range is achieved for the unbalanced graphs.

4. NUMERICAL EXAMPLE

In order to illustrate the effectiveness of the proposed algorithm, we give a numerical example on distributed solving algebraic equation $A\mathbf{x} = \mathbf{b}$ with $A = [A_1^\top, \cdots, A_n^\top] \in \mathbb{R}^{5\times 5}$, and $\mathbf{b} = [b_1, \cdots, b_5]^\top \in \mathbb{R}^5$. The multi-agent system is connected by unbalanced communication graph (See Fig.1) and the ith agent has access the information of A_i, b_i,

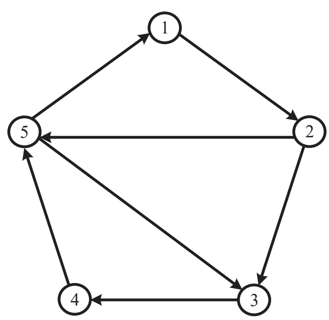

Fig. 1. The communication graph.

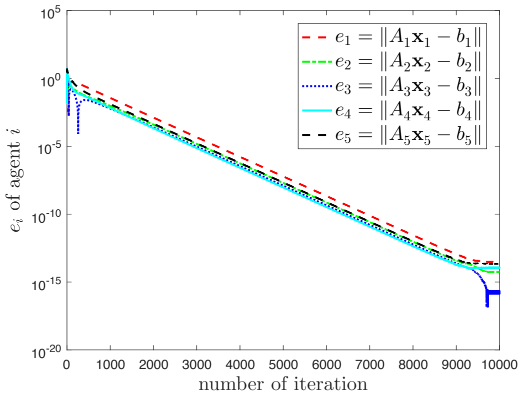

Fig. 2. The convergence result.

$\mathbf{x}_j$, and $\mathbf{g}_j$ with $j \in \mathcal{N}_i \cup i$. The entries of A and $\mathbf{b}$ are randomly generated according to $N(0, 5)$. As shown in Fig.2, the proposed algorithm can obtain the optimal solution with a linear convergence rate.

5. CONCLUSIONS

In this paper, a novel distributed optimization algorithm were proposed to handle unbalanced communication graphs. The conventional and universal push scheme was replaced by utilizing the in-degree and out-degree information, which can reduce the complexity of the algorithm and communication burden of the multi-agent system extensively. For a given strongly convex objective function, it was proved that the proposed algorithm converges to the optimal solution with a linear convergence rate. In the future, we will concentrate on extending the proposed method to the online distributed optimization [32] and remove the assumptions on strong convexity of the objective functions [15].

APPENDIX

Some related proofs of these intermediate lemmas are presented in this section.

Proof of Lemma 3.2

Considering the equilibrium of the proposed algorithm, we have

$$\frac{1}{n}\sum_{i=1}^{n} \nabla f_i(\mathbf{x}^*) = 0. \tag{41}$$

Therefore, we obtain

$$\begin{aligned} \Big\|\frac{1}{n}\sum_{i=1}^{n} \nabla f_i(\mathbf{x}^*)\Big\| &= \Big\|\frac{1}{n}\sum_{i=1}^{n} [\nabla f_i(\mathbf{x}_i^t) - \nabla f_i(\mathbf{x}^*)]\Big\| \\ &\leq \frac{1}{n}\sum_{i=1}^{n} \|\nabla f_i(\mathbf{x}_i^t) - \nabla f_i(\mathbf{x}^*)\| \\ &\leq \frac{1}{n} L_{\mathrm{m}}\|\mathbf{p}^t\|, \end{aligned} \tag{42}$$

where the last inequality is deduced by the Lipschitz gradient of each $f_i(\mathbf{x}_i^t)$ and $\mathbf{p}^t = \mathbf{x}^t - \mathbf{1} \otimes \mathbf{x}^*$.

Besides, According to the definition of $\mathbf{y}^t$ in (20) and the updating of $\mathbf{g}^t$ in (10b), we have

$$\mathbf{g}^t = (\Psi \otimes I_m)\mathbf{g}^{t-1} - \mathbf{y}^t. \tag{43}$$

Applying the recursive of (43), (12b) followed.

In addition, submitting $\mathbf{y}^t = \nabla f(\mathbf{x}^t) - \nabla f(\mathbf{x}^{t-1})$ into (12b), yields,

$$\begin{aligned} \mathbf{g}^t &= -\sum_{k=0}^{t} (\Psi^{t-k} \otimes I_m)[\nabla f(\mathbf{x}^t) - \nabla f(\mathbf{x}^{t-1})] \\ &= \sum_{k=0}^{t-1} \big[(\Psi^{t-k-1} - \Psi^{t-k}) \otimes I_m\big]\nabla f(\mathbf{x}^k) - \nabla f(\mathbf{x}^t). \end{aligned} \tag{44}$$

Therefore, we have

$$\begin{aligned} \|\mathbf{g}^t\| &\leq mL_m \sum_{k=0}^{t-1} \|\Psi^k - \Psi^{k+1}\| + L_m \\ &\leq 2mn^2\xi L_m \sum_{k=0}^{t-1} \xi^k + L_m \\ &\leq \tfrac{2mn^2\xi L_m}{1-\zeta} + L_m. \end{aligned} \tag{45}$$

which implies (12c). □

Proof of Lemma 3.3

Based on Lemma 3 in [18], we have

$$|\Psi_{ij}^t - \Psi_{ij}^*| \leq \tfrac{2(2-\gamma)}{1-\gamma}\gamma^t. \tag{46}$$

Therefore, (13b) can be obtained with $c_2 = \frac{2(2-\gamma)\sqrt{nm}}{1-\gamma}$.

Since $\mathbf{v}^t$ is governed by $\mathbf{v}^{t+1} = (\Psi \otimes I_m)\mathbf{v}^t$ with $\mathbf{v}^0 = \mathbf{1}_{nm}$, we consider the corresponding equilibrium

$$\mathbf{v}^* = (\Psi^* \otimes I_m)\mathbf{v}^0 = n\boldsymbol{\psi}^* \otimes \mathbf{1}_m. \tag{47}$$

where each column vector of Ψ^* is $\boldsymbol{\psi}^*$. Then, we can establish $(V^*)^{-1}$ as

$$(V^*)^{-1} = \tfrac{1}{n}(\mathrm{diag}\{\boldsymbol{\psi}^*\})^{-1} \otimes I_m. \tag{48}$$

Therefore, we have

$$\begin{aligned} I_m^n(V^*)^{-1}(\Psi^* \otimes I_m) &= (\mathbf{1}_n^\top \otimes I_m)[\tfrac{1}{n}(\mathrm{diag}\{\boldsymbol{\psi}^*\})^{-1} \otimes I_m](\Psi^* \otimes I_m) \\ &= \tfrac{1}{n}[\mathbf{1}_n^\top(\mathrm{diag}\{\boldsymbol{\psi}^*\})^{-1}\boldsymbol{\psi}^*\mathbf{1}^\top] \otimes I_m \\ &= I_m^n. \end{aligned} \tag{49}$$

Then, (13c) can be formulated.

Since $\mathbf{v}^t = (\Psi^t \otimes I_m)\mathbf{v}^0 = (\Psi^t \otimes I_m)\mathbf{1}_{nm}$ and $\mathbf{v}^* = (\Psi^* \otimes I_m)\mathbf{1}_{nm}$, consider $\mathbf{v}_i$, which is the ith sub-vector of $\mathbf{v}^t$

$$(v_i^t - v_i^*)\mathbf{1}_m = \sum_{j=1}^{n} (\Psi_{ij}^t - \Psi_{ij}^*)\mathbf{1}_m. \tag{50}$$

Furthermore, we have

$$\left(\tfrac{1}{v_i^t} - \tfrac{1}{v_i^*}\right)\mathbf{1}_m = \tfrac{1}{\mathbf{v}_i^t \mathbf{v}_i^*} \sum_{j=1}^{n} (\Psi_{ij}^t - \Psi_{ij}^*)\mathbf{1}_m. \tag{51}$$

Therefore, standing on (46) and (51), we achieve

$$\|I_m^n[(V^t)^{-1} - (V^*)^{-1}]\| = \left\| \left[\tfrac{1}{v_1^t} - \tfrac{1}{v_1^*}, \cdots, \tfrac{1}{v_n^t} - \tfrac{1}{v_n^*}\right] \mathbf{1}_m \otimes I_m \right\| \leq \tfrac{2(2-\gamma)\sqrt{nm}}{(1-\gamma)v_{\mathrm{m}}}\gamma^t. \tag{52}$$

Then (13a) can be obtained and the proof of Lemma 3.3 is completed. □

Proof of Lemma 3.5

According to Lemma 3.1, $\hat{\Psi} = (V^{t+1})^{-1}(\Psi \otimes I_m)V^t$ is a row stochastic matrix. Therefore, by multiplying matrix $(V^{t+1})^{-1}$ at both two sides of (10b), yields

$$\mathbf{w}^{t+1} = \hat{\Psi}\mathbf{w}^t - (V^{t+1})^{-1}\mathbf{y}^{t+1}, \tag{53}$$

where $\mathbf{w}^t = (V^t)^{-1}\mathbf{g}^t$. Then, we have

$$\|\tilde{\mathbf{w}}^{t+1}\| = \|\mathbf{w}^{t+1}\|_{\mathcal{L}} \leq \|\hat{\Psi}\mathbf{w}^t\|_{\mathcal{L}} + \|(V^{t+1})^{-1}\mathbf{y}^{t+1}\|_{\mathcal{L}} \leq \|\mathcal{L}\hat{\Psi}\mathbf{w}^t\| + \delta_1\|\mathbf{y}^{t+1}\|, \tag{54}$$

where $\delta_1 = \sup\{\|(V^{t+1})^{-1}\|_{\mathcal{L}}\}$. Since $\hat{\Psi}$ is a row stochastic matrix, it is easy to derive

$$\mathcal{L}\hat{\Psi} = \mathcal{L}\hat{\Psi}\mathcal{L}. \tag{55}$$

Submitting (55) into (54), we have

$$\|\tilde{\mathbf{w}}^{t+1}\| \leq \delta_2\|\tilde{\mathbf{w}}^t\| + \delta_1\|\mathbf{y}^{t+1}\|, \tag{56}$$

where $\delta_2 = \sup\{\|\hat{\Psi}\|_{\mathcal{L}}\}$. Considering $\frac{1}{\lambda^{t+1}}\|\tilde{\mathbf{w}}^{t+1}\|$, we obtain

$$\tfrac{1}{\lambda^{t+1}}\|\tilde{\mathbf{w}}^{t+1}\| \leq \tfrac{\delta_2}{\lambda}\tfrac{1}{\lambda^t}\|\tilde{\mathbf{w}}^t\| + \delta_1\tfrac{1}{\lambda^{t+1}}\|\mathbf{y}^{t+1}\|. \tag{57}$$

Taking maximization on the two sides of (57) from $t = 0$ to $t = T - 1$, it yields

$$\|\tilde{\mathbf{w}}\|_{\lambda}^{T} \leq \tfrac{\delta_2}{\lambda}\|\tilde{\mathbf{w}}\|_{\lambda}^{T-1} + \delta_1\|\mathbf{y}\|_{\lambda}^{T} \leq \tfrac{\delta_2}{\lambda}\|\tilde{\mathbf{w}}\|_{\lambda}^{T} + \delta_1\|\mathbf{y}\|_{\lambda}^{T}. \tag{58}$$

Rearrange some related terms, the desired results can be followed. □

Proof of Lemma 3.6

Considering the fact $\mathbf{w}^t = (V^t)^{-1}\mathbf{g}^t$, (10a) can be transformed as follows

$$\mathbf{x}^{t+1} = (\Phi \otimes I_m)\mathbf{x}^t + \alpha \mathbf{v}^t \cdot \mathbf{w}^t, \tag{59}$$

where $\mathbf{v}^t \cdot \mathbf{w}^t$ is the component wise multiplication of $\mathbf{v}^t$ and $\mathbf{w}^t$. Note that

$$\begin{aligned} \|\tilde{\mathbf{x}}^{t+1}\| &= \|\mathbf{x}^{t+1}\|_{\mathcal{L}} \\ &= \|(\Phi \otimes I_m)\mathbf{x}^t + \alpha \mathbf{v}^t \cdot \mathbf{w}^t\|_{\mathcal{L}} \\ &\le \|(\Phi \otimes I_m)\mathbf{x}^t\|_{\mathcal{L}} + \alpha\|\mathbf{v}^t \cdot \mathbf{w}^t\|_{\mathcal{L}} \\ &\le \delta_3\|\tilde{\mathbf{x}}^t\| + \alpha\|\mathbf{v}^t \cdot \mathbf{w}^t\|_{\mathcal{L}}, \end{aligned} \tag{60}$$

where the last inequality can be achieved similar to (56) with $\delta_3 = \sup\{\|\Phi \otimes I_m\|_{\mathcal{L}}\}$.

By defining $\bar{\mathbf{v}}^t = \frac{1}{n}\mathbf{1}_n \otimes (I_m^n \mathbf{v}^t)$ and $\bar{\mathbf{w}}^t = \frac{1}{n}\mathbf{1}_n \otimes (I_m^n \mathbf{w}^t)$, it follows

$$\begin{aligned} \big\|\mathbf{v}^t \cdot \mathbf{w}^t\big\|_{\mathcal{L}} &= \big\|\mathcal{L}(\mathbf{v}^t \cdot \mathbf{w}^t - \mathbf{v}^t \cdot \bar{\mathbf{w}}^t + \mathbf{v}^t \cdot \bar{\mathbf{w}}^t)\big\| \\ &\le \big\|\mathcal{L}(\mathbf{v}^t \cdot \tilde{\mathbf{w}}^t)\big\| + \big\|\mathcal{L}(\mathbf{v}^t \cdot \bar{\mathbf{w}}^t)\big\| \\ &\le \delta_4\big\|\tilde{\mathbf{w}}^t\big\| + \delta_4\big\|\bar{\mathbf{w}}^t\big\|, \end{aligned} \tag{61}$$

where $\delta_4 = \sup\{\|\mathbf{v}^t\|_{\mathcal{L}}\} < nm$ since the largest eigenvalue of $\mathcal{L}$ is 1 and $\mathbf{v}^t$ is governed by $\mathbf{v}^{t+1} = (\Psi \otimes I_m)\mathbf{v}^t$ with $\mathbf{v}^0 = \mathbf{1}$.

Combining $\mathbf{w}^t = (V^t)^{-1}\mathbf{g}^t$ and $\bar{\mathbf{w}}^t = \frac{1}{n}(\mathbf{1}_n \otimes I_m^n)\mathbf{w}^t$, we have

$$\begin{aligned} \|\bar{\mathbf{w}}^t\| &= \|\tfrac{1}{n}(\mathbf{1}_n \otimes I_m^n)(V^t)^{-1}\mathbf{g}^t\| \\ &\le \tfrac{1}{\sqrt{n}}\|I_m^n\big[(V^t)^{-1} - (V^*)^{-1}\big]\mathbf{g}^t\| + \tfrac{1}{\sqrt{n}}\|I_m^n(V^*)^{-1}\mathbf{g}^t\|. \end{aligned} \tag{62}$$

For the first term of (62), we have

$$\tfrac{1}{\sqrt{n}}\big\|I_m^n\big[(V^t)^{-1} - (V^*)^{-1}\big]\mathbf{g}^t\big\| \le \tfrac{c_3}{\sqrt{n}}\big\|I_m^n\big[(V^t)^{-1} - (V^*)^{-1}\big]\big\| \le \tfrac{c_1 c_3}{\sqrt{n}}\gamma^t, \tag{63}$$

where the first inequality follows from $\|\mathbf{g}^t\| \le c_3$ and the last one follows from Lemma 3.3.

For the second term of (62), we have

$$\big\|I_m^n(V^*)^{-1}\mathbf{g}^t\big\| = \big\|I_m^n(V^*)^{-1}\sum_{k=0}^{t}(\Psi^{t-k} \otimes I_m)\mathbf{y}^k\big\| \le M_1 + M_2, \tag{64}$$

where $M_1 = \|I_m^n(V^*)^{-1}\sum_{k=0}^t(\Psi^{t-k} \otimes I_m - \Psi^* \otimes I_m)\mathbf{y}^k\|$ and $M_2 = \|I_m^n(V^*)^{-1}\sum_{k=0}^t(\Psi^* \otimes I_m)\mathbf{y}^k\|$.

For M_1, we have

$$\begin{aligned} \tfrac{1}{\lambda^t}M_1 &= \tfrac{1}{\lambda^t}\|I_m^n(V^*)^{-1}\sum_{k=0}^{t}(\Psi^{t-k} \otimes I_m - \Psi^* \otimes I_m)\mathbf{y}^k\| \\ &\le \tfrac{1}{\lambda^t}\|I_m^n(V^*)^{-1}\|\sum_{k=0}^{t}\|(\Psi^{t-k} - \Psi^*) \otimes I_m\|\|\mathbf{y}^k\| \\ &\le c_4\sum_{k=0}^{t}\tfrac{\|(\Psi^{t-k} - \Psi^*) \otimes I_m\|}{\lambda^{t-k}}\tfrac{\|\mathbf{y}^k\|}{\lambda^k} \\ &\le c_2 c_4\|\mathbf{y}\|_\lambda^T\sum_{k=0}^{t}\left(\tfrac{\gamma}{\lambda}\right)^{t-k} \\ &\le \tfrac{c_2 c_4 \lambda}{\lambda - \gamma}\|\mathbf{y}\|_\lambda^T, \end{aligned} \tag{65}$$

where the second inequality is obtained by considering the bound $\|I_m^n(V^*)^{-1}\| = c_4$ and decomposing $\frac{1}{\lambda^t}$ as $\frac{1}{\lambda^{t-k}}\frac{1}{\lambda^k}$, the third inequality is attained by taking the maximization for $\frac{\|\mathbf{y}^k\|}{\lambda^k}$ from $k = 0$ to T and combining Lemma 3.3 to bound $\|(\Psi^{t-k} - \Psi^*) \otimes I_m\|$. For M_2, we have

$$\begin{aligned} M_2 &= \left\|I_m^n(V^*)^{-1}\sum_{k=0}^{t}(\Psi^* \otimes I_m)\mathbf{y}^k\right\| \\ &= \left\|I_m^n\sum_{k=0}^{t}\mathbf{y}^k\right\| \\ &= \left\|I_m^n[\nabla f(\mathbf{x}^t) - \nabla f(\mathbf{x}^*)]\right\| \\ &= L_{\mathrm{m}}\left\|\mathbf{x}^t - \mathbf{x}^*\right\| \\ &= L_{\mathrm{m}}\left\|\mathbf{p}^t\right\|, \end{aligned} \tag{66}$$

where the second equality is obtained according to (13c) in Lemma 3.3, the third equality is achieved by expanding $\mathbf{y}^k = \nabla f(\mathbf{x}^t) - \nabla f(\mathbf{x}^{t-1})$ and the fact $I_m^n\nabla f(\mathbf{x}^*) = 0$, and the fourth equality is deduced from the Lipschitz of $\nabla f(\mathbf{x}^t)$.

Submitting (61) – (66) into (60), and dividing λ^{t+1} both sides, yields

$$\frac{1}{\lambda^{t+1}}\|\tilde{\mathbf{x}}^{t+1}\| \leq \frac{\delta_3}{\lambda}\frac{\|\tilde{\mathbf{x}}^t\|}{\lambda^t} + \frac{\alpha\delta_4}{\lambda}\left[\frac{\|\tilde{\mathbf{w}}^t\|}{\lambda^t} + \frac{c_1c_3}{\sqrt{n}}\frac{\gamma^t}{\lambda^t} + \frac{c_2c_4\lambda\|\mathbf{y}\|_\lambda^T}{\sqrt{n}(\lambda-\gamma)} + \frac{L_{\mathrm{m}}\|\mathbf{p}^t\|}{\sqrt{n}\lambda^t}\right]. \tag{67}$$

If $\gamma \leq \lambda$ hold, taking maximization from $t = 0$ to $t = T - 1$, it follows

$$\begin{aligned} \|\tilde{\mathbf{x}}\|_\lambda^T &\leq \frac{\delta_3}{\lambda}\|\tilde{\mathbf{x}}\|_\lambda^{T-1} + \frac{\alpha\delta_4}{\lambda}\left[\|\tilde{\mathbf{w}}\|_\lambda^{T-1} + \frac{c_1c_3}{\sqrt{n}} + \frac{c_2c_4\lambda}{\sqrt{n}(\lambda-\gamma)}\|\mathbf{y}\|_\lambda^T + \frac{L_{\mathrm{m}}}{\sqrt{n}}\|\mathbf{p}\|_\lambda^{T-1}\right] \\ &\leq \frac{\delta_3}{\lambda}\|\tilde{\mathbf{x}}\|_\lambda^T + \frac{\alpha\delta_4}{\lambda}\left[\|\tilde{\mathbf{w}}\|_\lambda^T + \frac{c_1c_3}{\sqrt{n}} + \frac{c_2c_4\lambda}{\sqrt{n}(\lambda-\gamma)}\|\mathbf{y}\|_\lambda^T + \frac{L_{\mathrm{m}}}{\sqrt{n}}\|\mathbf{p}\|_\lambda^T\right]. \end{aligned} \tag{68}$$

Submitting the results of Lemma 3.4 into (68) and rearranging related terms, we have

$$\|\tilde{\mathbf{x}}\|_\lambda^T \leq \alpha\tau_1\|\tilde{\mathbf{w}}\|_\lambda^T + \alpha\bar{\tau}_2\|\mathbf{p}\|_\lambda^T + \tau_3, \tag{69}$$

where $\bar{\tau}_2 = \left[\frac{c_2c_4(\lambda+1)}{\lambda-\gamma} + 1\right]\frac{L_{\mathrm{m}}\tau_1}{\sqrt{n}} < \frac{(c_2c_4+1)(\lambda+1)}{\lambda-\gamma}\frac{L_{\mathrm{m}}\tau_1}{\sqrt{n}} = \tau_2$ because $\lambda - \gamma < \lambda + 1$.

Therefore, the desired results followed. □

Proof of Lemma 3.9

Considering $\|\bar{\mathbf{x}}^t - \mathbf{x}^*\|$ with $\bar{\mathbf{x}}^t = \frac{1}{n}I_m^n\mathbf{x}^t$, we have

$$\begin{aligned} \|\bar{\mathbf{x}}^t - \mathbf{x}^*\|^2 &= \|\bar{\mathbf{x}}^t - \bar{\mathbf{x}}^{t+1} + \bar{\mathbf{x}}^{t+1} - \mathbf{x}^*\|^2 \\ &= \|\bar{\mathbf{x}}^t - \bar{\mathbf{x}}^{t+1}\|^2 + \|\bar{\mathbf{x}}^{t+1} - \mathbf{x}^*\|^2 + 2\langle\bar{\mathbf{x}}^t - \bar{\mathbf{x}}^{t+1}, \bar{\mathbf{x}}^{t+1} - \mathbf{x}^*\rangle, \end{aligned} \tag{70}$$

therefore

$$\|\bar{\mathbf{x}}^{t+1} - \mathbf{x}^*\|^2 = \|\bar{\mathbf{x}}^t - \mathbf{x}^*\|^2 - \|\bar{\mathbf{x}}^t - \bar{\mathbf{x}}^{t+1}\|^2 - 2\langle\bar{\mathbf{x}}^t - \bar{\mathbf{x}}^{t+1}, \bar{\mathbf{x}}^{t+1} - \mathbf{x}^*\rangle. \tag{71}$$

Multiplying $\frac{1}{n}I_m^n$ on the both sides of (10a) in the proposed algorithm, we have

$$\bar{\mathbf{x}}^{t+1} = \frac{1}{n}I_m^n(\Phi \otimes I_m)\mathbf{x}^t + \alpha\bar{\mathbf{g}}^t = \bar{\mathbf{x}}^t + \alpha\bar{\mathbf{g}}^t + \frac{1}{n}\mathbf{k}\mathbf{p}^t, \tag{72}$$

where $\mathbf{k} = I_m^n(\Phi \otimes I_m - I_{nm})$. Submitting $\bar{\mathbf{x}}^{t+1} - \bar{\mathbf{x}}^t = \alpha \bar{\mathbf{g}}^t + \frac{1}{n}\mathbf{k}\mathbf{p}^t$ into the third term of (71), we have

$$\begin{aligned}\|\bar{\mathbf{x}}^{t+1} - \mathbf{x}^*\|^2 = \|\bar{\mathbf{x}}^t - \mathbf{x}^*\|^2 - \|\bar{\mathbf{x}}^t - \bar{\mathbf{x}}^{t+1}\|^2 + 2\langle \tfrac{\mathbf{k}}{n}\mathbf{p}^t, \bar{\mathbf{x}}^{t+1} - \mathbf{x}^*\rangle \\ +2\alpha\langle \bar{\mathbf{g}}^t, \bar{\mathbf{x}}^{t+1} - \mathbf{x}^*\rangle.\end{aligned} \tag{73}$$

Standing on $\bar{\mathbf{g}}^t = -\frac{1}{n}\sum_{i=1}^n \nabla f_i(x_i^t)$ and according to the results on convex function in Lemma 2.1, $\|\bar{\mathbf{x}}^{t+1} - \mathbf{x}^*\|^2$ in (73) can be bounded by

$$\begin{aligned}\|\bar{\mathbf{x}}^{t+1} - \mathbf{x}^*\|^2 \leq \|\bar{\mathbf{x}}^t - \mathbf{x}^*\|^2 - \|\bar{\mathbf{x}}^t - \bar{\mathbf{x}}^{t+1}\|^2 + 2\langle \tfrac{\mathbf{k}}{n}\mathbf{p}^t, \bar{\mathbf{x}}^{t+1} - x^*\rangle \\ +2\alpha[f(\mathbf{x}^*) - f(\bar{\mathbf{x}}^{t+1})] - \tfrac{\alpha\beta\bar{\mu}}{(1+\beta)}\|\bar{\mathbf{x}}^t - \mathbf{x}^*\|^2 \\ +\tfrac{1}{n}\sum_{i=1}^n \left[\alpha\beta\mu_i + \tfrac{\alpha(1+\eta)L_i}{\eta}\right]\|\mathbf{x}_i^t - \bar{\mathbf{x}}^t\|^2 + \alpha(1+\eta)\bar{L}\|\bar{\mathbf{x}}^{t+1} - \bar{\mathbf{x}}^t\|^2.\end{aligned} \tag{74}$$

Submitting $2\langle \frac{1}{n}\mathbf{k}\mathbf{p}^t, \bar{\mathbf{x}}^{t+1} - \mathbf{x}^*\rangle \leq \frac{1}{n^2\xi}\|\mathbf{k}\|^2\|\mathbf{p}^t\|^2 + \xi\|\bar{\mathbf{x}}^{t+1} - \mathbf{x}^*\|^2$ with $0 < \xi < 1$ into (74) and rearranging some related terms, we have

$$\begin{aligned}&(1-\xi)\|\bar{\mathbf{x}}^{t+1} - \mathbf{x}^*\|^2 \\ &\leq \left(1 - \tfrac{\alpha\beta\bar{\mu}}{(1+\beta)}\right)\|\bar{\mathbf{x}}^t - \mathbf{x}^*\|^2 + \tfrac{1}{n^2\xi}\|\mathbf{k}\|^2\|\mathbf{p}^t\|^2 + \left[\tfrac{\alpha L_{\mathrm{m}}(1+\eta)}{\eta} + \alpha\beta\mu_{\mathrm{m}}\right]\tfrac{1}{n}\sum_{i=1}^n \|\bar{\mathbf{x}}^t - \mathbf{x}_i^t\|^2 \\ &\quad -\left[1 - \alpha(1+\eta)\bar{L}\right]\|\bar{\mathbf{x}}^{t+1} - \bar{\mathbf{x}}^t\|^2 - 2\alpha\left[f(\bar{\mathbf{x}}^{t+1}) - f(\mathbf{x}^*)\right] \\ &\leq \left(1 - \tfrac{\alpha\beta\bar{\mu}}{(1+\beta)}\right)\|\bar{\mathbf{x}}^t - \mathbf{x}^*\|^2 - 2\alpha\left[f(\bar{\mathbf{x}}^{t+1}) - f(\mathbf{x}^*)\right] + \tfrac{1}{n^2\xi}\|\mathbf{k}\|^2\|\mathbf{p}^t\|^2 \\ &\quad + \tfrac{\alpha L_{\mathrm{m}}(1+\eta) + \alpha\beta\mu_{\mathrm{m}}n\eta}{n\eta}\|\tilde{\mathbf{x}}^t\|^2,\end{aligned} \tag{75}$$

where the last inequality come from the condition on step size $1 - \alpha(1+\eta)\bar{L} \geq 0$ and the fact $\|\tilde{\mathbf{x}}^t\|^2 = \sum_{i=1}^n \|\bar{\mathbf{x}}^t - \mathbf{x}_i^t\|^2$.

For $\forall t \in \mathbb{N}^+$, if following inequality holds

$$\|\bar{\mathbf{x}}^{t+1} - \mathbf{x}^*\|^2 \geq \tfrac{1}{\alpha\bar{\mu}n^2\xi}\|\mathbf{k}\|^2\|\mathbf{p}^t\|^2 + \tfrac{L_{\mathrm{m}}(1+\eta) + \beta\mu_{\mathrm{m}}n\eta}{n\bar{\mu}\eta}\|\tilde{\mathbf{x}}^t\|^2, \tag{76}$$

we have

$$\begin{aligned}2\alpha\left[f(\bar{\mathbf{x}}^{t+1}) - f(\mathbf{x}^*)\right] &\geq \alpha\bar{\mu}\|\bar{\mathbf{x}}^{t+1} - \mathbf{x}^*\|^2 \\ &\geq \tfrac{1}{n^2\xi}\|\mathbf{k}\|^2\|\mathbf{p}^t\|^2 + \tfrac{\alpha L_{\mathrm{m}}(1+\eta) + \alpha\beta\mu_{\mathrm{m}}n\eta}{n\eta}\|\tilde{\mathbf{x}}^t\|^2.\end{aligned} \tag{77}$$

Combining (75) with (77), it follows

$$\|\bar{\mathbf{x}}^{t+1} - \mathbf{x}^*\|^2 \leq \kappa_1^2\|\bar{\mathbf{x}}^t - \mathbf{x}^*\|^2, \tag{78}$$

where $\kappa_1^2 = \frac{1}{1-\xi}\left[1 - \frac{\alpha\beta\bar{\mu}}{(1+\beta)}\right]$.

Otherwise, if (76) not holds, it yields

$$\|\bar{\mathbf{x}}^{t+1} - \mathbf{x}^*\|^2 < \tfrac{1}{n^2}\kappa_2^2\|\mathbf{p}^t\|^2 + \tfrac{1}{n}\kappa_3^2\|\tilde{\mathbf{x}}^t\|^2, \tag{79}$$

where $\kappa_2^2 = \frac{1}{\alpha\bar{\mu}\xi}\|\mathbf{k}\|^2$ and $\kappa_3^2 = \frac{L_{\mathrm{m}}(1+\eta) + \beta\mu_{\mathrm{m}}n\eta}{\bar{\mu}\eta}$.

Based on (78) and (79), if $\kappa_1 \le 1$ holds, we have

$$\begin{aligned} \|\bar{\mathbf{x}}^{t+1} - \mathbf{x}^*\| &\le \max\Big\{\kappa_1\|\bar{\mathbf{x}}^t - \mathbf{x}^*\|, \tfrac{\kappa_2}{n}\|\mathbf{p}^t\| + \tfrac{\kappa_3}{\sqrt{n}}\|\tilde{\mathbf{x}}^t\|\Big\} \\ &\le \max\Big\{\kappa_1^{t+1}\|\bar{\mathbf{x}}^0 - \mathbf{x}^*\|, \max_{s=1\cdots t}\big\{\tfrac{\kappa_2}{n}\kappa_1^{t-s}\|\mathbf{p}^s\| + \tfrac{\kappa_3}{\sqrt{n}}\kappa_1^{t-s}\|\tilde{\mathbf{x}}^s\|\big\}\Big\} \\ &\le \kappa_1^{t+1}\|\bar{\mathbf{x}}^0 - \mathbf{x}^*\| + \max_{s=1\cdots t}\big\{\tfrac{\kappa_2}{n}\kappa_1^{t-s}\|\mathbf{p}^s\| + \tfrac{\kappa_3}{\sqrt{n}}\kappa_1^{t-s}\|\tilde{\mathbf{x}}^s\|\big\}, \end{aligned} \tag{80}$$

where the second inequality can be obtained by recursing the first one.

Dividing λ^{t+1} on the both sides of (80) with $\kappa_1 \le \lambda$ and taking maximization from $t = 0$ to $t = T - 1$, we obtain

$$\|\bar{\mathbf{x}} - \mathbf{x}^*\|_\lambda^T \le \|\bar{\mathbf{x}}^0 - \mathbf{x}^*\| + \tfrac{\kappa_2\lambda}{n}\|\mathbf{p}\|_\lambda^T + \tfrac{\kappa_3\lambda}{\sqrt{n}}\|\tilde{\mathbf{x}}\|_\lambda^T. \tag{81}$$

Considering $\mathbf{p}^t = \mathbf{x}^t - \mathbf{1}_n \otimes \mathbf{x}^*$, yields

$$\begin{aligned} \|\mathbf{p}^t\| &\le \|\mathbf{x}^t - \mathbf{1}_n \otimes \bar{\mathbf{x}}^t\| + \|\mathbf{1}_n \otimes \bar{\mathbf{x}}^t - \mathbf{1}_n \otimes \mathbf{x}^*\| \\ &= \|\tilde{\mathbf{x}}^t\| + \sqrt{n}\|\bar{\mathbf{x}}^t - \mathbf{x}^*\|. \end{aligned} \tag{82}$$

Apparently, we have

$$\|\mathbf{p}\|_\lambda^T \le \|\tilde{\mathbf{x}}\|_\lambda^T + \sqrt{n}\|\bar{\mathbf{x}} - \mathbf{x}^*\|_\lambda^T. \tag{83}$$

According to (81) and (83), the desired results can be achieved. □

ACKNOWLEDGEMENT

This work was supported in part by the National Key Research and Development Program of China (2016YFB0901900), and in part by the National Natural Science Foundation of China under Grant 61903027, 61733018, 61603378, and 61573344, and in part by the National Postdoctoral Program for Innovative Talents (BX20180346), and in part by the General Financial Grant from the China Postdoctoral Science Foundation (2019M660834), and in part by the Fundamental Research Funds for the China Central Universities of USTB under Grant FRF-TP-17-088A1.

(Received August 27, 2019)

REFERENCES

[1] K. Cai and H. Ishii: Average consensus on general strongly connected digraphs. Automatica *48* (2012), 2750–2761. DOI:10.1016/j.automatica.2012.08.003

[2] T. Chang, A. Nedić, and A. Scaglione: Distributed constrained optimization by consensus-based primal-dual perturbation method. IEEE Trans. Automat. Control *59* (2014), 1524–1538. DOI:10.1109/TAC.2014.2308612

[3] C. Chen, R. Chan, S. Ma, and J. Yang: Inertial proximal ADMM for linearly constrained separable convex optimization. SIAM J. Imaging Sci. *8* (2015), 2239–2267. DOI:10.1137/15100463X

[4] A. Dominguez-Garcia and C. Hadjicostis: Distributed matrix scaling and application to average consensus in directed graphs. IEEE Trans. Automat. Control *58* (2013), 667–681. DOI:10.1109/TAC.2012.2219953

[5] J. Duchi, and A. Agarwal, and M. Wainwright: Dual averaging for distributed optimization: Convergence analysis and network scaling. IEEE Trans. Automat. Control *57* (2012), 592–606. DOI:10.1109/TAC.2011.2161027

[6] B. Gharesifard and J. Cortés: Distributed strategies for generating weight-balanced and doubly stochastic digraphs. Europ. J. Control *18* (2012), 539–557. DOI:10.3166/EJC.18.539-557

[7] B. Gharesifard and J. Cortés, Jorge: Distributed continuous-time convex optimization on weight-balanced digraphs. IEEE Trans. Automat. Control *59* (2014), 781–786. DOI:10.1109/TAC.2013.2278132

[8] H. Loh: On a class of directed graphswith an application to traffic-flow problems. Oper. Res. *18* (1970), 87–94. DOI:10.1287/opre.18.1.87

[9] H. Halabian: Distributed Resource Allocation Optimization in 5G Virtualized Networks. IEEE J. Selected Areas Commun. *37* (2019), 627–642. DOI:10.1109/JSAC.2019.2894305

[10] L. Jian, J. Hu, J. W. Wang, and K. Shi: Distributed inexact dual consensus ADMM for network resource allocation. Optimal Control, Appl. Methods *40* (2019), 6, 1071–1087. DOI:10.1002/oca.2538

[11] L. Jian, Y. Zhao, J. Hu, and P. Li: Distributed inexact consensus-based ADMM method for multi-agent unconstrained optimization problem. IEEE Access *7* (2019), 1, 79311–79319. DOI:10.1109/ACCESS.2019.2923269

[12] D. Kempe, and A. Dobra, and J. Gehrke: Gossip-based computation of aggregate information. 44th Annual IEEE Symposium on Foundations of Computer Science, 2003. Proceedings. (2003), 482–491. DOI:10.1109/SFCS.2003.1238221

[13] S. Liang, and X. Zeng, and Y. Hong: Distributed nonsmooth optimization with coupled inequality constraints via modified lagrangian function. IEEE Trans. Automat. Control *63* (2018), 1753–1759. DOI:10.1109/TAC.2017.2752001

[14] S. Liang, and L. Wang, and G. Yin: Distributed quasi-monotone subgradient algorithm for nonsmooth convex optimization over directed graphs. Automatica *101* (2019), 175–181. DOI:10.1016/j.automatica.2018.11.056

[15] S. Liang, and L. Wang, and G. Yin: Exponential convergence of distributed primal-dual convex optimization algorithm without strong convexity. Automatica *105* (2019), 298–306. DOI:10.1016/j.automatica.2019.04.004

[16] P. Lin, Peng, Y. Wang, H. Qi, and Y. Hong: Distributed consensus-based K-means algorithm in switching multi-agent networks. J. Systems Sci. Complex. *31* (2018), 1128–1145. DOI:10.1007/s11424-018-7102-3

[17] S. Liu, Z. Qiu, and L. Xie: Convergence rate analysis of distributed optimization with projected subgradient algorithm. Automatica *83* (2017), 162–169. DOI:10.1016/j.automatica.2017.06.011

[18] A. Nedić and A. Ozdaglar: Distributed subgradient methods for multi-agent optimization. IEEE Trans. Automat. Control. *54* (2009), 48–61. DOI:10.1109/TAC.2008.2009515

[19] A. Nedić, A. Ozdaglar, and P. A. Parrilo: Constrained consensus and optimization in multi-agent networks. IEEE Trans. Automat. Control. *55* (2010), 922–938. DOI:10.1109/TAC.2010.2041686

[20] A. Nedić and A. Olshevsky: Stochastic gradient-push for strongly convex functions on time-varying directed graphs. IEEE Trans. Automat. Control *61* (2016), 3936–3947. DOI:10.1109/TAC.2016.2529285

[21] A. Nedić, A. Olshevsky, and W. Shi: Achieving geometric convergence for distributed optimization over time-varying graphs. SIAM J. Optim. *27* (2017), 2597–2633. DOI:10.1137/16M1084316

[22] G. Shi, B. D. Anderson, and U. Helmke: Network flows that solve linear equations. IEEE Trans. Automat. Control. *62* (2017), 2659–2674. DOI:10.1109/TAC.2016.2612819

[23] W. Shi, Q. Ling, G. Wu, and W. Yin: Extra: An exact first-order algorithm for decentralized consensus optimization. SIAM J. Optim. *25* (2015), 944–966. DOI:10.1137/14096668X

[24] K. I. Tsianos, S. Lawlor, and M. G. Rabbat: Push-sum distributed dual averaging for convex optimization. In: 2012 IEEE 51st IEEE Conference on Decision and Control (CDC) 2012, pp. 5453–5458. DOI:10.1109/CDC.2012.6426375

[25] K. I. Tsianos and M. G. Rabbat: Distributed dual averaging for convex optimization under communication delays. In: 2012 American Control Conference (ACC) 2012, pp. 1067–1072. DOI:10.1109/ACC.2012.6315289

[26] C. Xi and U. A. Khan: On the linear convergence of distributed optimization over directed graphs. arXiv preprint arXiv:1510.02149 (2015).

[27] C. Xi and U. A. Khan: Distributed subgradient projection algorithm over directed graphs. IEEE Trans. Automat. Control *62* (2017), 3986–3992. DOI:10.1109/TAC.2016.2615066

[28] R. Xin, C. Xi, and U. A. Khan: Distributed Subgradient Projection Algorithm over Directed Graphs: Alternate Proof. arXiv preprint arXiv:1706.07707 (2017).

[29] R. Xin, C. Xi, and U. A. Khan: FROST-Fast row-stochastic optimization with uncoordinated step-sizes. EURASIP J. Advances Signal Process. *1* (2019), 1–14. DOI:10.1186/s13634-018-0596-y

[30] T. Yang, J. George, J. Qin, X. Yi, and J. Wu: Distributed finite-time least squares solver for network linear equations. arXiv preprint arXiv:1810.00156(2018).

[31] P. Yi, Y. Hong, F. and Liu: Initialization-free distributed algorithms for optimal resource allocation with feasibility constraints and its application to economic dispatch of power systems. Automatica *74* (2016), 259–269. DOI:10.1016/j.automatica.2016.08.007

[32] D. Yuan, A. Proutiere, and G. Shi: Distributed Online Linear Regression. arXiv preprint arXiv:1902.04774.

[33] X. Zeng, P. Yi, and Y. Hong: Distributed continuous-time algorithm for constrained convex optimizations via nonsmooth analysis approach. IEEE Trans. Automat. Control *62* (2017), 5227–5233. DOI:10.1109/TAC.2016.2628807

Songsong Cheng, The School of Electrical Engineering and Automation, Anhui University, Hefei 230601. P. R. China.
e-mail: sscheng@amss.ac.cn

Shu Liang, The Key Laboratory of Knowledge Automation for Industrial Processes of Ministry of Education, School of Automation and Electrical Engineering, University of Science and Technology Beijing, Beijing 100083. P. R. China.
e-mail: sliang@ustb.edu.cn.