An Improved Distributed Optimization Algorithm over Unbalanced Directed Graph

Zhenteng Gao and Lipo Mo(✉)

School of Mathematics and Statistics, Beijing Technology and Business University, Beijing 100048, China
beihangmlp@126.com, molipo@th.btbu.edu.cn

Abstract. This paper mainly discusses the common distributed optimization problem over unbalanced directed graph. Assumed that the local objective function of each agent is strongly convex and has a Lipschitz continuous gradient. An improved distributed algorithm is proposed by introducing a momentum term and different local step lengths. Then we prove that all agents would find the optimal value under our algorithm when the maximum step length and the momentum parameter satisfy a certain range and are positive. At last, we illustrate the effectiveness of the obtained results by a numerical experiment.

Keywords: Distributed optimization · Random weight · Distributed step size · Unbalanced directed graph

1 Introduction

Recently, distributed optimization problems have received extensive attention, and it is very helpful to solve this problem by the distributed consensus algorithm. We improved the consensus algorithm to solve the distributed optimization problem in this paper, where each agent can access to one cost function $f_i : \mathbb{R}^m \rightarrow \mathbb{R}$, and all agents collaboratively minimize the entire function $\frac{1}{n}\sum_{i=1}^{n} f_i(x)$ through the exchange of information between agents. This paper focuses on the situation of unbalanced directed graph. Early works about distributed optimization problems mainly included distributed gradient descent [1] and distributed dual averaging [2] over undirected graphs. It was proved that the optimal value could be found at a linear rate of $O(\frac{lnk}{\sqrt{k}})$ for any convex function, and the rate of $O(\frac{lnk}{k})$for any strongly convex function, where k is the number of iterations. Under the conditions of strong convexity and Lipschitz continuous gradient, algorithms were improved with faster convergence speed. For example, the algorithm with a constant step size geometrically converged to an error ball around the optimal solution, there was another method that requires symmetric weights to achieve global geometric convergence. In [4–6], the imprecise gradient method and the gradient estimation method were introduced to deal with this problem.

Z. Ren et al. (Eds.): *Proceedings of 2021 5th Chinese Conference on Swarm Intelligence and Cooperative Control*, LNEE 934, pp. 515–525, 2023.
https://doi.org/10.1007/978-981-19-3998-3_49

The aforementioned methods were all for undirected graph. When the communication capabilities between agents were inconsistent, the algorithm for undirected graph would no longer be applicable. Therefore, algorithms suitable for directed graphs need to be developed. The Push-sum method and the DGD (distributed sub-gradient descent) method were introduced for directed graphs in [7–10]. However, the effect of the reduction of the step size resulted in a relatively slow convergence rate. Literature [11] assumed that the objective function has a Lipschitz continuous gradient and is strongly convex. It was shown that all agents would converge to the optimal value at geometrical rate. By constructing a row stochastic matrix and a column stochastic matrix, another type of algorithm was proposed [12–14], where the row random matrix ensured the consistency of the algorithm, and the column randomness matrix was used to guarantee the optimality. In [12,13], the cases of fixed strong connectivity and time-varying strong connectivity were considered, based on which, the gravity ball was introduced to improve the convergence rate of algorithms.

For second order and heterogeneous multi-agent systems, some improve algorithms were also proposed in [16,17]. Inspired by the literature [12], we studies an improved fully distributed algorithm to optimize all objective functions in a distributed manner, where the momentum term is borrowed to improve the convergence rate. It is shown that the position states of every agents would converge to the optimal solution of the objective function by the nature of the random matrix.

I_n represents an n-dimensional unit matrix, and 1_n represents a column vector whose components are all ones. $\rho(x)$ represents the spectral radius of the vector x, and X_∞ represents the infinite power of the matrix X. For the row random matrix A, π_r and 1_n to represent the left and right eigenvectors of A respectively, such that $\pi_r^T 1_n = 1$. Similarly, for the column random matrix B, 1_n and π_c to represent the left and right eigenvectors of B respectively, such that $\pi_c^T 1_n = 1$. $\|\cdot\|_2$ represents the 2-Norm of the vector. $|||\cdot|||_2$ represents the spectral norm of the matrix.

2 Graph Theory Foundation and Problem Description

$\mathcal{G} = (\mathcal{V}, \mathcal{E})$ denotes a directed graph, where $\mathcal{V} = \{1, 2, \ldots, n\}$ represents the set of network agents, and $\mathcal{E}$ represents the set of edges between agents in the network. (j, i) or $j \to i$ indicates that there is a directed edge that transmits information from the agent j to the agent i. For $\forall\ i, j$, if there is a directed path $(i_1, i_{s1}), (i_{s1}, i_{s2}), \ldots, (i_{sk}, j)$, then it is called a strongly connected graph. In addition, $N_i^{in} = \{j \mid (j, i) \in \mathcal{E}\}$ represents the into-neighbor set of the agent i, that is the set of agents that the agent i can receives information from. Similarly, $N_i^{out} = \{j \mid (i, j) \in \mathcal{E}\}$ represents the out-neighbors set of the agent i, that is the set of agents that can receive information from agent i. Note that both N_i^{in} and N_i^{out} contain node i.

In the distributed convex optimization problem, each agent i can access to a local decision variable $x_i \in \mathbb{R}^m$ and a convex cost function $f_i(x)$. The goal of this problem is to minimize the following integral objective function.

$$\min f(x) = \frac{1}{n}\sum_{i=1}^{n} f_i(x), x \in \mathbb{R}^m \tag{1}$$

Each agent i can only obtain its own cost function $f_i(x) : \mathbb{R}^m \to \mathbb{R}$. Assume that each cost function is strongly convex and its gradient is Lipschitz.

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

Assumption 2. The gradient of the objective function of each agent satisfies Lipschitz condition, that is, for any agent i and $x, y \in \mathbb{R}^m$, there is a constant l_i such that:

$$\|\nabla f_i(x) - \nabla f_i(y)\| \leq l_i\|x - y\| \tag{2}$$

Assumption 3. The cost function of each agent is strongly convex, that is, for any agent i and $x, y \in \mathbb{R}^m$, there is a positive constant μ such that:

$$f_i(x) - f_i(y) \geq \nabla f_i(x)^{\mathrm{T}}(x - y) - \frac{\mu}{2}\|x - y\|_2^2 \tag{3}$$

Remark 1: Assumption 2 and Assumption 3 ensure that the global optimal solution x^* exists and is unique respectively. Assumption 3 is conducive to the subsequent proof of the convergence of the algorithm.

3 Algorithm Design

We propose the following algorithm to solve problem (1) in this paper. Each agent i contains two variables $x_{i,k}, s_{i,k}$ in the network, and k represents iteration step, where $i \in \mathcal{V}$, $x_{i,k}, s_{i,k} \in \mathbb{R}^m$. The system satisfies the initial state $s_{i,0} = \nabla f_i(x_{i,0}), i \in \mathcal{V}$.

$$x_{i,k+1} = \sum_{j=1}^{n} a_{ij}x_{j,k} - \alpha_i s_{i,k} + \beta[\sum_{j=1}^{n} a_{ij}(x_{j,k} - x_{i,k})]_- \tag{4a}$$

$$s_{i,k+1} = \sum_{j=1}^{n} b_{ij}[s_{j,k} + \nabla f_j(x_{j,k+1}) - \nabla f_j(x_{j,k})] \tag{4b}$$

where α_i and β are both positive constants. The weights a_{ij} and b_{ij} satisfy the following:

$$a_{ij} = \begin{cases} > 0, & j \in N_i^{in}, \\ 0, & j \notin N_i^{in}, \end{cases} \quad \sum_{j=1}^{n} a_{ij} = 1, \forall i, \tag{5}$$

$$b_{ij} = \begin{cases} > 0, & i \in N_j^{out}, \\ 0, & i \notin N_j^{out}, \end{cases} \quad \sum_{i=1}^{n} b_{ij} = 1, \forall j, \tag{6}$$

$\bar{A} = \{a_{ij}\}$ represents the row random matrix, and $\bar{B} = \{b_{ij}\}$ represents the column random matrix.

Denote $x_k = [x_{1,k}^T, \ldots, x_{n,k}^T]^T$, $s_k = [s_{1,k}^T, \ldots, s_{n,k}^T]^T$, $\nabla f(x_k) = [\nabla f_1(x_{1,k})^T, \ldots, \nabla f_n(x_{n,k})^T]^T$. Let$A = \bar{A} \otimes I_m$,$B = \bar{B} \otimes I_m$, then Eq. (4) can be rewritten the following form:

$$x_{k+1} = Ax_k - D_\alpha s_k + \beta[Ax_k - x_k]_- \tag{7a}$$

$$s_{k+1} = B[s_k + \nabla f(x_{k+1}) - \nabla f(x_k)] \tag{7b}$$

where D_α represents a diagonal matrix whose diagonal elements are α_i, and other elements are 0, where $s_0 = \nabla f(x_0)$, and x_0 is arbitrary.

4 Algorithm Convergence Analysis

First, let us prove a key lemma, which involves the shrinkage of the consistency process of the row and column random matrix respectively.

Lemma 1. $A = \bar{A} \otimes I_m$ and $B = \bar{B} \otimes I_m$ are weight matrices, there are vector norms $\| \cdot \|_A$ and $\| \cdot \|_B$ such that for $\forall x \in \mathbb{R}^{mn}$,

$$\|Ax - A_\infty\|_A \leq \sigma_A \|x - A_\infty x\|_A \tag{8}$$

$$\|Bx - B_\infty\|_B \leq \sigma_B \|x - B_\infty x\|_B. \tag{9}$$

Proof. Since $\bar{A}$ is irreducible, its diagonal elements are all positive, and the rows are random. According to the Perro-Frobenius theorem, $\rho(\bar{A}) = 1$. Every eigenvalue except 1 is strictly less than $\rho(\bar{A})$, π_r^T is a strictly positive left eigenvector corresponding to eigenvalue 1, and $\pi_r^T 1_n = 1$. Therefore, $\lim_{k\to\infty} \bar{A}^k = 1_n \pi_r^T$, and

$$A_\infty = \lim_{k\to\infty} A^k = (\lim_{k\to\infty} \bar{A}^k) \otimes I_m = (1_n \pi_r^T) \otimes I_m.$$

Then

$$AA_\infty = (\bar{A} \otimes I_m)((1_n \pi_r^T) \otimes I_m) = A_\infty$$
$$A_\infty A_\infty = ((1_n \pi_r^T) \otimes I_m)((1_n \pi_r^T) \otimes I_m) = A_\infty$$

Therefore,$AA_\infty - A_\infty A_\infty = 0$, then there are the following formulas

$$Ax - A_\infty x = (A - A_\infty)(x - A_\infty x). \tag{10}$$

Because $\rho(A - A_\infty) = \rho((\bar{A} - 1_n \pi_r^T) \otimes I_m) < 1$, according to [15], there is a matrix norm $|||\cdot|||_A$ such that $\sigma_A = |||A - A_\infty|||_A < 1$. In addition, according to Theorem 5.7.13 in [15], there is a corresponding vector norm $|||\cdot|||_A$ for any matrix norm $\| \cdot \|_A$, such that for all matrices Y and vectors y , $\|Yy\|_A \leq |||Y|||_A \|y\|_A$. Therefore, Eq. (10) leads to:

$$\begin{aligned} \|Ax - A_\infty x\|_A &= \|(A - A_\infty)(x - A_\infty x)\|_A \\ &\leq |||A - A_\infty|||_A \|x - A_\infty x\|_A = \sigma_A \|x - A_\infty x\|_A. \end{aligned}$$

The Eq. (8) of Lemma 1 is proved. The same is true for Eq. (9).

Lemma 2.

$$(1_n^T \otimes I_m)s_k = (1_n^T \otimes I_m)\nabla f(x_k), \forall k.$$

Proof. $(1_n^T \otimes I_m)s_k = (1_n^T \otimes I_m)(\bar{B} \otimes I_m)[s_k + \nabla f(x_{k+1}) - \nabla f(x_k)] = (1_n^T \otimes I_m)s_k + (1_n^T \otimes I_m)(\nabla f(x_{k+1}) - \nabla f(x_k)) = (1_n^T \otimes I_m)(s_0 - \nabla f(x_0)) + (1_n^T \otimes I_m)\nabla f(x_k) = (1_n^T \otimes I_m)\nabla f(x_k)$.

Lemma 3 [18]**.** If the function f satisfies Assumptions 2 and 3, and l and μ are respectively strongly convex and Lipschitz continuous coefficients, then for $\forall x \in \mathbb{R}^m, 0 < \alpha < \frac{1}{l}$,

$$\|x - \alpha\nabla f(x) - x^*\| \le (1 - \mu\alpha)\|x - x^*\|.$$

Lemma 4 [15]**.** Suppose $W \in \mathbb{R}^{n\times n}$ is non-negative, and $w \in \mathbb{R}^n$ is positive. If $Ww < \zeta w$ with $\zeta > 0$, then $\rho(W) < \zeta$.

The subsequent analysis of convergence is carried out from the contraction relationship of the following four quantities.

1) $\|x_{k+1} - A_\infty x_{k+1}\|_A$;
2) $\|x_{k+1} - Ax_{k+1}\|_2$;
3) $\|A_\infty x_{k+1} - 1_n \otimes x^*\|_2$;
4) $\|s_{k+1} - B_\infty s_{k+1}\|_B$.

Norms in finite-dimensional linear space are equivalent, that is, there are positive constants c, d, h, q, g, p such that the vector norm satisfies the following inequality:

$$\|\cdot\|_A \le c\|\cdot\|_B, \quad \|\cdot\|_2 \le h\|\cdot\|_B, \quad \|\cdot\|_2 \le g\|\cdot\|_A,$$

$$\|\cdot\|_B \le d\|\cdot\|_A, \quad \|\cdot\|_B \le q\|\cdot\|_2, \quad \|\cdot\|_A \le p\|\cdot\|_2.$$

Lemma 5. For $\forall k \ge 0$, the following inequality holds,

$$\|s_k\|_2 \le h\|s_k\|_B + \vert\vert\vert B\vert\vert\vert_2 \bar{l} g\|s_k\|_A + \vert\vert\vert B\vert\vert\vert_2 \bar{l}\|A_\infty x_k - 1_n \otimes x^*\|_2$$

where $\bar{l} = \max\{l_i\}$.

Proof.

$$\|s_k\|_2 \le h\|s_k - B_\infty s_k\|_B + \|B_\infty s_k\|_2$$

$\|B_\infty s_k\|_2 = \|(\pi_c \otimes I_m)(1_n^T \otimes I_m)s_k\|_2 = \|\pi_c\|_2\|(1_n^T \otimes I_m)s_k\|_2 = \|\pi_c\|_2\|\sum_{i=1}^n \nabla f_i(x_{i,k}) - \sum_{i=1}^n \nabla f_i(x^*)\|_2 \le \|\pi_c\|_2\bar{l}\sum_{i=1}^n \|x_{i,k} - x^*\|_2 \le \|\pi_c\|_2\bar{l}\sqrt{n}\|x_k - 1_n \otimes x^*\|_2 \le \vert\vert\vert B\vert\vert\vert_2 \bar{l} g\|x_k - A_\infty x_k\|_A + \vert\vert\vert B\vert\vert\vert_2 \bar{l}\|A_\infty x_k - 1_n \otimes x^*\|_2$.

The proof is completed.

Lemma 6. For $\forall k \geq 0$, we have the following inequality,
$\|x_{k+1} - A_\infty x_{k+1}\|_A \leq \sigma_A \|x_k - A_\infty x_k\|_A + \bar{\alpha} p |||I_{mn} - A_\infty|||_2 \|s_k\|_2 + \beta |||I_{mn} - A_\infty|||_A \|Ax_k - x_k\|_A$.
where $\bar{\alpha} = \max\{\alpha_i\}$.

Proof. $\|x_{k+1} - A_\infty x_{k+1}\|_A = \|Ax_k - D_\alpha s_k + \beta[Ax_k - x_k]_- - A_\infty x_k + A_\infty D_\alpha s_k \quad - \beta A_\infty [Ax_k - x_k]_- \|_A \leq \sigma_A \|x_k - A_\infty x_k\|_A + \bar{\alpha}\|s_k - A_\infty s_k\|_A + \beta |||I_{mn} - A_\infty|||_A \|Ax_k - x_k\|_A \leq \sigma_A \|x_k - A_\infty x_k\|_A + \bar{\alpha} p |||I_{mn} - A_\infty|||_2 \|s_k\|_2 \quad + \beta |||I_{mn} - A_\infty|||_A \|Ax_k - x_k\|_A$.

Lemma 7. For $\forall k \geq 0$, we have the following inequality,
$\|x_{k+1} - Ax_{k+1}\|_2 \leq (\sigma_A + \sigma_A^2) g \|x_k - A_\infty x_k\|_A + \bar{\alpha} |||I_{mn} - A|||_2 \|s_k\|_2 + \beta |||I_{mn} - A|||_2 \|x_k - Ax_k\|_2$.

Proof. $\|x_{k+1} - Ax_{k+1}\|_2 = \|Ax_k - D_\alpha s_k + \beta[Ax_k - x_k]_- \quad - A^2 x_k + AD_\alpha s_k - \beta A[Ax_k - x_k]_- \|_A = \|Ax_k - A_\infty x_k - A^2 x_k + A_\infty x_k - D_\alpha s_k + AD_\alpha s_k + \beta[Ax_k - x_k]_- - \beta A[Ax_k - x_k]_- \|_A \leq \sigma_A g \|x_k - A_\infty x_k\|_A + \sigma_A^2 g \|x_k - A_\infty x_k\|_A + \bar{\alpha}\|s_k - As_k\|_2 + \beta |||I_{mn} - A|||_2 \|x_k - Ax_k\|_2 \leq (\sigma_A + \sigma_A^2) g \|x_k - A_\infty x_k\|_A + \bar{\alpha} |||I_{mn} - A|||_2 \|s_k\|_2 + \beta |||I_{mn} - A|||_2 \|x_k - Ax_k\|_2$.

Lemma 8. When $0 < \bar{\alpha} < \frac{1}{nl\pi_r^T \pi_c}$, for $\forall k \geq 0$, we have the following inequality:
$\|A_\infty x_{k+1} - 1_n \otimes x^*\|_2 \leq (1 - n\mu(\pi_r^T \pi_c)\bar{\alpha}) \|A_\infty x_k - 1_n \otimes x^*\|_2 + \bar{\alpha}(\pi_r^T \pi_c) nlg \|x_k - A_\infty x_k\|_A + \bar{\alpha} h \|s_k - B_\infty s_k\|_B + \beta |||A_\infty|||_2 \|x_k - Ax_k\|_2$

Proof. $\|A_\infty x_{k+1} - 1_n \otimes x^*\|_2 = \|A_\infty (Ax_k - D_\alpha s_k + (D_\alpha - D_\alpha) B_\infty s(k) + \beta[Ax_k - x_k]_-) - 1_n \otimes x^*\|_2 \leq \|A_\infty x_k - A_\infty D_\alpha B_\infty \nabla f(x_k) - (1_n \otimes I_m) x^*\|_2 + \bar{\alpha} h \|s_k - B_\infty s_k\|_B + \beta |||A_\infty|||_2 \|x_k - Ax_k\|_2$.

$A_\infty B_\infty = ((1_n \pi_r^T) \otimes I_m)((\pi_c 1_n^T) \otimes I_m) = (\pi_r^T \pi_c)(1_n 1_n^T) \otimes I_m$.

$\|((1_n \pi_r^T) \otimes I_m) x_k - (1_n \otimes I_m) x^* A_\infty D_\alpha B_\infty \nabla f(x_k)\|_2 = \|(1_n \otimes I_m)((\pi_r^T \otimes I_m) x_k - (\pi_r^T diag(\alpha) \pi_c)(1_n^T \otimes I_m) \nabla f(x_k) - x^*\|_2) \leq \|(1_n \otimes I_m)((\pi_r^T \otimes I_m) x_k - n\pi_r^T \pi_c \bar{\alpha} \nabla f(\pi_r^T \otimes I_m x_k) - x^*)\|_2 + n\pi_r^T \pi_c \bar{\alpha} \|(1_n \otimes I_m)(n \nabla f((\pi_r^T \otimes I_m) x_k - (1_n \otimes I_m) \nabla f_k)\|_2 \triangleq s_1 + s_2$.

From Lemma 3, if $0 < n(\pi_r^T \pi_c)\bar{\alpha} < \frac{1}{l}$.

$s_1 = \sqrt{n}\|(\pi_r^T \otimes I_m) x_k - n\pi_r^T \pi_c \bar{\alpha} \nabla F((\pi_r^T \otimes I_m) x_k) - x^*\|_2 \leq \sqrt{n}(1 - n(\pi_r^T \pi_c)\bar{\alpha}) \|(\pi_r^T \otimes I_m) x_k - x^*\|_2 = (1 - n\mu(\pi_r^T \pi_c)\bar{\alpha}) \|A_\infty x_k - 1_n \otimes x^*\|_2$,

$s_2 \leq \bar{\alpha}(\pi_r^T \pi_c) n \|\nabla f((1_n \otimes I_m)(\pi_r^T \otimes I_m) x_k) - \nabla f(x_k)\|_2 \leq \bar{\alpha}(\pi_r^T \pi_c) nlg \|x_k - A_\infty x_k\|_A$.

Lemma 9. For $\forall k \geq 0$, we have the following inequality
$\|s_{k+1} - B_\infty s_{k+1}\|_B \leq \sigma_B \|s_k - B_\infty s_k\|_B + \sigma_B q\bar{l} g |||A - I_{mn}|||_2 \|x_k - A_\infty x_k\|_2 + \sigma_B q\bar{l}\beta \|x_k - Ax_k\|_2 + \sigma_B q\bar{l}\bar{\alpha}\|s_k\|_2$.

Proof. $\|s_{k+1} - B_\infty s_{k+1}\|_B = \|B[s_k + \nabla f(x_{k+1}) - \nabla f(x_k)] - B_\infty B[s_k + \nabla f(x_{k+1}) - \nabla f(x_k)]\|_B \leq \sigma_B \|s_k - B_\infty s_k\|_B + \sigma_B \bar{l} q \|x_{k+1} - x_k\|_2$,

$\|x_{k+1} - x(k)\|_2 = \|Ax_k - D_\alpha s(k) + \beta[Ax_k - x_k]_- - x_k\|_2 \leq \|(A - I_{mn})(x_k - A_\infty x_k)\|_2 + \beta\|x_k - Ax_k\|_2 + \bar{\alpha}\|s_k\|_2$.

5 Analysis of Convergence Results

The analysis of the convergence results is given below.

Theorem 1.

$$t_{k+1} < J_{\bar{\alpha},\beta} t_k, \forall k \geq 0$$

where $t_k \in \mathbb{R}^4, J_{\bar{\alpha},\beta} \in \mathbb{R}^{4\times 4}$ are given by.

$$t_k = \begin{pmatrix} \|x_k - A_\infty x_k\|_A \\ \|A_\infty x_k - 1_n \otimes x^*\|_2 \\ \|x_k - Ax_k\|_2 \\ \|s_k - B_\infty s_k\|_B \end{pmatrix}$$

$$J_{\bar{\alpha},\beta} = \begin{pmatrix} \sigma_A + a_1\bar{\alpha} & a_2\bar{\alpha} & a_3\beta & a_4\bar{\alpha} \\ a_5\bar{\alpha} & 1 - a_6\bar{\alpha} & a_7\beta & a_8\bar{\alpha} \\ (\sigma_A + \sigma_A^2)a_9 + a_{10}\bar{\alpha} & a_{11}\bar{\alpha} & a_{12}\beta & a_{13}\bar{\alpha} \\ \sigma_B a_{14} + \sigma_B a_{15}\bar{\alpha} & \sigma_B a_{16}\bar{\alpha} & \sigma_B a_{17}\beta & \sigma_B + \sigma_B a_{18}\bar{\alpha} \end{pmatrix}$$

where a_i in the above expression are $a_1 = |||B|||_2 \bar{l} gm |||I_{mn} - A_\infty|||_2$, $a_2 = p\bar{l}|||B|||_2|||I_{mn} - A_\infty|||_2$, $a_3 = |||I_{mn} - A_\infty|||_A$, $a_4 = ph|||I_{mn} - A_\infty|||_2$, $a_5 = (\pi_r^T \pi_c)nlg$, $a_6 = n\mu(\pi_r^T\pi_c)$, $a_7 = |||A_\infty|||_2$, $a_8 = h$, $a_9 = g$, $a_{10} = \bar{l}g|||B_\infty|||_2|||I_{mn} - A_\infty|||_2$, $a_{11} = |||I_{mn} - A_\infty|||_2|||B_\infty|||_2$, $a_{12} = |||I_{mn} - A_\infty|||_2$, $a_{13} = h|||I_{mn} - A_\infty|||_2$, $a_{14} = \bar{l}qg|||I_{mn} - A_\infty|||_2$, $a_{15} = \bar{l}^2qg|||B_\infty|||_2$, $a_{16} = \bar{l}^2q|||B_\infty|||_2$, $a_{17} = \bar{l}q$, $a_{18} = \bar{l}qh$

Define the positive vector $\boldsymbol{\delta} = [\delta_1, \delta_2, \delta_3, \delta_4]$, where

$\delta_1 = 1 - \sigma_B, \quad \delta_2 = 2\frac{a_5(a-\sigma_B)+2\sigma_B a_{14}}{a_6}, \delta_3 = 2(\sigma_A + \sigma_A^2)a_9, \quad \delta_4 = 2\sigma_B a_{14}$

If $\bar{\alpha}$ and β are within:

$$0 < \bar{\alpha} < \min\{\frac{1}{nl\pi_r^T\pi_c}, \frac{(1-\sigma_A)\delta_1}{a_1\delta_1 + a_2\delta_2 + a_4\delta_4}, \frac{\delta_3 - (\sigma_A + \sigma_A^2)a_9}{a_{10}\delta_1 + a_{11}\delta_2 + a_{13}\delta_4}, \frac{(1-\sigma_B)\delta_4 - \sigma_B a_{14}\delta_1}{a_{15}\sigma_B\delta_1 + a_{16}\sigma_B\delta_2 + a_{18}\sigma_B\delta_4}\} \tag{11}$$

$$0 < \beta < \min\{\frac{(1-\sigma_A) - (a_1\delta_1 + a_2\delta_2 + a_4\delta_4)\bar{\alpha}}{a_3\delta_3}, \frac{a_6\delta_2 - a_5\delta_1 + a_8\delta_4}{a_7\delta_3}, \frac{\delta_3 - (\sigma_A + \sigma_A^2)a_9 - a_{10}\bar{\alpha}\delta_1 - (a_{11}\delta_2 + a_{13}\delta_4)\bar{\alpha}}{a_{12}\delta_3}, \frac{(1-\sigma_B)\delta_4 - \sigma_B a_{14}\delta_1 - (a_{15}\sigma_B\delta_1 + a_{16}\sigma_B\delta_2 + a_{18}\sigma_B\delta_4)\bar{\alpha}}{\sigma_B a_{17}\delta_3}\} \tag{12}$$

then, $\rho(J_{\bar{\alpha},\beta}) < 1$. Therefore, $\|x(k) - 1_n \bigotimes x^*\|_2$ converges linearly to 0 at the rate of $\mathcal{O}(\rho(J_{\bar{\alpha},\beta}))^k$.

Proof. It is easy to verify $t_{k+1} < J_{\bar{\alpha},\beta} t_k, \forall k \geq 0$ from Lemmas 5–9. To prove that $\|x_k - 1_n \bigotimes x^*\|_2$ linearly converges, just prove that there are $\bar{\alpha}$ and β such that $\rho(J_{\bar{\alpha},\beta}) < 1$. From Lemma 4, we need to proof there exist $\bar{\alpha}$ and β that satisfies $J_{\bar{\alpha},\beta}\boldsymbol{\delta} < \boldsymbol{\delta}$ for some positive vector $\boldsymbol{\delta} = [\delta_1, \delta_2, \delta_3, \delta_4]$ and solve the range of $\bar{\alpha}$ and β. The inequality is changed into.

$$a_3\delta_3\beta < (1-\sigma_A)\delta_1 - (a_1\delta_1 + a_2\delta_2 + a_4\delta_4)\bar{\alpha} \tag{13}$$

$$a_7\delta_3\beta < (a_6\delta_2 - a_5\delta_1 - a_8\delta_4)\bar{\alpha} \tag{14}$$

$$a_{12}\delta_3\beta < \delta_3 - ((\sigma_A + \sigma_A^2)a_9 + a_{10}\bar{\alpha}\delta_1) - a_{11}\delta_2\bar{\alpha} - a_{13}\delta_4\bar{\alpha} \tag{15}$$

$$\sigma_B a_{17}\delta_3\beta < -\sigma_B a_{14}\delta_1 + (1-\sigma_B)\delta_4 - (\sigma_B a_{18}\delta_4 + \sigma_B a_{15}\delta_1 + \sigma_B a_{16}\delta_2)\bar{\alpha}. \tag{16}$$

Since $\beta > 0$, the right side of the above four inequalities are positive, which can derive the range of $\bar{\alpha}, \delta_1, \delta_2, \delta_3, \delta_4$.

$$\bar{\alpha} < \frac{(1-\sigma_A)\delta_1}{a_1\delta_1 + a_2\delta_2 + a_4\delta_4} \tag{17}$$

$$\bar{\alpha} < \frac{\delta_3 - (\sigma_A + \sigma_A^2)a_9}{a_{10}\delta_1 + a_{11}\delta_2 + a_{13}\delta_4} \tag{18}$$

$$\bar{\alpha} < \frac{(1-\sigma_B)\delta_4 - \sigma_B a_{14}\delta_1}{a_{15}\sigma_B\delta_1 + a_{16}\sigma_B\delta_2 + a_{18}\sigma_B\delta_4} \tag{19}$$

$$\delta_2 > \frac{a_5\delta_1 + a_8\delta_4}{a_6} \tag{20}$$

Because $\bar{\alpha} > 0$, we can choose $\delta_1, \delta_2, \delta_3, \delta_4$ to make $\bar{\alpha}$ positive. According to formulas (17)–(20), select the value of δ_i as follows: $\delta_1 = 1 - \sigma_B, \quad \delta_2 = 2\frac{a_5(a-\sigma_B)+2\sigma_B a_1 4}{a_6}, \delta_3 = 2(\sigma_A + \sigma_A^2)a_9, \quad \delta_4 = 2\sigma_B a_1 4$

After determining δ_i, the upper bound of $\bar{\alpha}$ can be determined according to inequality (13)–(16), and the upper bound of β can be determined according to inequality (17)–(20). Theorem 1 is finally proved.

Remark 2: From the above theorem, we can obtain the linear convergence rate of the algorithm. However, since the equivalent constants between σ_A and σ_B and the norm are unknown, the boundary between α and β cannot be clearly given, it is necessary to manually adjust the parameters to get the best performance.

6 Numerical Experiment

This paper use Matlab to demonstrate the simulation effect to check the effectiveness of the algorithm.

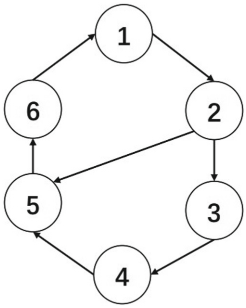

Fig. 1. Directed graph with six agents

Figure 1 denotes a communication topology between 6 agents. We consider the distributed convex optimization problem in the directed strongly connected network with 6 agents. And the local objective function of each agent is

$f_1(x_1) = x_1^2 - 2x_1 + cos(x_1) + 3, f_2(x_2) = x_2^2 - 5x_2 + e^{-0.1x_2} - 1,$
$f_3(x_3) = x_3^2 - 3x_3 - 0.5sin(x_3) - 3, f_4(x_4) = x_4^2 + 2x_4^4 - 3.$
$f_5(x_5) = x_5^2 + 3x_5 + 1, f_6(x_6) = 4x_6^2 + 2x_6 - cos(x_6) + 3$

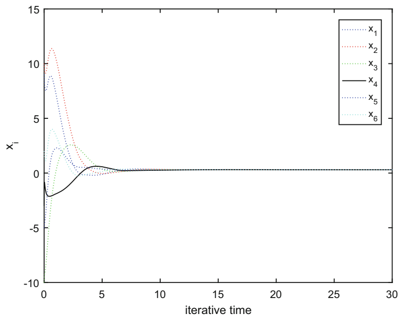

Fig. 2. Agent state trajectory

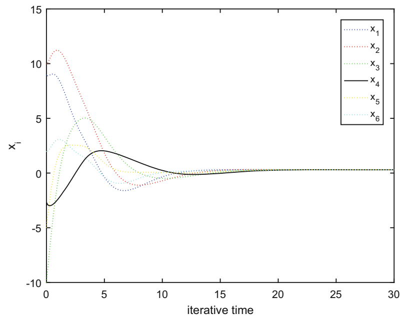

Fig. 3. Agent state trajectory (extra)

The optimal solution in this case is $x^* = 0.2980$. Figure 2 shows that the algorithm proposed in this paper finally find the optimal value. Figure 3 is the agent trajectory diagram of the algorithm proposed in [12], and shows that the algorithm proposed in this article has a faster convergence rate. Distributed optimization can also be applied to drone formation, if a drone wants to control its own position, it can make a decision based on the position information of the nearby drones to determine its own position.

7 Conclusion

We proposed an improved fully distributed optimization algorithm for the directed strongly connected graph in this paper. Assume that the objective function is strongly convex with Lipschitz continuous gradient, all agents can be forced to converge to the optimal point at geometric rate under the algorithm. By introducing row stochastic, column stochastic matrix, and a momentum term, the rate of convergence of our algorithm is higher than that of the literatures.

References

1. Nedic, A., Ozdaglar, A.: Distributed subgradient methods for multi-agent optimization. IEEE Trans. Autom. Control **54**(1), 48–61 (2009). https://doi.org/10.1109/TAC.2008.2009515
2. Duchi, J.C., Agarwal, A., Wainwright, M.J.: Dual averaging for distributed optimization: convergence analysis and network scaling. IEEE Trans. Autom. Control **57**(3), 592–606 (2012). https://doi.org/10.1109/TAC.2011.2161027
3. Yuan, K., Ling, Q., Yin, W.: On the convergence of decentralized gradient descent. SIAM J. Optim. **26**(3), 1835–1854 (2016). https://doi.org/10.1137/130943170
4. Xu, J., Zhu, S., Soh, Y.C., Xie, L.: Augmented distributed gradient methods for multi-agent optimization under uncoordinated constant stepsizes. In: 2015 54th IEEE Conference on Decision and Control (CDC), pp. 2055–2060 (2015). https://doi.org/10.1109/CDC.2015.7402509
5. Qu, G., Li, N.: Harnessing smoothness to accelerate distributed optimization. IEEE Trans. Control Netw. Syst. **5**(3), 1245–1260 (2018). https://doi.org/10.1109/TCNS.2017.2698261
6. Qu, G., Li, N.: Accelerated distributed Nesterov gradient descent. IEEE Trans. Autom. Control **65**(6), 2566–2581 (2020). https://doi.org/10.1109/TAC.2019.2937496
7. Tsianos, K.I., Lawlor, S., Rabbat, M.G.: Push-sum distributed dual averaging for convex optimization. In: 2012 IEEE 51st IEEE Conference on Decision and Control (CDC), pp. 5453–5458 (2012). https://doi.org/10.1109/CDC.2012.6426375
8. Nedic, A., Olshevsky, A.: Distributed optimization over time-varying directed graphs. IEEE Trans. Autom. Control **60**(3), 601–615 (2015). https://doi.org/10.1109/TAC.2014.2364096
9. Mai, V.S., Abed, E.H.: Distributed optimization over weighted directed graphs using row stochastic matrix. In: 2016 American Control Conference (ACC), pp. 7165–7170 (2016). https://doi.org/10.1109/ACC.2016.7526803
10. Li, H., Lu, Q., Huang, T.: Distributed projection subgradient algorithm over time-varying general unbalanced directed graphs. IEEE Trans. Autom. Control **64**(3), 1309–1316 (2019). https://doi.org/10.1109/TAC.2018.2849616
11. Xi, C., Khan, U.A.: DEXTRA: a fast algorithm for optimization over directed graphs. IEEE Trans. Autom. Control **62**(10), 4980–4993 (2017). https://doi.org/10.1109/TAC.2017.2672698
12. Xin, R., Khan, U.A.: A linear algorithm for optimization over directed graphs with geometric convergence. IEEE Control Syst. Lett. **2**(3), 315–320 (2018). https://doi.org/10.1109/LCSYS.2018.2834316

13. Saadatniaki, F., Xin, R., Khan, U.A.: Decentralized optimization over time-varying directed graphs with row and column-stochastic matrices. IEEE Trans. Autom. Control **65**(11), 4769–4780 (2020). https://doi.org/10.1109/TAC.2020.2969721
14. Xin, R., Khan, U.A.: Distributed heavy-ball: a generalization and acceleration of first-order methods with gradient tracking. IEEE Trans. Autom. Control **65**(6), 2627–2633 (2020). https://doi.org/10.1109/TAC.2019.2942513
15. Horn, R.A., Johnson, C.R.: Matrix Analysis, 2nd edn. Cambridge University Press, New York (2013)
16. Mo, L., Li, J., Huang, J.: Distributed optimization algorithm for discrete-time heterogeneous multi-agent systems with nonuniform stepsizes. IEEE Access **7**, 87303–87312 (2019). https://doi.org/10.1109/ACCESS.2019.2925414
17. Mo, L., Lin, P.: Distributed continuous-time optimization over second-order multi-agent networks with nonuniform gains. In: 2019 Chinese Control and Decision Conference (CCDC), pp. 35-38 (2019). https://doi.org/10.1109/CCDC.2019.8833262
18. Nesterov, Y.: Introductory Lectures on Convex Optimization. Springer, Boston (2004). https://doi.org/10.1007/978-1-4419-8853-9