SCIENCE CHINA
Information Sciences

• LETTER •

December 2018, Vol. 61 129204:1–129204:3
https://doi.org/10.1007/s11432-018-9502-1

Distributed optimization on unbalanced graphs via continuous-time methods

Zhenhong LI & Zhengtao DING*

School of Electrical and Electronic Engineering, University of Manchester, Manchester M13 9PL, *UK*

Received 25 April 2018/Accepted 24 May 2018/Published online 20 November 2018

Citation Li Z H, Ding Z T. Distributed optimization on unbalanced graphs via continuous-time methods. Sci China Inf Sci, 2018, 61(12): 129204, https://doi.org/10.1007/s11432-018-9502-1

Dear editor,
Distributed optimization problems (DOPs) have attracted significant attention in the past decade, owing to their potential applications in a variety of scenarios such as sensor networks, distributed parameters estimation, and power system economic dispatch. An important class of DOPs refers to minimizing the sum of local objective functions (e.g., [1–3])

$$\min_{\omega\in\mathbb{R}^n} f(\omega) = \min_{\omega\in\mathbb{R}^n} \sum_{i=1}^{N} f_i(\omega), \qquad (1)$$

where N is the number of agents, $f_i : \mathbb{R}^n \to \mathbb{R}$ is the local cost function of agent i, and $f = \sum_{i=1}^{N} f_i(\omega)$ is the global cost function of the network.

To solve problem (1), two continuous-time schemes are designed from a control perspective in [4] to find the optimal solution with centralized and distributed structures, respectively. The distributed scheme achieves asymptotic convergence for constrained optimization problem on directed graphs. For the system with twice differentiable local cost functions, zero-gradient-sum method in [5] achieves exponential convergence if the initial value of states are the optimal solution of local cost functions. To remove the restriction on the initial condition, Lagrangian based algorithms are proposed in [2]. A remarkable feature of Lagrangian based algorithms is the use of auxiliary states which can also be regarded as Lagrangian multipliers. However the algorithms in [2] need to transmit the auxiliary states over the network. To reduce the communication cost, a new Lagrangian based algorithm is designed in [3]. Sufficient conditions are established to guarantee the exponential convergence of the algorithm.

The aforementioned studies require that the communication structures are undirected or at least balanced. Moreover, to eliminate the communication of auxiliary states, the lower bound of local convexity constants are used to establish the convergence of algorithms [3, 6]. Designing optimization algorithms on a more general communication structure and relaxing the assumptions on the local gradients remain as ongoing research issues.

We consider the distributed optimization problem where each agent has a strongly convex cost function with globally Lipschitz gradients. A continuous-time algorithm is presented for unbalanced directed graphs. Sufficient conditions for the convergence are derived based on invariance and Lyapunov stability theory. By introducing a semi-positive definite term to the Lyapunov function and exploring the invariant projection of Laplacian matrix, the requirement of the lower bound of local convexity constants is removed. Finally, we build a experiment on a distributed microcomputer platform to validate the results.

Methodology. Consider a group of N agents. The communication topology among agents is described by the directed graph $\mathcal{G}$. The set of agents is defined as $\mathcal{V} = \{1, \ldots, N\}$. The adjacency matrix is defined as $\mathcal{A} = [\alpha_{ij}] \in \mathbb{R}^{N\times N}$, where $\alpha_{ii} = 0$

* Corresponding author (email: zhengtao.ding@manchester.ac.uk)

and $\alpha_{ij} = 1$ if the i-th agent can get the information from the j-th agent, otherwise $\alpha_{ij} = 0$. The Laplacian matrix $\mathcal{L} = [l_{ij}] \in \mathbb{R}^{N\times N}$ associated with $\mathcal{G}$ is defined as $l_{ii} = \sum_{j=1}^{N} a_{ij}$ and $l_{ij} = -a_{ij}$, for $i \neq j$. A directed graph is strongly connected if there exists a directed path from every agent to every other agent.

Throughout this study, we make Assumptions 1 and 2.

Assumption 1. Each local objective function f_i is strongly convex and differentiable with globally Lipschitz gradient, i.e., there exists $K_i \in \mathbb{R}_{\geqslant 0}$ such that $\| \bigtriangledown f_i(x) - \bigtriangledown f_i(y)\| \leqslant K_i\|x-y\|, \forall x, y \in \mathbb{R}^n$.

With Assumption 1, the global objective function f is also strongly convex, and the solution of problem (1) is unique.

Assumption 2. The communication topology $\mathcal{G}$ is strongly connected.

With Assumption 2 and [7, Lemma 2.1], following similar steps in [1], we transform the problem (1) into a minimization problem under a consensus condition.

$$\min_{x\in\mathbb{R}^{Nn}} \tilde{f}(x) = \sum_{i=1}^{N} f_i(x_i),$$
$$\text{s.t. } (\mathcal{L}\otimes I_n)x = 0_{Nn}, \tag{2}$$

where $x_i \in \mathbb{R}^n$ is the state of i-th agent, x is the aggregated variable of x_i, and 0_{Nn} denotes a column vector of size Nn with all entries equal to zero. In the following, we will introduce the main results of this study.

A continuous-time optimization algorithm is designed as

$$\dot{x} = -\gamma \bigtriangledown \tilde{f}(x) - \alpha\left(\Xi\mathcal{L}\otimes I_n\right)x - v, \tag{3a}$$
$$\dot{v} = \alpha\beta\left(\Xi\mathcal{L}\otimes I_n\right)x, \tag{3b}$$

where $v_i \in \mathbb{R}^n$ is the auxiliary state of i-th agent, v is the aggregated variable of v_i, $\gamma, \alpha, \beta \in \mathbb{R}_{>0}$ are constant gains, $\bigtriangledown\tilde{f}(x) = [\bigtriangledown f_1(x_1)^{\mathrm{T}}, \bigtriangledown f_2(x_2)^{\mathrm{T}}, \ldots, \bigtriangledown f_N(x_N)^{\mathrm{T}}]^{\mathrm{T}}$ is the gradient of $\tilde{f}$, and Ξ is defined in [7, Lemma 2.2].

Lemma 1. Under Assumptions 1 and 2, the equilibrium point of (3) satisfying $(\bar{x}, \bar{v}) \in \mathcal{P}_0(0)$ is an optimal solution of problem (2), where $\mathcal{P}_0(0) = \{(x, v) \in \mathbb{R}^{Nn}\times\mathbb{R}^{Nn} \mid (1_N^{\mathrm{T}}\otimes I_n)v = 0_n\}$.

Proof. Note that the equilibrium point $(\bar{x}, \bar{v})$ of (3) satisfies

$$\bar{x} = 1_N\otimes a, \quad \forall a \in \mathbb{R}^n,$$
$$\bar{v} = -\gamma \bigtriangledown \tilde{f}(\bar{x}).$$

Applying $(1_N^{\mathrm{T}}\otimes I_n)\bar{v} = -\gamma\sum_{i=1}^{N} f_i(\bar{x}) = 0_n$, we have that $(\bar{x}, \bar{v})$ is the optimal solution $(x^\star, v^\star)$.

Theorem 1. Under Assumptions 1 and 2, algorithm (3) solves the distributed optimization problem (2) for $(x(0), v(0)) \in \mathcal{P}_0(0)$, if $\alpha, \beta, \gamma \in \mathbb{R}_{>0}$ satisfy

$$(\delta+1)\gamma\beta - 2\gamma^2\bar{K} > 0, \tag{4a}$$
$$(2\delta+1)\lambda_2(\bar{\mathcal{L}})\alpha\beta - \frac{17}{2}(\delta+1)^2\beta^2 > 0, \tag{4b}$$

where $\delta \in \mathbb{R}_{>0}$, $\bar{K} = \max\{K_1, K_2, \ldots, K_N\}$, $\bar{\mathcal{L}} = \Xi\mathcal{L} + \mathcal{L}^{\mathrm{T}}\Xi$ and $\lambda_2(\bar{\mathcal{L}})$ denotes the smallest nonzero eigenvalues of $\bar{\mathcal{L}}$.

Proof. Define $\rho = x - \bar{x}$, $\varrho = v - \bar{v}$. We can get the network dynamics

$$\dot{\rho} = -\gamma h - \alpha\left(\Xi\mathcal{L}\otimes I_n\right)\rho - \varrho,$$
$$\dot{\varrho} = \alpha\beta\left(\Xi\mathcal{L}\otimes I_n\right)\rho,$$

where $h = \bigtriangledown\tilde{f}(x) - \bigtriangledown\tilde{f}(\bar{x})$.

Consider the following Lyapunov function candidate

$$V_2 = \frac{1}{2}\rho^{\mathrm{T}}\left(\left((\delta+1)\beta\Pi + \delta\beta I_N\right)\otimes I_n\right)\rho + \frac{1}{2\beta}(\beta\rho+\varrho)^{\mathrm{T}}(\beta\rho+\varrho),$$

where $\Pi = I_N - \frac{1}{N}1_N 1_N^{\mathrm{T}}$.

The time derivative of V_2 along (3) is given by

$$\begin{aligned}\dot{V}_2 = &-(\delta+1)\gamma\beta\rho^{\mathrm{T}}h - (\delta+1)\gamma\beta\rho^{\mathrm{T}}\left(\Pi\otimes I_n\right)h \\ &- \frac{(2\delta+1)}{2}\alpha\beta\rho^{\mathrm{T}}\left(\bar{\mathcal{L}}\otimes I_n\right)\rho - (\delta+1)\beta\rho^{\mathrm{T}}\varrho \\ &- (\delta+1)\beta\rho^{\mathrm{T}}\left(\Pi\otimes I_n\right)\varrho - \gamma\varrho^{\mathrm{T}}h - \varrho^{\mathrm{T}}\varrho.\end{aligned} \tag{5}$$

Note that $\mathcal{P}_0(0)$ is positive invariant under (3). Furthermore, for $(x, v) \in \mathcal{P}_0(0)$, we have $\rho^{\mathrm{T}}\varrho = \rho^{\mathrm{T}}\left(\Pi\otimes I_n\right)\varrho$. Using this fact, we can rewrite (5) as follows:

$$\begin{aligned}\dot{V}_2 = &-(\delta+1)\gamma\beta\rho^{\mathrm{T}}h - (\delta+1)\gamma\beta\rho^{\mathrm{T}}\left(\Pi\otimes I_n\right)h \\ &- 2(\delta+1)\beta\rho^{\mathrm{T}}\left(\Pi\otimes I_n\right)\varrho - \gamma\varrho^{\mathrm{T}}h - \varrho^{\mathrm{T}}\varrho \\ &- \frac{(2\delta+1)}{2}\alpha\beta\rho^{\mathrm{T}}\left(\bar{\mathcal{L}}\otimes I_n\right)\rho \\ \leqslant &-\left((\delta+1)\gamma\beta - 2\gamma^2\bar{K}\right)\rho^{\mathrm{T}}h - \frac{1}{2}\varrho^{\mathrm{T}}\varrho \\ &- \left\|2(\delta+1)\beta\left(\Pi\otimes I_n\right)\rho + \frac{1}{2}\varrho\right\|^2 - \left\|\gamma h + \frac{1}{2}\varrho\right\|^2 \\ &- \left\|\frac{1}{2}(\delta+1)\beta\left(\Pi\otimes I_n\right)\rho + \gamma h\right\|^2 \\ &- \rho^{\mathrm{T}}\left(\left(\frac{(2\delta+1)}{2}\alpha\beta\bar{\mathcal{L}} - \frac{17}{4}(\delta+1)^2\beta^2\Pi\right)\otimes I_n\right)\rho.\end{aligned}$$

The last equality follows the facts that $\Pi\Pi = \Pi$ and $h^{\mathrm{T}}h \leqslant \bar{K}\rho^{\mathrm{T}}h$.

Because $\bar{\mathcal{L}}\Pi = \Pi\bar{\mathcal{L}}$, by [8, Theorem 4.1.6], there exists an orthogonal matrix $U \in \mathbb{R}^{N\times N}$ such that

$$\frac{(2\delta+1)}{2}\alpha\beta\bar{\mathcal{L}} - \frac{17}{4}(\delta+1)^2\beta^2\Pi$$

$$= U\left(\frac{(2\delta+1)}{2}\alpha\beta\begin{bmatrix}0 & 0 & \cdots & 0\\ 0 & \lambda_2(\bar{\mathcal{L}}) & \cdots & 0\\ \vdots & \vdots & & \vdots\\ 0 & 0 & \cdots & \lambda_N(\bar{\mathcal{L}})\end{bmatrix} - \frac{17}{4}(\delta+1)^2\beta^2\begin{bmatrix}0 & 0 & \cdots & 0\\ 0 & 1 & \cdots & 0\\ \vdots & \vdots & & \vdots\\ 0 & 0 & \cdots & 1\end{bmatrix}\right)U^{\mathrm{T}}. \quad (6)$$

Applying $\rho^{\mathrm{T}}h \geqslant 0$, (4) and (6), we have $\dot{V}_2 \leqslant 0$. Therefore, we can conclude that the variables ρ and ϱ are bounded. From LaSalle's invariance principle, we have $\lim_{t\to\infty}\rho^{\mathrm{T}}h = 0$ and $\lim_{t\to\infty}\varrho = 0_{Nn}$, which implies that x_i converges to the optimal solution of problem (2) as $t \to \infty$, for $i = 1, \ldots, N$.

There always exist $\alpha, \beta, \gamma \in \mathbb{R}_{>0}$ satisfying (4); e.g., by choosing $\beta, \gamma, \delta \in \mathbb{R}_{>0}$ satisfying (4a), we can find sufficiently large α to have (4b). The tunning of α and γ is decoupled.

Experiment validation. A microcomputer platform is used to validate the design of algorithm (3). The microcomputer platform shown in Figure 1(a) consists of a router and 5 microcomputers. Each microcomputer has an onboard processor Cortex-A53 running at 1.2 GHz and a micro-SD card. The router is used to set up the wireless communication among microcomputers, each of which can get the state information of its neighbors through an 802.11n wireless LAN that is provided by the onboard chip BCM43438.

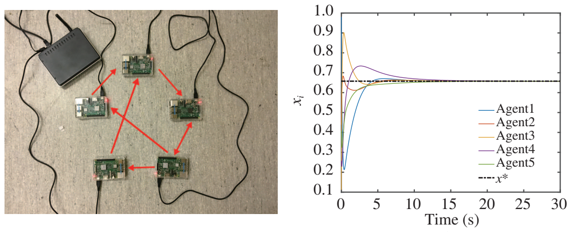
Consider a network of 5 agents with local cost functions given by

$$\begin{aligned} &f_1 = x^{\frac{4}{3}}, \quad f_2 = \mathrm{e}^{0.2x},\\ &f_3 = (x+2)^2, \quad f_4 = 0.1x^2 + \frac{x^2}{\sqrt{x^2+1}},\\ &f_5 = x^2 + \ln(x^2+1), \end{aligned} \quad (7)$$

where $x \in \mathbb{R}$. The initial states $x_i(0)$, $i = 1, 2, \ldots, 5$ are randomly selected within $[0.11, 1]$. Then we can calculate the optimal solution $x^\star = 0.6575$. To satisfy $(x(0), v(0)) \in \mathcal{P}_0(0)$, $v_i(0)$ are set as 0, $\forall i = 1, 2, \ldots, 5$. The parameters are chosen as $\alpha = 6$, $\beta = 1$, and $\gamma = 1$. The communication structure $\mathcal{G}$ is shown in Figure 1(a).

To implement the algorithms (3) on hardware platform, the gradients $\bigtriangledown f_i$, which include terms $x^{\frac{1}{3}}$, $\sqrt{x^2+1}$ and $\mathrm{e}^{0.2x}$, are approximated by a Newton iterative method and Taylor series. The accuracy of approximation is set as 10^{-5}. The integrations are calculated using a forward Euler method. Furthermore the frequency of algorithm (3) is 100 Hz.

The result in Figure 1(b) shows that all the trajectories of states converge to the optimal solution, implying that (3) can be implemented on embedded systems with limited computation capability.



(a) (b)

Figure 1 (Color online) (a) Communication structure $\mathcal{G}$; (b) the agent states x_i, $i = 1, 2, \ldots, 5$.

Conclusion. We consider the DOP on unbalanced directed graphs. Sufficient conditions for the convergence are established without the knowledge of the lower bound of local convexity constants. The experiment results show that our algorithm can be implemented on embedded systems with limited computation capability.

Acknowledgements This work was supported by Science and Technology Facilities Council (Grant No. ST/N006852/1).

References

1 Li Z H, Ding Z T, Sun J Y, et al. Distributed adaptive convex optimization on directed graphs via continuous-time algorithms. IEEE Trans Autom Control, 2018, 63: 1434–1441
2 Gharesifard B, Cortes J. Distributed continuous-time convex optimization on weight-balanced digraphs. IEEE Trans Autom Control, 2014, 59: 781–786
3 Kia S S, Cortés J, Martínez S. Distributed convex optimization via continuous-time coordination algorithms with discrete-time communication. Automatica, 2015, 55: 254–264
4 Wang J, Elia N. A control perspective for centralized and distributed convex optimization. In: Proceedings of the 50th IEEE Conference on Decision and Control and European Control Conference, Orlando, 2011. 3800–3805
5 Lu J, Tang C Y. Zero-gradient-sum algorithms for distributed convex optimization: the continuous-time case. IEEE Trans Autom Control, 2012, 57: 2348–2354
6 Yang S F, Liu Q S, Wang J. Distributed optimization based on a multiagent system in the presence of communication delays. IEEE Trans Syst Man Cybern Syst, 2017, 47: 717–728
7 Mei J, Ren W, Chen J. Distributed consensus of second-order multi-agent systems with heterogeneous unknown inertias and control gains under a directed graph. IEEE Trans Autom Control, 2016, 61: 2019–2034
8 Horn R A, Johnson C R. Matrix Analysis. Cambridge: Cambridge University Press, 1985