SCIENCE CHINA
Information Sciences

• LETTER •

September 2026, Vol. 69, Iss. 9, 199203:1–199203:2
https://doi.org/10.1007/s11432-025-4848-2

Distributed optimization algorithm with superlinear convergence rate

Yeming XU[1], Ziyuan GUO[1], Kaihong LU[1], Hongxia WANG[1] & Huanshui ZHANG[1,2*]

[1]*School of Electrical and Automation Engineering, Shandong University of Science and Technology, Qingdao 266590, China*
[2]*School of Control Science and Engineering, Shandong University, Jinan 250061, China*

Received 21 October 2025/Revised 31 January 2026/Accepted 3 March 2026/Published online 15 April 2026

Citation Xu Y M, Guo Z Y, Lu K H, et al. Distributed optimization algorithm with superlinear convergence rate. Sci China Inf Sci, 2026, 69(9): 199203, https://doi.org/10.1007/s11432-025-4848-2

We consider a consensus optimization problem in which n agents in a connected network collaboratively solve

$$\min_{y\in\mathbb{R}^p} f(y) \triangleq \sum\nolimits_{i=1}^{n} f_i(y), \tag{1}$$

where $f:\mathbb{R}^p \mapsto \mathbb{R}$ is the objective function, where each agent i holds a private local objective $f_i:\mathbb{R}^p\to\mathbb{R}$ and communicates only with neighbors. Problem (1) has been widely studied in recent years for its broad applicability to sensor networks, distributed control, and machine learning.

Common approaches for solving problem (1) typically use only first-order information, which makes them simple and computationally efficient. However, this limits their convergence to at most linear [1]. Second-order methods are attractive for distributed optimization due to their faster convergence. However, the inverse of the Hessian matrix is typically dense, which limits the direct application of second-order methods in fully distributed settings. The network Newton (NN) algorithm approximates the Newton step by truncating the Taylor series of the Hessian inverse [2]. Qu et al. [3] developed two second-order methods, but also used local Hessian inversions and only achieved linear convergence. The dual inexact nonsmooth Newton method guarantees superlinear convergence [4], but it requires solving an internal optimization at each step, increasing computational cost. Existing methods still face limitations in convergence speed or computational complexity, highlighting the need for more flexible strategies to improve algorithm performance. We briefly summarize the motivation of this study here, and a more comprehensive literature review is given in Appendix A.

Based on the above discussion, this study proposes a new second-order distributed optimization algorithm based on optimal control (DOAOC). We first reformulate problem (1) using a penalty method, and then construct the DOAOC by combining the optimal control principle (OCP) method [5] and replacing the inverse of the Hessian matrix. The main contributions of this study can be summarized as follows. (i) The DOAOC incorporates second-order information of the objective functions through a special optimization structure, revealing a new way to utilize second-order information without explicitly computing the Hessian inverse. (ii) Different from [1], which requires all the information from the other agents for optimization, the proposed DOAOC uses only neighbor information, which eliminates the need for a central coordinator or global consensus variable. (iii) Our theoretical analysis shows that the algorithm has a superlinear convergence rate. In contrast to prior studies [2, 3] that establish only linear convergence for second-order methods, the achieved convergence rate is much faster. (iv) By limiting the number of communication rounds among agents to K during each iteration, we can obtain a variant of the DOAOC, named the DOAOC-K, which further reduces the communication cost.

Problem reformulation. The communication network among nodes is described by an undirected graph $\mathcal{G}=(\mathcal{V},\mathcal{E})$, where $\mathcal{V}=\{1,\ldots,n\}$ denotes the set of nodes and $\mathcal{E}\subseteq\mathcal{V}\times\mathcal{V}$ is the set of edges. For $(i,j)\in\mathcal{E}$, nodes i and j can directly exchange messages, and $\mathcal{N}_i$ denotes the neighbor set of node i. The weight matrix of $\mathcal{G}$ is a symmetric, doubly stochastic $n\times n$ matrix W with elements $w_{ij}\geqslant 0$. The weight $w_{ij}=0$ if and only if $j\notin\mathcal{N}_i\cup\{i\}$, and $w_{ii}=1-\sum_{j\in\mathcal{N}_i\setminus\{i\}} w_{ij}>0$.

We employ a penalty coefficient λ to embed the consensus constraints as a penalty term inside the objective function. The penalty reformulation of problem (1) is

$$\min_{x\in\mathbb{R}^{np}} F(x) \triangleq h(x) + \frac{1}{2\lambda} x^\top (I_{np}-Z)x, \tag{2}$$

where $F:\mathbb{R}^{np}\mapsto\mathbb{R}$ is the penalty objective function, $x=(x^1,\ldots,x^n)\in\mathbb{R}^{np}$, $x^i\in\mathbb{R}^p$ is the local decision vector of agent i, $h(x)=\sum_{i=1}^n f_i(x^i)$, and $Z=W\otimes I_p\in\mathbb{R}^{np\times np}$ is the Kronecker product of W and I. This study aims to develop a distributed optimization algorithm to find the solution x_* of problem (2). For clarity and rigor in the subsequent analysis, the assumptions below are introduced.

Assumption 1. The local objective functions $f_i,\ i=1,\ldots,n$, are assumed to be twice continuously differentiable, and their Hessian matrices satisfy $mI_p\preceq\nabla^2 f_i(y)\preceq MI_p$ for some constants $0<m\leqslant M<\infty$ and all $y\in\mathbb{R}^p$.

Assumption 2. The Hessian matrices $\nabla^2 f_i,\ i=1,\ldots,n$, are L-Lipschitz continuous, i.e., for any $y,\hat{y}\in\mathbb{R}^p$, there exists a constant $L>0$ such that $||\nabla^2 f_i(y)-\nabla^2 f_i(\hat{y})||\leqslant L||y-\hat{y}||$.

For the notations, the explanations of the above assumptions,

* Corresponding author (email: hszhang@sdu.edu.cn)

and the discussion on the relationship between problems (1) and (2), please refer to Appendix B.

Distributed optimization via OCP method. We develop a novel distributed optimization algorithm that attains a superlinear convergence rate. Firstly, we review the OCP method. A general iteration for solving optimization problem (2) is given by

$$x_{k+1} = x_k - g_k(x_k), \tag{3}$$

where $g_k(x_k) \in \mathbb{R}^{np}$ is the update applied to the current optimization variable $x_k \in \mathbb{R}^{np}$. The changes of optimization variable x_k in (3) can be understood from the perspective of a discrete-time linear time-invariant system. Based on this, Ref. [5] proposed the OCP method, in which the update $g_k(x_k)$ is computed using

$$g_t(x_k) = \big(R + \nabla^2 F(x_k)\big)^{-1}\big(\nabla F(x_k) + R g_{t-1}(x_k)\big), \tag{4}$$

where $t = 1, \ldots, k$, $g_0(x_k) = \big(R + \nabla^2 F(x_k)\big)^{-1} \nabla F(x_k)$, and $R \in \mathbb{S}_{++}^{np}$ is an adjustable matrix. With an appropriate choice of R, the method can be shown to converge superlinearly while maintaining iteration stability.

Remark 1. Since the computation of $(R + \nabla^2 F(x_k))^{-1}$ cannot be distributed, the OCP method is confined to a centralized setting. In addition, computing the inverse of $R + \nabla^2 F(x_k)$ greatly increases the computational burden. To overcome these challenges, we develop a distributed optimization algorithm by modifying (4).

We replace $(R + \nabla^2 F(x_k))^{-1}$ in (4) with a scalar matrix ηI_{np}, where η is an adjustable parameter. Due to the relationship $(R + \nabla^2 F(x_k))^{-1} R = I_{np} - (R + \nabla^2 F(x_k))^{-1} \nabla^2 F(x_k)$, the matrix $(R + \nabla^2 F(x_k))^{-1} R$ can be substituted with $I_{np} - \eta \nabla^2 F(x_k)$. For this, we propose the DOAOC for solving problem (2), as outlined in the following algorithm:

$$x_{k+1} = x_k - \hat{g}_k(x_k), \tag{5}$$

$$\hat{g}_t(x_k) = \eta \nabla F(x_k) + \big(I_{np} - \eta \nabla^2 F(x_k)\big)\hat{g}_{t-1}(x_k), \tag{6}$$

where $t = 1, \ldots, k$ and $\hat{g}_0(x_k) = \eta \nabla F(x_k)$. The DOAOC uses the second-order information $\nabla^2 F(x_k)$, which is as sparse as the network, allowing for its distributed implementation. Appendix C.1 explains in more detail why the proposed method can be implemented in a distributed manner. The subsequent theorem establishes the convergence rate of DOAOC.

Theorem 1. Under Assumption 1, if $\{x_k\}$ generated by the DOAOC converges and η satisfies $0 < \eta < \frac{2}{a}$, where $a = M + \frac{2(1 - w_{\min})}{\lambda}$ and $w_{\min} = \min_i\{w_{ii}\}$, then $\{x_k\}$ converges superlinearly to x_*.

The choice of η is related to the penalty coefficient λ, revealing a fundamental trade-off between convergence rate and solution accuracy. When λ is large, the sequence generated by DOAOC converges more rapidly to the solution of problem (2), but this comes at the cost of enlarging the discrepancy between problems (1) and (2). In contrast, a smaller λ tends to narrow this gap, but it also imposes constraints on the choice of η, thereby slowing down the convergence rate of DOAOC. Under the requirement of a given solution accuracy while ensuring stable convergence, the step size η can be chosen as large as possible to reduce the number of iterations. The following remark is provided to better illustrate the connection between DOAOC and existing methods.

Remark 2. The decentralized gradient descent is included in DOAOC as a special case if we only use $\hat{g}_0(x_k)$ to update x_k. Furthermore, if η is chosen appropriately, then $\hat{g}_t(x_k)$ converges to the Newton step as $t \to \infty$.

The details of Remark 2 are provided in Appendix C.2, while the proofs of the convergence and convergence rate of the DOAOC are presented in Appendix D.

The variant of DOAOC and its convergence analysis. To further reduce the number of communications, the DOAOC-K is given as

$$x_{k+1} = x_k - \hat{g}_{K-1}(x_k), \tag{7}$$

$$\hat{g}_t(x_k) = \eta \nabla F(x_k) + \big(I_{np} - \eta \nabla^2 F(x_k)\big)\hat{g}_{t-1}(x_k), \tag{8}$$

where $t = 1, \ldots, K-1$, $\hat{g}_0(x_k) = \eta \nabla F(x_k)$ and K is a given positive integer. The parameter K represents the restricted number of communications among agents in each iteration. For $K = 1$, meaning that agent i only needs to transmit x_k^i to its neighbors at each iteration, the DOAOC-K reduces to the decentralized gradient descent algorithm. For $K = 2, \ldots, k$, agent i should additionally transmit $\hat{g}_0^i, \ldots, \hat{g}_{k-2}^i$ to its neighbors at each iteration.

We prove that by using DOAOC-K, the function value $F(x_k)$ gradually converges to $F(x_*)$ as the iterations progress.

Theorem 2. Suppose Assumptions 1 and 2 hold, and let $\{x_k\}$ be generated by the DOAOC-K algorithm. If the step size satisfies $\eta < \min\Big\{1, \frac{1}{a}, \frac{m}{a^2 K^2}, \Big[\frac{6m^5}{a^4 (2a)^{\frac{2}{3}} K^3 L (F(x_0) - F(x_*))^{\frac{1}{2}}}\Big]^{\frac{1}{2}}\Big\}$, then $\{F(x_k)\}$ converges to $F(x_*)$ at least linearly as

$$F(x_k) - F(x_*) \leqslant (1 - \bar{\epsilon})^k (F(x_0) - F(x_*)), \tag{9}$$

where the constant $0 < \bar{\epsilon} < 1$ is given by $\bar{\epsilon} = \frac{2m^2\eta - ma^2\eta^2K^2}{a} - \frac{a^3 (2a)^{\frac{3}{2}} \eta^3 K^3 L}{6m^3}(F(x_0) - F(x_*))^{\frac{1}{2}}$.

The proof of Theorem 2 can be found in Appendix E. As shown in (9), $F(x_k)$ exhibits at least linear convergence toward $F(x_*)$ when communication rounds between agents are restricted. It is clear that a smaller λ leads to a larger value of a, which in turn reduces $\bar{\epsilon}$. Consequently, this reduction in $\bar{\epsilon}$ slows down the convergence rate of DOAOC-K. The parameter K should be flexibly adjusted according to the specific requirements of different problems to achieve favorable convergence behavior.

Numerical simulation. We offer the pseudocodes of the proposed algorithms and compare them with other methods. The results of the numerical simulation can be observed in Appendix F.

Conclusion. This study proposes a novel distributed optimization algorithm called DOAOC, which integrates the penalty method interpretation of distributed optimization with the advantages of the OCP method, effectively incorporating second-order information without requiring the Hessian matrix inverse. A connection is established between DOAOC and traditional optimization methods. To reduce the communication burden, we propose a variant of DOAOC, referred to as DOAOC-K. The convergence properties of DOAOC and DOAOC-K are analyzed, and their superior performance is validated through numerical simulations.

Acknowledgements This work was supported by Original Exploratory Program Project of National Natural Science Foundation of China (Grant No. 62450004), Joint Funds of National Natural Science Foundation of China (Grant No. U23A20325), and Natural Science Foundation of Shandong Province (Grant No. ZR2024MF045).

Supporting information Appendixes A–F. The supporting information is available online at info.scichina.com and link.springer.com. The supporting materials are published as submitted, without typesetting or editing. The responsibility for scientific accuracy and content remains entirely with the authors.

References

1 Zhang J, You K, Başar T. Distributed adaptive Newton methods with global superlinear convergence. Automatica, 2022, 138: 110156
2 Mokhtari A, Ling Q, Ribeiro A. Network Newton distributed optimization methods. IEEE Trans Signal Process, 2016, 65: 146–161
3 Qu Z, Li X, Li L, et al. Linearly convergent second-order distributed optimization algorithms. IEEE Trans Automat Contr, 2024, 69: 5431–5438
4 Niu D, Hong Y, Song E. A dual inexact nonsmooth Newton method for distributed optimization. IEEE Trans Signal Process, 2025, 73: 188–203
5 Zhang H S, Wang H X, Xu Y, et al. Optimization methods rooted in optimal control. Sci China Inf Sci, 2024, 67: 222208