Sublinear and Linear Convergence of Modified ADMM for Distributed Nonconvex Optimization

Xinlei Yi, Shengjun Zhang, Tao Yang, *Senior Member, IEEE*, Tianyou Chai, *Life Fellow, IEEE*, and Karl Henrik Johansson, *Fellow, IEEE*

***Abstract*—In this article, we consider distributed nonconvex optimization over an undirected connected network. Each agent can only access to its own local nonconvex cost function and all agents collaborate to minimize the sum of these functions by using local information exchange. We first propose a modified alternating direction method of multipliers (ADMM) algorithm. We show that the proposed algorithm converges to a stationary point with the sublinear rate $\mathcal{O}(1/T)$ if each local cost function is smooth and the algorithm parameters are chosen appropriately. We also show that the proposed algorithm linearly converges to a global optimum under an additional condition that the global cost function satisfies the Polyak–Łojasiewicz condition, which is weaker than the commonly used conditions for showing linear convergence rates including strong convexity. We then propose a distributed linearized ADMM (L-ADMM) algorithm, derived from the modified ADMM algorithm, by linearizing the local cost function at each iteration. We show that the L-ADMM algorithm has the same convergence properties as the modified ADMM algorithm under the same conditions. Numerical simulations are included to verify the correctness and efficiency of the proposed algorithms.**

***Index Terms*—Alternating direction method of multipliers (ADMM), distributed optimization, linear convergence, linearized ADMM, Polyak–Łojasiewicz condition.**

I. Introduction

CONSIDER a group of n agents that are connected via a communication network. Each agent is associated with a local (possibly nonconvex) cost function $f_i(x)$, where $x \in \mathbb{R}^p$ is the decision variable and p is its dimension. The local cost function f_i is known to agent i only. By exchanging information with their neighbors through the underlying communication network, all agents collaborate to solve the following unconstrained optimization problem:

$$\min_{x\in\mathbb{R}^p} f(x) = \frac{1}{n}\sum_{i=1}^{n} f_i(x). \tag{1}$$

This is the distributed nonconvex optimization problem. It is a fundamental component of distributed decision-making and has a wide range of applications, for example, power allocation in wireless *ad hoc* networks [1], distributed clustering [2], compressed sensing [3], dictionary learning [4], and empirical risk minimization [5]. Various algorithms have been proposed to solve (1); see, e.g., [1], [4], [6]–[16]. The convergence properties have also been analyzed. For instance, in [12], [15], and [16], it was shown that the first-order stationary point can be found with the sublinear convergence rate $\mathcal{O}(1/T)$ when each local cost function is smooth, where T is the total number of iterations; in [9], [10], and [13], it was shown that the second-order stationary points can be found under additional assumptions, such as Lipschitz-continuous Hessian and/or a suitably chosen initialization; in [16], it was shown that the global optima can be found linearly if the global cost function satisfies the Polyak–Łojasiewicz (P–Ł) condition.

We are interested in proposing the alternating direction method of multipliers (ADMM) method to solve (1). The ADMM is very effective at numerically solving many practical convex and nonconvex optimization problems [17]–[19]. However, existing distributed ADMM algorithms with provable convergence analysis to solve (1) normally require that cost functions are convex or the communication network is a star graph, i.e., hub/leaf topology. If cost functions are convex, many distributed ADMM algorithms have been proposed to solve (1); see, e.g., [20]–[32]. The convergence property of these algorithms has also been analyzed. For instance, the $\mathcal{O}(1/T)$ and the linear convergence rates were established in [20], [21], and [27] and [22]–[24], [26], [28], and [29], respectively. If the communication network is a star graph, the authors of [33]–[35] proposed distributed ADMM algorithms and proved that the first-order stationary points can be found with the sublinear convergence rate $\mathcal{O}(1/T)$ when each local cost function is smooth. One advantage of these algorithms is that they are

Manuscript received 25 April 2021; revised 25 September 2021 and 27 December 2021; accepted 1 May 2022. Date of publication 27 June 2022; date of current version 13 March 2023. This work was supported in part by the Knut and Alice Wallenberg Foundation, in part by the Swedish Foundation for Strategic Research, in part by the Swedish Research Council, in part by the National Natural Science Foundation of China under Grant 62133003, Grant 61991403, Grant 61991404, and Grant 61991400, and in part by the 2020 Science and Technology Major Project of Liaoning Province under Grant 2020JH1/10100008. Recommended by Associate Editor C. Altafini. *(Corresponding author: Tao Yang.)*

Xinlei Yi and Karl Henrik Johansson are with the Division of Decision and Control Systems, School of Electrical Engineering and Computer Science, Digital Futures, KTH Royal Institute of Technology, 10044 Stockholm, Sweden (e-mail: xinleiy@kth.se; kallej@kth.se).

Shengjun Zhang is with the Department of Electrical Engineering, University of North Texas, Denton, TX 76203 USA (e-mail: shengjunzhang@my.unt.edu).

Tao Yang and Tianyou Chai are with the State Key Laboratory of Synthetical Automation for Process Industries, Northeastern University, Shenyang 110819, China (e-mail: taoyang.work@gmail.com; tychai@mail.neu.edu.cn).

Digital Object Identifier 10.1109/TCNS.2022.3186653

asynchronous. However, in addition to the star graph restriction, the algorithms proposed in [33] and [34] require that each leaf agent communicates both primal and dual variables to the hub agent. Moreover, the algorithm proposed in [35] is based on the standard master/worker model. Specifically, the master (hub agent) executes all of the updating, while each worker (leaf agent) only computes the gradient of its own local cost function and sends it to the master. In other words, all decisions are made by a single agent, the master, which suffers from a single point of failure, high communication, and computation cost, etc. To the best of our knowledge, the distributed proximal primal-dual algorithm (Prox-PDA) proposed in [36], which is a generalization of the distributed ADMM algorithms proposed in [22] and [29], is the only distributed ADMM algorithm with provable convergence analysis to solve (1) when cost functions are nonconvex and the communication network is arbitrarily connected. Through a lower bounded potential function, it was shown that the Prox-PDA finds a first-order stationary point with the sublinear convergence rate $\mathcal{O}(1/T)$ when each local cost function is smooth. To the best of our knowledge, there are no existing results to guarantee that the global optima can be found by ADMM algorithms when cost functions are nonconvex.

In this article, we first propose a modified ADMM algorithm to solve the nonconvex optimization problem (1), which is modified from the classic ADMM algorithm. We have the following contributions.

1) The proposed modified ADMM algorithm is suitable for arbitrarily undirected connected communication networks, not necessarily a star graph.
2) When each local cost function is smooth, we appropriately choose the algorithm parameters and construct a nonnegative potential function. With this nonnegative potential function, we show that the proposed algorithm can find a first-order stationary point with the well-known sublinear convergence rate $\mathcal{O}(1/T)$.
3) If the global cost function satisfies the P–Ł condition in addition, with the same algorithm parameters and potential function, we show that not only the modified ADMM algorithm can find a global optimum but also its convergence rate is linear, which is our main contribution. The P–Ł condition is weaker than the strong convexity condition assumed in [22]–[24], [26], [28], and [29] since it does not require convexity and the global minimizer is not necessarily unique or finite. To the best of our knowledge, the proposed distributed ADMM algorithm is the first ADMM algorithm with provable convergence rate analysis to find the global optima of nonconvex cost function. The closely related studies, e.g., [10], [33]–[36], used lower bounded potential functions to only show that their algorithms can find a stationary point sublinearly at a rate $\mathcal{O}(1/T)$, but they did not consider the scenario when the global cost function satisfies the P–Ł condition. It is unclear whether those lower bounded potential functions can be used or the analysis can be extended to show linear convergence under the P–Ł condition or not.

Note that the modified ADMM algorithm has the same potential drawback as existing distributed ADMM algorithms, such as [20]–[22], [24], [25], [28]–[35], i.e., each agent has to solve a local optimization problem at each iteration, which results in a heavy computational burden to each agent. To tackle this potential drawback, we then propose a distributed linearized ADMM (L-ADMM) algorithm, derived from the proposed distributed ADMM algorithm by linearizing the local cost function at each iteration. As a result, in the proposed distributed L-ADMM, the explicit closed-form solution to each local optimization problem is available. We show that the proposed distributed L-ADMM algorithm has the same convergence properties as the proposed distributed ADMM algorithm under the same conditions.

The rest of this article is organized as follows. Section II introduces some preliminaries. Sections III and IV provide the distributed ADMM and L-ADMM algorithms, respectively, and present their convergence properties. Simulations are given in Section V. Finally, Section VI concludes this article.

Notations: $[n]$ denotes the set $\{1,\dots,n\}$ for any positive integer n. $\mathrm{col}(z_1,\dots,z_k)$ is the concatenated column vector of vectors $z_i\in\mathbb{R}^{p_i},\ i\in[k]$. $\mathbf{1}_n$ ($\mathbf{0}_n$) denotes the column one (zero) vector of dimension n. $\mathbf{I}_n$ is the n-dimensional identity matrix. Given a vector $[x_1,\dots,x_n]^\top\in\mathbb{R}^n$, $\mathrm{diag}([x_1,\dots,x_n])$ is a diagonal matrix with the ith diagonal element being x_i. The notation $A\otimes B$ denotes the Kronecker product of matrices A and B. $\mathrm{null}(A)$ is the null space of matrix A. Given two symmetric matrices M and N, $M\geq N$ means that $M-N$ is positive semidefinite. $\rho(\cdot)$ stands for the spectral radius for matrices and $\rho_2(\cdot)$ indicates the minimum positive eigenvalue for matrices having positive eigenvalues. $\|\cdot\|$ represents the Euclidean norm for vectors or the induced two-norm for matrices. For any square matrix A, denote $\|x\|_A^2=x^\top Ax$. Given a differentiable function f, ∇f denotes the gradient of f. $\mathcal{R}$

II. Preliminaries

In this section, we present some definitions from algebraic graph theory, smooth functions, and the P–Ł condition.

A. Algebraic Graph Theory

Let $\mathcal{G}=(\mathcal{V},\mathcal{E},A)$ denote a weighted undirected graph with the set of vertices (nodes) $\mathcal{V}=[n]$, the set of links (edges) $\mathcal{E}\subseteq\mathcal{V}\times\mathcal{V}$, and the weighted adjacency matrix $A=A^\top=(a_{ij})$ with non-negative elements a_{ij}. A link of $\mathcal{G}$ is denoted by $(i,j)\in\mathcal{E}$ if $a_{ij}>0$, i.e., if vertices i and j can communicate with each other. It is assumed that $a_{ii}=0$ for all $i\in[n]$. Let $\mathcal{N}_i=\{j\in[n]:\ a_{ij}>0\}$ and $\deg_i=\sum_{j=1}^n a_{ij}$ denote the neighbor set and weighted degree of vertex i, respectively. The degree matrix of graph $\mathcal{G}$ is $\mathrm{Deg}=\mathrm{diag}([\deg_1,\dots,\deg_n])$. The Laplacian matrix is $L=(L_{ij})=\mathrm{Deg}-A$. A path of length k between vertices i and j is a subgraph with distinct vertices $i_0=i,\dots,i_k=j\in[n]$ and edges $(i_j,i_{j+1})\in\mathcal{E},\ j=0,\dots,k-1$. An undirected graph is connected if there exists at least one path between any two distinct vertices. The star graph is a special undirected graph, in which there is one and only one agent (hub agent) that connects to all of the rest agents (leaf agents) and each leaf agent only connects to the hub agent.

For a connected undirected graph, we have the following results.

Lemma 1: ([37, Lemmas 1 and 2]) Let L be the Laplacian matrix associated with a connected undirected graph $\mathcal{G}$ and $K_n = \mathbf{I}_n - \frac{1}{n}\mathbf{1}_n\mathbf{1}_n^\top$. Then L and K_n are positive semidefinite, $\mathrm{null}(L) = \mathrm{null}(K_n) = \{\mathbf{1}_n\}$, $L \le \rho(L)\mathbf{I}_n$, $\rho(K_n) = 1$

$$K_n L = L K_n = L \tag{2}$$

$$0 \le \rho_2(L) K_n \le L \le \rho(L) K_n. \tag{3}$$

Moreover, there exists an orthogonal matrix $[r\ R] \in \mathbb{R}^{n\times n}$ with $r = \frac{1}{\sqrt{n}}\mathbf{1}_n$ and $R \in \mathbb{R}^{n\times(n-1)}$ such that

$$R\Lambda_1^{-1}R^\top L = LR\Lambda_1^{-1}R^\top = K_n \tag{4}$$

$$\frac{1}{\rho(L)}K_n \le R\Lambda_1^{-1}R^\top \le \frac{1}{\rho_2(L)}K_n \tag{5}$$

where $\Lambda_1 = \mathrm{diag}([\lambda_2, \dots, \lambda_n])$ with $0 < \lambda_2 \le \dots \le \lambda_n$ being the nonzero eigenvalues of the Laplacian matrix L.

B. Smooth Function

Definition 1: The function $f(x): \mathbb{R}^p \mapsto \mathbb{R}$ is smooth with constant $L_f > 0$ if it is differentiable and

$$\|\nabla f(x) - \nabla f(y)\| \le L_f\|x - y\|\ \forall x, y \in \mathbb{R}^p. \tag{6}$$

From [38, Lemma 1.2.3], we know that (6) implies

$$|f(y) - f(x) - (y - x)^\top\nabla f(x)| \le \frac{L_f}{2}\|y - x\|^2\ \forall x, y \in \mathbb{R}^p. \tag{7}$$

Moreover, we have the following lemma.

Lemma 2: If $f(x): \mathbb{R}^p \mapsto \mathbb{R}$ is smooth with constant $L_f > 0$, then, for any $a > L_f$, the function $g(x) = f(x) + \frac{a}{2}\|x\|^2$ is strongly convex with convex parameter $a - L_f$.

Proof: From (6), we have

$$\langle\nabla f(x) - \nabla f(y), x - y\rangle \ge -\|\nabla f(x) - \nabla f(y)\|\|x - y\| \ge -L_f\|x - y\|^2.$$

Then

$$\begin{aligned}\langle\nabla g(x) - \nabla g(y), x - y\rangle &= \langle\nabla f(x) + ax - \nabla f(y) - ay, x - y\rangle\\ &= \langle\nabla f(x) - \nabla f(y), x - y\rangle + a\|x - y\|^2\\ &\ge (a - L_f)\|x - y\|^2.\end{aligned}$$

Then, from [38, Theorem 2.1.9], we know that this lemma holds. □

C. Polyak–Łojasiewicz Condition

Let $f(x): \mathbb{R}^p \mapsto \mathbb{R}$ be a differentiable function. Let $\mathbb{X}^* = \arg\min_{x\in\mathbb{R}^p} f(x)$ and $f^* = \min_{x\in\mathbb{R}^p} f(x)$. Moreover, we assume that $f^* > -\infty$.

Definition 2: The function f satisfies the P–Ł condition with constant $\nu > 0$ if

$$\frac{1}{2}\|\nabla f(x)\|^2 \ge \nu(f(x) - f^*)\ \forall x \in \mathbb{R}^p. \tag{8}$$

It is straightforward to see that if a function is strongly convex with convex parameter ν, then it also satisfies the P–Ł condition with the same constant ν. Moreover, it was shown in [39] that the P–Ł condition is weaker than the commonly used conditions that have been explored to show linear convergence rates without strong convexity, such as essential strong convexity, weak strong convexity, and restricted strong convexity. The P–Ł condition implies that every stationary point is a global minimizer, i.e., $\mathbb{X}^* = \{x \in \mathbb{R}^p : \nabla f(x) = \mathbf{0}_p\}$. But unlike the (essentially, weakly, or restricted) strong convexity, the P–Ł condition does not imply the convexity of f. Moreover, it does not imply that $\mathbb{X}^*$ is a singleton either.

It was also given in [39] that the function $f(x) = x^2 + 3\sin^2(x)$ is an example of nonconvex functions satisfying the P–Ł condition with $\nu = 1/32$. Although it is difficult to precisely characterize the general class of functions satisfying the P–Ł condition, in [39], one special case was given as follows.

Lemma 3: Let $f(x) = g(Ax)$, where $g: \mathbb{R}^p \to \mathbb{R}$ is a strongly convex function and $A \in \mathbb{R}^{p\times p}$ is a matrix; then f satisfies the P–Ł condition.

Moreover, the loss functions in some applications may satisfy the P–Ł condition in the local region near a local minimum; see [40]. For example, [41] and [42] showed strong convexity in the neighborhood of the ground truth solution in some simple neural networks. Moreover, the P–Ł condition holds in certain reinforcement learning problems; see [43] and [44]. For example, [45] proved that the cost function of the policy optimization for the linear quadratic regulator problem is nonconvex and satisfies the P–Ł condition.

III. Distributed Alternating Direction Method of Multipliers

In this section, we propose a distributed ADMM algorithm to solve optimization (1) and analyze its convergence rate under different conditions.

We assume that the communication network among agents is described by a weighted undirected graph $\mathcal{G}$. Let $\mathbb{X}^*$ and f^* denote the optimal set and the minimum function value of the optimization problem (1), respectively. The following standard assumptions are made.

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

Assumption 2: The optimal set $\mathbb{X}^*$ is nonempty and $f^* > -\infty$.

Assumption 3: Each local cost function is smooth with constant $L_f > 0$.

Remark 1: It should be highlighted that the boundedness of the gradients of the cost functions are not assumed. Moreover, we do not assume that $\mathbb{X}^*$ is a singleton or finite set either.

A. Distributed ADMM Algorithm

Denote $\boldsymbol{x} = \mathrm{col}(x_1, \dots, x_n)$ and $\tilde{f}(\boldsymbol{x}) = \sum_{i=1}^n f_i(x_i)$, and then the optimization problem (1) is equivalent to the following constrained optimization problem:

$$\begin{aligned}&\boldsymbol{x} \in \mathbb{R}^{np}, x_0 \in \mathbb{R}^p\, \tilde{f}(\boldsymbol{x})\\ &\text{s.t} \qquad \boldsymbol{x} - \mathbf{1}_n \otimes x_0 = \mathbf{0}_{np}.\end{aligned} \tag{9}$$

The augmented Lagrangian of (9) is

$$\mathcal{L}(\boldsymbol{x}, x_0, \boldsymbol{v}) = \tilde{f}(\boldsymbol{x}) + \beta\langle \boldsymbol{v}, \boldsymbol{x} - \mathbf{1}_n \otimes x_0\rangle + \frac{\gamma}{2}\|\boldsymbol{x} - \mathbf{1}_n \otimes x_0\|^2 \tag{10}$$

where $\boldsymbol{v} = \mathrm{col}(v_1, \ldots, v_n) \in \mathbb{R}^{np}$ is the Lagrange multiplier, and $\beta > 0$ and $\gamma > 0$ are constants. Then, applying the classic ADMM algorithm [17], [18], we get the following ADMM algorithm to solve (9):

$$\boldsymbol{x}_{k+1} = \arg\min_{\boldsymbol{x}\in\mathbb{R}^{np}} \mathcal{L}(\boldsymbol{x}, x_{0,k}, \boldsymbol{v}_k) \tag{11a}$$

$$x_{0,k+1} = \arg\min_{x_0\in\mathbb{R}^{p}} \mathcal{L}(\boldsymbol{x}_{k+1}, x_0, \boldsymbol{v}_k) \tag{11b}$$

$$\boldsymbol{v}_{k+1} = \boldsymbol{v}_k + \frac{\gamma}{\beta}(\boldsymbol{x}_{k+1} - \mathbf{1}_n \otimes x_{0,k+1}). \tag{11c}$$

If there exists a virtual agent, denoted as agent 0, that can communicate with all of the n agents, which corresponds to that the underlying communication graph $\mathcal{G}$ of the n agents is a star graph, then the ADMM algorithm (11) can be written agentwise as

$$x_{i,k+1} = \arg\min_{x\in\mathbb{R}^p} f_i(x) + \beta\langle v_{i,k}, x\rangle + \frac{\gamma}{2}\|x - x_{0,k}\|^2 \tag{12a}$$

$$x_{0,k+1} = \frac{1}{n}\sum_{i=1}^{n}\left(x_{i,k+1} + \frac{\beta}{\gamma}v_{i,k}\right) \tag{12b}$$

$$v_{i,k+1} = v_{i,k} + \frac{\gamma}{\beta}(x_{i,k+1} - x_{0,k+1})\ \forall i \in [n]. \tag{12c}$$

It has been shown in [33]–[35] that for star graphs, the ADMM algorithm (12) can find a first-order stationary point of the optimization problem (1) with a rate $\mathcal{O}(1/k)$ if γ is large enough, $\beta = 1$, and Assumptions 2 and 3 hold. If the communication graph $\mathcal{G}$ is a general connected graph, then each agent i cannot execute (12a) and (12c) since $x_{0,k+1}$ is not available in this case. In other words, the ADMM algorithm (12) is restricted to a star graph. In order to remove this restriction, we modify the ADMM algorithm (12) as follows:

$$x_{i,k+1} = \arg\min\nolimits_{x\in\mathbb{R}^p} f_i(x) + \beta\langle v_{i,k}, x\rangle + \frac{\gamma}{2}\|x - x_{i,k} + \frac{\alpha}{\gamma}\sum_{j=1}^{n}L_{ij}x_{j,k}\|^2 \tag{13a}$$

$$v_{i,k+1} = v_{i,k} + \frac{\beta}{\gamma}\sum_{j=1}^{n}L_{ij}x_{j,k+1},\ \sum_{j=1}^{n}v_{j,0} = \mathbf{0}_p\ \forall i \in [n] \tag{13b}$$

where $\alpha > 0$ is a constant.

Remark 2: The intuition of the modification from (12) to (13) is as follows. When γ is large enough, then from (12b), we know $x_{0,k+1} \approx (1/n)\sum_{i=1}^{n}x_{i,k+1}$. In multiagent systems, for each agent i, $\frac{1}{n}\sum_{i=1}^{n}x_{i,k}$ can be estimated by $x_{i,k} - b\sum_{j=1}^{n}L_{ij}x_{j,k}$ with some positive gains b. Thus, replacing $x_{0,k}$ in (12a) by its estimation $x_{i,k} - (\alpha/\gamma)\sum_{j=1}^{n}L_{ij}x_{j,k}$ gives (13a). Then, each $x_{i,k+1}$ is available to each agent i, and, through communication, it is also available to agent j if $j \in \mathcal{N}_i$. Thus, replacing $x_{0,k+1}$ in (12c) by its estimation $x_{i,k+1} - \frac{\beta^2}{\gamma^2}\sum_{j=1}^{n}L_{ij}x_{j,k+1}$ gives (13b). Here, we used different gains $\frac{\alpha}{\gamma}$ and $\frac{\beta^2}{\gamma^2}$ since such a setting facilitates the convergence analysis. Moreover, the extra initialization condition $\sum_{j=1}^{n}v_{j,0} = \mathbf{0}_p$ is also used to facilitate the convergence analysis. This initialization condition is easy to be satisfied, for example, $v_{i,0} = \mathbf{0}_p\ \forall i \in [n]$, or $v_{i,0} = \sum_{j=1}^{n}L_{ij}x_{j,0}\ \forall i \in [n]$.

Algorithm 1: Distributed ADMM Algorithm.

Input: constants $\alpha > 0$, $\beta > 0$, and $\gamma > 0$.
Initialize: $x_{i,0} \in \mathbb{R}^p$ and $v_{i,0} = \mathbf{0}_p,\ \forall i \in [n]$.
Broadcast $x_{i,0}$ to $\mathcal{N}_i$ and receive $x_{j,0}$ from $j \in \mathcal{N}_i$;
for $k = 0, 1, \ldots$ **do**
  for $i = 1, \ldots, n$ in parallel**do**
    Update $x_{i,k+1}$ by (13a);
    Broadcast $x_{i,k+1}$ to $\mathcal{N}_i$ and receive $x_{j,k+1}$ from $j \in \mathcal{N}_i$;
    Update $v_{i,k+1}$ by (13b).
  end for
end for
Output: $\{\boldsymbol{x}_k\}$.

Remark 3: The objective function in subproblem (13a) may be not convex since each f_i is possibly nonconvex. However, if Assumption 3 holds and $\gamma > L_f$, then from Lemma 2, we know that the objective function is strongly convex with convexity parameter $\gamma - L_f$. Hence, subproblem (13a) is solvable.

We write the distributed ADMM algorithm (13) in pseudocode as Algorithm 1.

For simplicity, denote $\boldsymbol{x}_k = \mathrm{col}(x_{1,k}, \ldots, x_{n,k})$, $\boldsymbol{v}_k = \mathrm{col}(v_{1,k}, \ldots, v_{n,k})$, $\boldsymbol{L} = L \otimes \mathbf{I}_p$, $\boldsymbol{K} = K_n \otimes \mathbf{I}_p$, $\boldsymbol{H} = \frac{1}{n}(\mathbf{1}_n\mathbf{1}_n^\top \otimes \mathbf{I}_p)$, $\boldsymbol{Q} = R\Lambda_1^{-1}R^\top \otimes \mathbf{I}_p$, $\bar{x}_k = \frac{1}{n}(\mathbf{1}_n^\top \otimes \mathbf{I}_p)\boldsymbol{x}_k$, $\bar{\boldsymbol{x}}_k = \mathbf{1}_n \otimes \bar{x}_k$, $\boldsymbol{g}_k = \nabla\tilde{f}(\boldsymbol{x}_k)$, $\bar{\boldsymbol{g}}_k = \boldsymbol{H}\boldsymbol{g}_k$, $\boldsymbol{g}_k^0 = \nabla\tilde{f}(\bar{\boldsymbol{x}}_k)$, $\bar{\boldsymbol{g}}_k^0 = \boldsymbol{H}\boldsymbol{g}_k^0 = \mathbf{1}_n \otimes \nabla f(\bar{x}_k)$, and $\boldsymbol{y}_k = \boldsymbol{v}_k + \frac{1}{\beta}\boldsymbol{g}_k^0$.

B. Convergence Analysis

In this section, we present convergence analysis for Algorithm 1. We first present a preliminary result regarding the general relations of two consecutive outputs of Algorithm 1.

Lemma 4: Let $\{\boldsymbol{x}_k\}$ be the sequence generated by Algorithm 1. If Assumptions 1–3 hold and $\gamma > L_f$, then

$$\tilde{V}_{k+1} \le \tilde{V}_k - \|\boldsymbol{x}_k\|^2_{\frac{1}{\gamma}(\epsilon_3 - \frac{1}{\gamma}\epsilon_4)\boldsymbol{K}} - \|\boldsymbol{y}_k\|^2_{\frac{1}{\gamma}(\epsilon_5 - \frac{1}{\gamma}\epsilon_6)\boldsymbol{K}} - \frac{1}{4\gamma}\|\bar{\boldsymbol{g}}_k^0\|^2 - \frac{1}{\gamma}\left(\epsilon_7 - \frac{1}{\gamma}\epsilon_8 - \frac{1}{\gamma^2}\epsilon_9 - \frac{1}{\gamma^3}\epsilon_{10}\right)\|\bar{\boldsymbol{g}}_{k+1}\|^2 \tag{14}$$

where $\tilde{V}_k = V_k - \|\boldsymbol{x}_k\|^2_{\frac{1}{\gamma}(\epsilon_1 + \frac{1}{\gamma}\epsilon_2)\boldsymbol{K}}$, $V_k = \sum_{i=1}^{4}V_{i,k}$, and

$$V_{1,k} = \frac{1}{2}\|\boldsymbol{x}_k\|^2_{\boldsymbol{K}},\ V_{2,k} = \frac{1}{2}\|\boldsymbol{y}_k\|^2_{\boldsymbol{Q} + \frac{\alpha}{\beta}\boldsymbol{K}}$$

$$V_{3,k} = \boldsymbol{x}_k^\top\boldsymbol{K}\left(\boldsymbol{v}_k + \frac{1}{\beta}\boldsymbol{g}_k^0\right),\ V_{4,k} = n(f(\bar{x}_k) - f^*)$$

$$\epsilon_1 = \frac{3}{2} + 2L_f^2 + \beta\rho(L),$$
$$\epsilon_2 = (2+\rho(L^2))3L_f^2 + \beta^2\rho(L) + \alpha\beta\rho(L^2),$$
$$\epsilon_3 = \alpha\rho_2(L) - \frac{1}{2} - \epsilon_1,$$
$$\epsilon_4 = \left(1 + \frac{1}{2}\rho(L^2)\right) 3\alpha^2\rho(L^2) + \epsilon_2,$$
$$\epsilon_5 = \beta - \frac{1}{2} - \frac{\alpha}{2\beta^2} - \frac{1}{2\beta\rho_2(L)},$$
$$\epsilon_6 = \frac{1}{2}(\alpha^2 + (7+3\rho(L^2))\beta^2),$$
$$\epsilon_7 = \frac{1}{4} - \frac{1}{2\beta}\left(\frac{1}{\rho_2(L)} + \frac{\alpha+1}{\beta}\right) L_f^2,$$
$$\epsilon_8 = \left(\frac{1}{2} + \frac{1}{\beta^2}\left(\frac{1}{\rho_2(L)} + \frac{\alpha}{\beta}\right) L_f\right) L_f,$$
$$\epsilon_9 = 3L_f^2,\ \epsilon_{10} = 3(2+\rho(L^2))L_f^2.$$

Proof: The proof is given in Appendix A. ∎

Remark 4: From Lemma 4, we know that $\tilde{V}_k$ can serve as the potential function for Algorithm 1. This potential function has a good property that it is nonnegative if the parameters α, β, and γ are appropriately chosen. With this nonnegative potential function, we can establish convergence rates for Algorithm 1 under different assumptions as shown in the following.

The first main result is stated below.

Theorem 1: Let $\{\boldsymbol{x}_k\}$ be the sequence generated by Algorithm 1. If Assumptions 1–3 hold, $\frac{1}{\rho_2(L)}(\rho(L)\beta + \kappa_1) < \alpha \le \kappa_2\beta$, $\beta > \max\{\frac{\kappa_1}{\kappa_2\rho_2(L)-\rho(L)},\ \kappa_3,\ \kappa_4\}$, and $\gamma > \max\{\frac{\epsilon_4}{\epsilon_3},\ \frac{\epsilon_6}{\epsilon_5},\ \frac{\epsilon_8+\epsilon_9+\epsilon_{10}}{\epsilon_7},\ \frac{1}{\epsilon_{15}}\}$, then

$$\sum_{k=0}^{T}(\|\boldsymbol{x}_k\|^2_{\boldsymbol{K}} + \|\boldsymbol{y}_k\|^2_{\boldsymbol{K}} + \|\bar{\boldsymbol{g}}_k^0\|^2) \le \frac{\tilde{V}_0}{\epsilon_{16}}\ \forall T \ge 0 \tag{15}$$

where

$$\kappa_1 = 2L_f^2 + 2,\ \kappa_2 > \frac{\rho(L)}{\rho_2(L)},$$
$$\kappa_3 = \frac{1}{4}(1 + (1 + 8\kappa_2 + \frac{8}{\rho_2(L)})^{\frac{1}{2}}),$$
$$\kappa_4 = \left(\kappa_2 + \frac{1}{\rho_2(L)}\right) L_f^2 + \left(\left(\kappa_2 + \frac{1}{\rho_2(L)}\right)^2 L_f^4 + 2L_f^2\right)^{\frac{1}{2}}$$
$$\epsilon_{11} = \frac{1}{2} - \frac{1}{\gamma}\epsilon_1 - \frac{1}{\gamma^2}\epsilon_2 > 0,\ \epsilon_{12} = \frac{1}{2}\left(\frac{1}{\rho(L)} + \frac{\alpha}{\beta}\right)$$
$$\epsilon_{13} = \frac{1}{2}(\epsilon_{11} - \epsilon_{12} + ((\epsilon_{11}-\epsilon_{12})^2 + 1)^{\frac{1}{2}})$$
$$\epsilon_{14} = \frac{\alpha+\beta}{2\beta} + \frac{1}{2\rho_2(L)}$$
$$\epsilon_{15} = \frac{1}{2\epsilon_2}\left(-\epsilon_1 + \left(\epsilon_1^2 + 2 - \frac{1}{\epsilon_{12}}\right)^{\frac{1}{2}}\right) > 0$$
$$\epsilon_{16} = \frac{1}{\gamma}\min\left\{\epsilon_3 - \frac{1}{\gamma}\epsilon_4,\ \epsilon_5 - \frac{1}{\gamma}\epsilon_6,\ \frac{1}{4}\right\} > 0.$$

Proof: **(i)** We first show that all of the used constants are positive.

From $\frac{1}{\rho_2(L)}(\rho(L)\beta + \kappa_1) < \alpha$, we have $\frac{\alpha}{\beta} > \frac{\rho(L)}{\rho_2(L)} \ge 1$. Then, we know $\epsilon_{12} > \frac{1}{2}$. Thus, $2 - \frac{1}{\epsilon_{12}} > 0$. Hence

$$\epsilon_{15} > 0. \tag{16}$$

Then, from $0 < \frac{1}{\gamma} < \epsilon_{15}$, we have $4\epsilon_{11}\epsilon_{12} > 1$. Hence

$$\frac{1}{2} > \epsilon_{11} - \epsilon_{13} > 0. \tag{17}$$

From $\frac{1}{\rho_2(L)}(\rho(L)\beta + \kappa_1) < \alpha$, we have

$$\epsilon_3 > \kappa_1 - 2L_f^2 - 2 = 0. \tag{18}$$

Hence, from $0 < \frac{1}{\gamma} < \frac{\epsilon_3}{\epsilon_4}$ and (18), we have

$$\frac{1}{\gamma}\left(\epsilon_3 - \frac{1}{\gamma}\epsilon_4\right) > 0. \tag{19}$$

From $\alpha \le \kappa_2\beta$ and $\beta > \kappa_3$, we have

$$\epsilon_5 \ge \left(\beta - \frac{1}{2} - \frac{\kappa_2}{2\beta}\right) - \frac{1}{2\beta\rho_2(L)} > 0. \tag{20}$$

Hence, from $0 < \frac{1}{\gamma} < \frac{\epsilon_5}{\epsilon_6}$ and (20), we have

$$\frac{1}{\gamma}\left(\epsilon_5 - \frac{1}{\gamma}\epsilon_6\right) > 0. \tag{21}$$

From (19) and (21), we have

$$\epsilon_{16} > 0. \tag{22}$$

From $\alpha \le \kappa_2\beta$ and $\beta > \kappa_4$, we have

$$\epsilon_7 \ge \frac{1}{4} - \frac{1}{2\beta}\left(\frac{1}{\beta} + \frac{1}{\rho_2(L)} + \kappa_2\right) L_f^2 > 0. \tag{23}$$

From $\kappa_2 > 1$, we have $\kappa_3 > 1$. Thus, $\beta > 1$. Thus, $\frac{1}{\gamma} < \frac{\epsilon_5}{\epsilon_6} < \frac{2}{7\beta} < \frac{2}{7}$. Hence, from $0 < \frac{1}{\gamma} < \frac{\epsilon_7}{\epsilon_8+\epsilon_9+\epsilon_{10}}$ and (23), we have

$$\begin{aligned}&\frac{1}{\gamma}\left(\epsilon_7 - \epsilon_8\frac{1}{\gamma} - \epsilon_9\frac{1}{\gamma^2} - \epsilon_{10}\frac{1}{\gamma^3}\right)\\ &> \frac{1}{\gamma}\left(\epsilon_7 - \epsilon_8\frac{1}{\gamma} - \epsilon_9\frac{1}{\gamma} - \epsilon_{10}\frac{1}{\gamma}\right) > 0.\end{aligned} \tag{24}$$

(ii) We then show that (15) holds.

Noting that $\beta > \kappa_4 > \sqrt{2}L_f$ and $0 < \epsilon_5 < \beta$, we know $\gamma > \frac{\epsilon_6}{\epsilon_5} > \frac{\epsilon_6}{\beta} > \frac{7\beta}{2} > \frac{7\sqrt{2}L_f}{2} > L_f$. Thus, the conditions needed in Lemma 4 are all satisfied. Thus, (14) holds.

Denote

$$\hat{V}_k = \|\boldsymbol{x}_k\|^2_{\boldsymbol{K}} + \|\boldsymbol{y}_k\|^2_{\boldsymbol{K}} + n(f(\bar{x}_k) - f^*). \tag{25}$$

We know

$$\begin{aligned}\tilde{V}_k = &\left(\frac{1}{2} - \epsilon_1\frac{1}{\gamma} - \epsilon_2\frac{1}{\gamma^2}\right)\|\boldsymbol{x}_k\|^2_{\boldsymbol{K}} + \frac{1}{2}\|\boldsymbol{y}_k\|^2_{\boldsymbol{Q}+\frac{\alpha}{\beta}\boldsymbol{K}}\\ &+ \boldsymbol{x}_k^\top\boldsymbol{K}\boldsymbol{y}_k + n(f(\bar{x}_k) - f^*)\end{aligned}$$

$$\geq \epsilon_{11}\|\boldsymbol{x}_k\|^2_{\boldsymbol{K}} + \epsilon_{12}\|\boldsymbol{y}_k\|^2_{\boldsymbol{K}} - \epsilon_{13}\|\boldsymbol{x}_k\|^2_{\boldsymbol{K}} - \frac{1}{4\epsilon_{13}}\|\boldsymbol{y}_k\|^2_{\boldsymbol{K}} + n(f(\bar{x}_k) - f^*)$$

$$= (\epsilon_{11} - \epsilon_{13})(\|\boldsymbol{x}_k\|^2_{\boldsymbol{K}} + \|\boldsymbol{y}_k\|^2_{\boldsymbol{K}}) + n(f(\bar{x}_k) - f^*) \quad (26)$$

$$\geq (\epsilon_{11} - \epsilon_{13})\hat{V}_k \geq 0 \quad (27)$$

where the first inequality holds due to (5) and the Cauchy–Schwarz inequality; the second equality holds due to $\epsilon_{11} - \epsilon_{13} = \epsilon_{12} - \frac{1}{4\epsilon_{13}}$; and the last inequality holds due to (17). Similarly, we have

$$\tilde{V}_k \leq V_k \leq \epsilon_{14}\hat{V}_k. \quad (28)$$

From (14), (24), and $\boldsymbol{K} \geq 0$, we know that

$$\tilde{V}_{k+1} \leq \tilde{V}_k - \|\boldsymbol{x}_k\|^2_{\frac{1}{\gamma}(\epsilon_1 - \epsilon_3 - \frac{1}{\gamma}(\epsilon_2 + \epsilon_4))\boldsymbol{K}} - \|\boldsymbol{y}_k\|^2_{\frac{1}{\gamma}(\epsilon_5 - \frac{1}{\gamma}\epsilon_6)\boldsymbol{K}} - \frac{1}{4\gamma}\|\bar{\boldsymbol{g}}^0_k\|^2$$

$$\leq \tilde{V}_k - \epsilon_{16}(\|\boldsymbol{x}_k\|^2_{\boldsymbol{K}} + \|\boldsymbol{y}_k\|^2_{\boldsymbol{K}} + \|\bar{\boldsymbol{g}}^0_k\|^2). \quad (29)$$

Then, (29) yields

$$\sum_{k=0}^{T}\tilde{V}_{k+1} \leq \sum_{k=0}^{T}\tilde{V}_k - \epsilon_{16}\sum_{k=0}^{T}(\|\boldsymbol{x}_k\|^2_{\boldsymbol{K}} + \|\boldsymbol{y}_k\|^2_{\boldsymbol{K}} + \|\bar{\boldsymbol{g}}^0_k\|^2). \quad (30)$$

Then, (30) yields

$$\tilde{V}_{T+1} + \epsilon_{16}\sum_{k=0}^{T}(\|\boldsymbol{x}_k\|^2_{\boldsymbol{K}} + \|\boldsymbol{y}_k\|^2_{\boldsymbol{K}} + \|\bar{\boldsymbol{g}}^0_k\|^2) \leq \tilde{V}_0. \quad (31)$$

From (31), (22), and (27), we know that (15) holds. ■

Remark 5: From (15), we know that $\min_{k\in[T]}\{\|\boldsymbol{x}_k\|^2_{\boldsymbol{K}} + \|\boldsymbol{v}_k + \frac{1}{\beta}\boldsymbol{g}^0_k\|^2_{\boldsymbol{K}} + \|\bar{\boldsymbol{g}}^0_k\|^2\} = \mathcal{O}(1/T)$. In other words, Theorem 1 shows that our distributed ADMM algorithm converges to a stationary point sublinearly at a rate $\mathcal{O}(1/T)$. This rate is the same as that achieved by the Prox-PDA proposed in [36] under the same conditions. The same convergence rate was also achieved by ADMM algorithms proposed in [10], [33]–[35]. However, these algorithms are restricted to a star graph. Moreover, the algorithms proposed in [10], [33], and [34] require that each leaf agent has to communicate both primal and dual variables to the hub agent and the algorithm proposed in [35] is based on the standard master/worker model. Compared with these algorithms, the advantages of Algorithm 1 are that it is suitable for general connected graphs and each agent only needs to communicate the primal variable with its neighbors, while one potential drawback is that our algorithm is synchronous. We leave the extension to the asynchronous communication setting for future studies.

Remark 6: The settings on the algorithm parameters α, β, and γ in Theorem 1 are instrumental in the convergence analysis of Algorithm 1. They are just sufficient conditions. In other words, the bounds for α, β, and γ are not tight. With some modifications of the proofs, for example choosing different coefficients when applying the Cauchy–Schwarz inequality in the proofs, other forms of bounds for these parameters can still guarantee the same kind of convergence rate as stated in (15) but with a different definition of ϵ_{16}.

If the following assumption holds, then Algorithm 1 can find a global optimum and the convergence rate is linear.

Assumption 4: The global cost function $f(x)$ satisfies the P–Ł condition with constant $\nu > 0$.

Theorem 2: Let $\{\boldsymbol{x}_k\}$ be the sequence generated by Algorithm 1. If Assumptions 1–4 hold, the settings on α, β, and γ are the same as those in Theorem 1, then

$$\|\boldsymbol{x}_k - \bar{\boldsymbol{x}}_k\|^2 + n(f(\bar{x}_k) - f^*) \leq (1-\epsilon)^k c \;\forall k \geq 0 \quad (32)$$

where

$$\epsilon = \frac{\epsilon_{17}}{\epsilon_{14}} \in (0,1),\; c = \frac{\tilde{V}_0}{\epsilon_{11} - \epsilon_{13}} \geq 0,$$

$$\epsilon_{17} = \frac{1}{\gamma}\min\{\epsilon_3 - \frac{1}{\gamma}\epsilon_4,\; \epsilon_5 - \frac{1}{\gamma}\epsilon_6,\; \frac{\nu}{2}\} > 0.$$

Proof: **(i)** We first show that $\epsilon \in (0,1)$ and $c \geq 0$.

From (19) and (21), we have

$$\epsilon_{17} > 0. \quad (33)$$

From Assumptions 2 and 4 as well as (8), we have that

$$\|\bar{\boldsymbol{g}}^0_k\|^2 = n\|\nabla f(\bar{x}_k)\|^2 \geq 2\nu n(f(\bar{x}_k) - f^*). \quad (34)$$

Then, from (24)–(25), (34), (33), and (28), we have

$$\tilde{V}_{k+1} \leq \tilde{V}_k - \epsilon_{17}\hat{V}_k \leq \tilde{V}_k - \frac{\epsilon_{17}}{\epsilon_{14}}\tilde{V}_k. \quad (35)$$

Noting that $\epsilon_5 < \beta$, $\epsilon_6 > \frac{7}{2}\beta^2$, and $\epsilon_{14} > \frac{\alpha+\beta}{2\beta} > 1$, we have

$$0 < \epsilon = \frac{\epsilon_{17}}{\epsilon_{14}} < \epsilon_{17} \leq \frac{1}{\gamma}(\epsilon_5 - \frac{1}{\gamma}\epsilon_6) \leq \frac{\epsilon_5^2}{4\epsilon_6} < \frac{1}{14}. \quad (36)$$

From (17), we have $c \geq 0$.

(ii) We then show that (32) holds.

From (35), (27), and (36), we have

$$\tilde{V}_{k+1} \leq (1-\epsilon)\tilde{V}_k \leq (1-\epsilon)^{k+1}\tilde{V}_0. \quad (37)$$

Hence, from (27) and (17), we have

$$\|\boldsymbol{x}_k - \bar{\boldsymbol{x}}_k\|^2 + n(f(\bar{x}_k) - f^*) = \|\boldsymbol{x}_k\|^2_{\boldsymbol{K}} + n(f(\bar{x}_k) - f^*) \leq \hat{V}_k \leq \frac{\tilde{V}_k}{\epsilon_{11} - \epsilon_{13}}. \quad (38)$$

Hence, (37) and (38) give (32). ■

Remark 7: From (32), we know that there exists a constant $\theta \in (0,1)$ such that $\|\boldsymbol{x}_k - \bar{\boldsymbol{x}}_k\|^2 + n(f(\bar{x}_k) - f^*) = \mathcal{O}(\theta^k)$. In other words, Theorem 2 shows that our distributed ADMM algorithm converges linearly under the P–Ł condition. Linear convergence was also established by the distributed ADMM algorithms proposed in [22], [24], [28], and [29]. However, they all assumed that each local cost function is convex. Moreover, in [22] and [28], it was assumed that each local cost function is strongly convex. In [24], it was assumed that the optimal set $\mathbb{X}^*$ is a singleton and the global cost function is locally strongly convex. In [29], it was assumed that the global cost function is strongly convex. In contrast, the linear convergence result established in Theorem 2 only requires that the global cost function satisfies the P–Ł condition, but the convexity assumption on cost functions and the singleton assumption on the optimal set are not required. Compared with the results established in [22], [24], [28], and [29], one potential drawback of our results is that we need to use some global information, such as the smooth constant and the eigenvalues of the Laplacian matrix associated with the communication graph to design the algorithm parameters α, β, and γ. Noting that [10], [33]–[36] which proposed distributed ADMM algorithms for nonconvex optimization problem also have such a kind of drawback, we think it may be caused by the lack of the convexity assumption. It is unclear how to overcome this drawback. It may be overcome with the studies on estimating the largest and the second smallest eigenvalues of the communication graph [46], [47].

Remark 8: A detailed expression for the theoretical convergence rate is stated in (32), although it is complicated. Note that ϵ_{17} is the only constant that depends on the P–Ł constant ν. From (32), we know that the larger the P–Ł constant, the faster the convergence. However, we cannot make similar kinds of conclusion for the smooth constant and the eigenvalues of the communication graph. Compared with the linear convergence rates achieved in [22] and [28], this is a potential drawback. We think that it may be caused because the weaker assumption (the P–Ł condition) rather than the stronger assumption (the strongly convex assumption for each local cost function) is used.

IV. Distributed Linearized Alternating Direction Method of Multipliers

One potential limitation of Algorithm 1 is the requirement that at each iteration, each subproblem (13a) needs to be solved exactly, which normally has no closed-form solution, and thus results in a heavy computational burden to each agent. To over come this, in this section, we propose a distributed L-ADMM algorithm and analyze its convergence rate under different conditions.

A. Distributed Linearized ADMM Algorithm

In this section, we present the modification of (13a). The main idea is that instead of minimizing exactly with respect to x, we take an inexact minimization in which the function $f_i(x)$ is replaced by a linearized approximation centered at the current iteration. Specifically, replacing the function $f_i(x)$ with $f_i(x_{i,k}) + \langle \nabla f_i(x_{i,k}), x - x_{i,k}\rangle$ in (13a) gives the inexact update for $x_{i,k+1}$ as follows:

$$x_{i,k+1} = \arg\min_{x\in\mathbb{R}^p} f_i(x_{i,k}) + \langle \nabla f_i(x_{i,k}), x - x_{i,k}\rangle + \beta\langle v_{i,k}, x\rangle + \frac{\gamma}{2}\|x - x_{i,k} + \frac{\alpha}{\gamma}\sum_{j=1}^{n} L_{ij}x_{j,k}\|^2. \tag{39}$$

The idea of using linearized approximation is standard and has also been used in [23], [36], and [48]–[50].

Noting that the objective function in subproblem (39) is strongly convex, from the first-order optimality conditions for convex optimization problem, we know that the explicit expression of $x_{i,k+1}$. Hence, we get the following distributed L-ADMM algorithm:

Algorithm 2: Distributed L-ADMM Algorithm.

1: **Input**: constants $\alpha > 0$, $\beta > 0$, and $\gamma > 0$.
2: **Initialize**: $x_{i,0} \in \mathbb{R}^p$ and $v_{i,0} = \mathbf{0}_p,\ \forall i \in [n]$.
3: Broadcast $x_{i,0}$ to $\mathcal{N}_i$ and receive $x_{j,0}$ from $j \in \mathcal{N}_i$;
4: **for** $k = 0, 1, \ldots$ **do**
5: **for** $i = 1, \ldots, n$ in parallel**do**
6: Update $x_{i,k+1}$ by (40a);
7: Broadcast $x_{i,k+1}$ to $\mathcal{N}_i$ and receive $x_{j,k+1}$ from $j \in \mathcal{N}_i$;
8: Update $v_{i,k+1}$ by (40b).
9: **end for**
10: **end for**
11: **Output**: $\{\boldsymbol{x}_k\}$.

$$x_{i,k+1} = x_{i,k} - \frac{1}{\gamma}\left(\alpha\sum_{j=1}^{n} L_{ij}x_{j,k} + \beta v_{i,k} + \nabla f_i(x_{i,k})\right), \tag{40a}$$

$$v_{i,k+1} = v_{i,k} + \frac{\beta}{\gamma}\sum_{j=1}^{n} L_{ij}x_{j,k+1},\ \sum_{j=1}^{n} v_{j,0} = \mathbf{0}_p\ \forall i \in [n]. \tag{40b}$$

We write the distributed L-ADMM algorithm (40) in pseudocode as Algorithm 2.

Remark 9: It is straightforward to check that the sequence $\{\boldsymbol{x}_k\}$ generated by the distributed L-ADMM algorithm (40) with the initialization condition $v_{i,0} = \frac{\beta}{\gamma}\sum_{j=1}^{n} L_{ij}x_{i,0}\ \forall i \in [n]$ is the same as the sequence generated by the EXTRA proposed in [51]

$$\boldsymbol{x}_1 = \boldsymbol{W}\boldsymbol{x}_0 - \frac{1}{\gamma}\nabla\tilde{f}(\boldsymbol{x}_0)\ \forall \boldsymbol{x}_0 \in \mathbb{R}^{np}$$

$$\boldsymbol{x}_{k+1} = (\mathbf{I}_{np} + \boldsymbol{W})\boldsymbol{x}_k - \tilde{\boldsymbol{W}}\boldsymbol{x}_{k-1} - \frac{1}{\gamma}(\nabla\tilde{f}(\boldsymbol{x}_k) - \nabla\tilde{f}(\boldsymbol{x}_{k-1}))$$

with mixing matrices $\boldsymbol{W} = \mathbf{I}_{np} - \frac{\alpha}{\gamma}\boldsymbol{L} - \frac{\beta^2}{\gamma^2}\boldsymbol{L}$ and $\tilde{\boldsymbol{W}} = \mathbf{I}_{np} - \frac{\alpha}{\gamma}\boldsymbol{L}$. However, in [51], it was assumed that each local cost function is convex, the global cost function is restricted strongly convex, and $\mathbb{X}^*$ is a singleton, while our proposed L-ADMM algorithm (40) is applicable to general nonconvex cost functions as shown later in Theorems 3 and 4.

B. Convergence Analysis

Similar to Lemma 4, we have the following lemma.

Lemma 5: Let $\{\boldsymbol{x}_k\}$ be the sequence generated by Algorithm 2. If Assumptions 1–3 hold, then

$$\check{V}_{k+1} \le \check{V}_k - \|\boldsymbol{x}_k\|^2_{\frac{1}{\gamma}(\check{\epsilon}_3 - \frac{1}{\gamma}\check{\epsilon}_4)\boldsymbol{K}} - \|\boldsymbol{y}_k\|^2_{\frac{1}{\gamma}(\epsilon_5 - \frac{1}{\gamma}\epsilon_6)\boldsymbol{K}}$$

$$-\frac{1}{4\gamma}\|\bar{\boldsymbol{g}}_k^0\|^2-\frac{1}{\gamma}\left(\epsilon_7-\frac{1}{\gamma}\epsilon_8\right)\|\bar{\boldsymbol{g}}_k\|^2 \tag{41}$$

where $\breve{V}_k=V_k-\|\boldsymbol{x}_k\|^2_{\frac{1}{\gamma}(\breve{\epsilon}_1+\frac{1}{\gamma}\breve{\epsilon}_2)\boldsymbol{K}}$, and

$$\breve{\epsilon}_1=\frac{1}{2}+\beta\rho(L),\ \breve{\epsilon}_2=\beta^2\rho(L)+\alpha\beta\rho(L^2),$$
$$\breve{\epsilon}_3=\frac{1}{2}(2\alpha\rho_2(L)-1-3L_f^2)-\breve{\epsilon}_1,$$
$$\breve{\epsilon}_4=3(1+\frac{1}{2}\rho(L^2))(\alpha^2\rho(L^2)+L_f^2)+\breve{\epsilon}_2.$$

Proof: The proof is similar to the proof of Lemma 4 and is thus omitted. ■

Similar to Theorem 1, we have the following result.

Theorem 3: Let $\{\boldsymbol{x}_k\}$ be the sequence generated by Algorithm 2. If Assumptions 1–3 hold, $\frac{1}{\rho_2(L)}(\rho(L)\beta+\breve{\kappa}_1)<\alpha\le\kappa_2\beta,\ \beta>\max\{\frac{\breve{\kappa}_1}{\kappa_2\rho_2(L)-\rho(L)},\ \kappa_3,\ \kappa_4\}$, and $\gamma>\max\{\frac{\breve{\epsilon}_4}{\breve{\epsilon}_3},\ \frac{\epsilon_6}{\epsilon_5},\ \frac{\epsilon_8}{\epsilon_7},\ \frac{1}{\breve{\epsilon}_{15}}\}$, then

$$\sum_{k=0}^{T}(\|\boldsymbol{x}_k\|^2_{\boldsymbol{K}}+\|\boldsymbol{y}_k\|^2_{\boldsymbol{K}}+\|\bar{\boldsymbol{g}}_k^0\|^2)\le\frac{\breve{V}_0}{\breve{\epsilon}_{16}}\ \forall T\ge 0 \tag{42}$$

where

$$\breve{\kappa}_1=\frac{3}{2}L_f^2+1,\ \breve{\epsilon}_{11}=\frac{1}{2}-\frac{1}{\gamma}\breve{\epsilon}_1-\frac{1}{\gamma^2}\breve{\epsilon}_2>0$$
$$\breve{\epsilon}_{13}=\frac{1}{2}(\breve{\epsilon}_{11}-\epsilon_{12}+((\breve{\epsilon}_{11}-\epsilon_{12})^2+1)^{\frac{1}{2}})$$
$$\breve{\epsilon}_{15}=\frac{1}{2\breve{\epsilon}_2}(-\breve{\epsilon}_1+(\breve{\epsilon}_1^2+2-\frac{1}{\epsilon_{12}})^{\frac{1}{2}})>0$$
$$\breve{\epsilon}_{16}=\frac{1}{\gamma}\min\{\breve{\epsilon}_3-\frac{1}{\gamma}\breve{\epsilon}_4,\ \epsilon_5-\frac{1}{\gamma}\epsilon_6,\ \frac{1}{4}\}>0.$$

Proof: The proof is similar to the proof of Theorem 1 and is thus omitted. ■

When Assumption 4 also holds, similar to Theorem 2 we have the following result.

Theorem 4: Let $\{\boldsymbol{x}_k\}$ be the sequence generated by Algorithm 2. If Assumptions 1–4 hold, the settings on α, β, and γ are the same as those in Theorem 3, and then

$$\|\boldsymbol{x}_k-\bar{\boldsymbol{x}}_k\|^2+n(f(\bar{x}_k)-f^*)\le(1-\breve{\epsilon})^k\breve{c}\ \forall k\ge 0 \tag{43}$$

where

$$\breve{\epsilon}=\frac{\breve{\epsilon}_{17}}{\epsilon_{14}}\in(0,1),\ \breve{c}=\frac{\breve{V}_0}{\breve{\epsilon}_{11}-\breve{\epsilon}_{13}}\ge 0$$
$$\breve{\epsilon}_{17}=\frac{1}{\gamma}\min\{\breve{\epsilon}_3-\frac{1}{\gamma}\breve{\epsilon}_4,\ \epsilon_5-\frac{1}{\gamma}\epsilon_6,\ \frac{\nu}{2}\}>0.$$

■

Proof: The proof is similar to the proof of Theorem 2 and is thus omitted. ■

Remark 10: The same convergence rate as stated in (42) has also been achieved by the linearized version of Prox-PDA, the distributed proximal gradient primal-dual algorithm (Prox-GPDA), proposed in [36] under the same conditions. However, we also show that our distributed L-ADMM algorithm achieves linear convergence under the P–Ł condition, which was not considered in [36].

Remark 11: Linear convergence was also established by the distributed L-ADMM algorithm proposed in [23]. However, in [23], it was assumed that each local cost function is strongly convex, while we assume that the global cost function satisfies the P–Ł condition, which is much weaker. Same as stated in Remark 7, compared with the results established in [23], one potential drawback of our results is that we need to use some global information, such as the eigenvalues of the communication graph.

Remark 12: By comparing Theorems 1 and 2 with Theorems 3 and 4, respectively, we see that, in theory, under the same conditions, the distributed L-ADMM algorithm (40) has the same convergence properties as those of the distributed ADMM algorithm (13). However, in numerical simulations, the distributed ADMM algorithm (13) normally requires less iterations than the distributed L-ADMM algorithm (40) to reach the same error bound at a cost of more computation resource being needed by each agent to solve the local optimization problem.

V. Simulations

This section evaluates the performance of Algorithms 1 and 2 in solving the phase retrieval problem [52].

Phase retrieval can be reformulated as the distributed optimization problem (1) with each component function f_i given by

$$f_i(x)=\frac{1}{m_i}\sum_{l=1}^{m_i}(y_{il}-|b_{il}^\top x|^2)^2$$
$$=\frac{1}{m_i}\sum_{l=1}^{m_i}(y_{il}-(x^\top b_{il}^R)^2-(x^\top b_{il}^I)^2)^2 \tag{44}$$

where m_i is the number of data points recorded by detector i, $b_{il}=b_{il}^R+\mathrm{i}b_{il}^I\in\mathbb{C}^p$ is the phase of the linear operator used in the lth measurement by detector i, and $y_{il}\in\mathbb{R}$ is the corresponding noisy squared magnitude.

All settings for cost functions and the communication graph are the same as those described in [53]. Specifically, $n=50$, $p=64$, and $m=30$. We independently and randomly generate the vectors b_{il}^R and b_{il}^I such that $(b_{il}^R,b_{il}^I)\sim\mathcal{N}(\mathbf{0}_{2p},\frac{1}{2}\mathbf{I}_{2p})$. The scalars y_{il} are generated by $y_{il}=|b_{il}^\top y_0|+\varepsilon_{i,l}$, where $y_0=(1,0,\dots,0)^\top$ and $\varepsilon_{i,l}\sim\mathcal{N}(0,0.01^2)$ are independent Gaussian noise. The graph used in the simulation is generated by uniformly randomly sampling n points on $\mathbb{S}^2$, and then connecting pairs of points with spherical distances less than $\pi/4$.

We compare Algorithms 1 and 2 with state-of-the-art algorithms: distributed gradient tracking algorithm (DGTA) [53], [54], distributed ADMM algorithm (Prox-PDA), and its linearized version (Prox-GPDA) [36]. Fig. 1 illustrates the convergence of $\min_{k\in[T]}\{\|\nabla f(\bar{x}_k)\|^2+\frac{1}{n}\sum_{i=1}^n\|x_{i,k}-\bar{x}_k\|^2\}$ with respect to the number of iterations T for these algorithms with the same initial condition. It can be seen that, in this numerical example, both distributed ADMM algorithms (Algorithms 1 and Prox-PDA) have almost the same performance and are better than the remaining algorithms. By comparing the two distributed

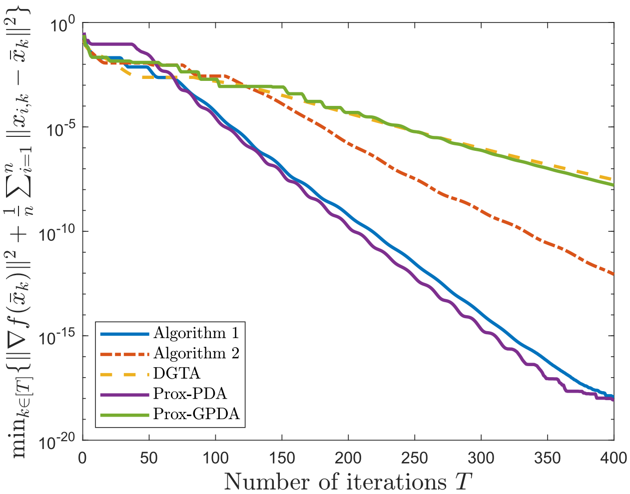


Fig. 1. Evolutions of $\min_{k\in[T]}\{\|\nabla f(\bar{x}_k)\|^2 + \frac{1}{n}\sum_{i=1}^n \|x_{i,k}-\bar{x}_k\|^2\}$ w.r.t. the number of iterations T.

L-ADMM algorithms (Algorithm 2 and Prox-GPDA), we see that Algorithm 2 converges faster. Moreover, Algorithm 2 also converges faster than DGTA.

VI. Conclusion

In this article, we proposed a ADMM algorithm to solve the distributed nonconvex optimization problem. We analyzed its convergence properties under different conditions. Especially, the linear convergence was established under the condition that the global cost function satisfies the P–Ł condition. Moreover, we extended the proposed distributed ADMM algorithm to a linearized version and established the same convergence properties under the same conditions. Interesting directions for future work include proving the convergence results for larger algorithm parameters, considering asynchronous and dynamic network setting, and studying constraints.

Appendix

A. Proof of Lemma 4

We first note that $V_{4,k}$ is well defined due to $f^* > -\infty$ as assumed in Assumption 2. Thus, V_k is well defined.

Noting $\gamma > L_f$, from Remark 3, we know that subproblem (13a) is solvable and $x_{i,k+1}$ is unique. Then noting first-order optimality conditions for convex optimization problem, we know that algorithm (13) can be rewritten as

$$x_{i,k+1} = x_{i,k} - \eta\left(\alpha\sum_{j=1}^n L_{ij}x_{j,k} + \beta v_{i,k} + \nabla f_i(x_{i,k+1})\right) \tag{45a}$$

$$v_{i,k+1} = v_{i,k} + \eta\beta\sum_{j=1}^n L_{ij}x_{j,k+1}$$

$$\forall x_{i,0} \in \mathbb{R}^p,\ \sum_{j=1}^n v_{j,0} = \mathbf{0}_p \tag{45b}$$

where $\eta = \frac{1}{\gamma}$.

Denote $\bar{v}_k = \frac{1}{n}(\mathbf{1}_n^\top \otimes \mathbf{I}_p)\boldsymbol{v}_k$. From (45b), we have $\bar{v}_{k+1} = \bar{v}_k$. Then, from $\sum_{i=1}^n v_{i,0} = \mathbf{0}_p$, we have $\bar{v}_k = \mathbf{0}_p$. Then, from (45a), we know that

$$\bar{\boldsymbol{x}}_{k+1} = \bar{\boldsymbol{x}}_k - \eta\bar{\boldsymbol{g}}_{k+1}. \tag{46}$$

Noting that $\nabla\tilde{f}$ is Lipschitz-continuous with constant $L_f > 0$ as assumed in Assumption 3, we have that

$$\begin{aligned}\|\boldsymbol{g}_k^0 - \boldsymbol{g}_k\|^2 &= \|\nabla\tilde{f}(\bar{\boldsymbol{x}}_k) - \nabla\tilde{f}(\boldsymbol{x}_k)\|^2\\ &\le L_f^2\|\bar{\boldsymbol{x}}_k - \boldsymbol{x}_k\|^2 = L_f^2\|\boldsymbol{x}_k\|^2_{\boldsymbol{K}}.\end{aligned} \tag{47}$$

We also have

$$\|\boldsymbol{g}_{k+1}^0 - \boldsymbol{g}_k^0\|^2 \le L_f^2\|\bar{\boldsymbol{x}}_{k+1} - \bar{\boldsymbol{x}}_k\|^2 = \eta^2L_f^2\|\bar{\boldsymbol{g}}_{k+1}\|^2 \tag{48}$$

where the equality holds due to (46). Then, we have

$$\begin{aligned}\|\boldsymbol{g}_k^0 - \boldsymbol{g}_{k+1}\|^2 &= \|\boldsymbol{g}_k^0 - \boldsymbol{g}_{k+1}^0 + \boldsymbol{g}_{k+1}^0 - \boldsymbol{g}_{k+1}\|^2\\ &\le 2\|\boldsymbol{g}_k^0 - \boldsymbol{g}_{k+1}^0\|^2 + 2\|\boldsymbol{g}_{k+1}^0 - \boldsymbol{g}_{k+1}\|^2\\ &\le 2\eta^2L_f^2\|\bar{\boldsymbol{g}}_{k+1}\|^2 + 2L_f^2\|\boldsymbol{x}_{k+1}\|^2_{\boldsymbol{K}}\end{aligned} \tag{49}$$

where the last inequality holds due to (47) and (48). Then, we have

$$\begin{aligned}&\|\bar{\boldsymbol{g}}_k^0 - \bar{\boldsymbol{g}}_{k+1}\|^2 = \|\boldsymbol{H}(\boldsymbol{g}_k^0 - \boldsymbol{g}_{k+1})\|^2 \le \|\boldsymbol{g}_k^0 - \boldsymbol{g}_{k+1}\|^2\\ &\le 2\eta^2L_f^2\|\bar{\boldsymbol{g}}_{k+1}\|^2 + 2L_f^2\|\boldsymbol{x}_{k+1}\|^2_{\boldsymbol{K}}\end{aligned} \tag{50}$$

where the first inequality holds due to $\rho(\boldsymbol{H}) = 1$; and the last inequality holds due to (49). Then, we have

$$\begin{aligned}\|\boldsymbol{x}_{k+1} - \boldsymbol{x}_k\|^2_{\boldsymbol{K}} &= \eta^2\|\alpha\boldsymbol{L}\boldsymbol{x}_k + \beta\boldsymbol{v}_k + \boldsymbol{g}_k^0 + \boldsymbol{g}_{k+1} - \boldsymbol{g}_k^0\|^2_{\boldsymbol{K}}\\ &\le 3\eta^2(\|\alpha\boldsymbol{L}\boldsymbol{x}_k\|^2 + \|\beta\boldsymbol{v}_k + \boldsymbol{g}_k^0\|^2_{\boldsymbol{K}} + \|\boldsymbol{g}_{k+1} - \boldsymbol{g}_k^0\|^2)\\ &\le \|\boldsymbol{x}_k\|^2_{3\eta^2\alpha^2\rho(L^2)\boldsymbol{K}} + \|\boldsymbol{y}_k\|^2_{3\eta^2\beta^2\boldsymbol{K}}\\ &\quad + 6\eta^4L_f^2\|\bar{\boldsymbol{g}}_{k+1}\|^2 + \|\boldsymbol{x}_{k+1}\|^2_{6\eta^2L_f^2\boldsymbol{K}}\end{aligned} \tag{51}$$

where the first equality holds due to (45a); the first inequality holds due to the Cauchy–Schwarz inequality, (2), and $\rho(\boldsymbol{K}) = 1$; and the last inequality holds due to (3) and (49).

We have

$$\begin{aligned}V_{1,k+1} &= \frac{1}{2}\|\boldsymbol{x}_k - \eta(\alpha\boldsymbol{L}\boldsymbol{x}_k + \beta\boldsymbol{v}_k + \boldsymbol{g}_{k+1})\|^2_{\boldsymbol{K}}\\ &= \frac{1}{2}\|\boldsymbol{x}_k\|^2_{\boldsymbol{K}} - \|\boldsymbol{x}_k\|^2_{\eta\alpha\boldsymbol{L} - \frac{\eta^2\alpha^2}{2}\boldsymbol{L}^2}\\ &\quad - \eta\beta\boldsymbol{x}_k^\top(\mathbf{I}_{np} - \eta\alpha\boldsymbol{L})\boldsymbol{K}\left(\boldsymbol{v}_k + \frac{1}{\beta}\boldsymbol{g}_k^0 + \frac{1}{\beta}\boldsymbol{g}_{k+1} - \frac{1}{\beta}\boldsymbol{g}_k^0\right)\\ &\quad + \frac{\eta^2\beta^2}{2}\|\boldsymbol{v}_k + \frac{1}{\beta}\boldsymbol{g}_k^0 + \frac{1}{\beta}\boldsymbol{g}_{k+1} - \frac{1}{\beta}\boldsymbol{g}_k^0\|^2_{\boldsymbol{K}}\\ &\le \frac{1}{2}\|\boldsymbol{x}_k\|^2_{\boldsymbol{K}} - \|\boldsymbol{x}_k\|^2_{\eta\alpha\boldsymbol{L} - \frac{\eta^2\alpha^2}{2}\boldsymbol{L}^2} - \eta\beta\boldsymbol{x}_k^\top\boldsymbol{K}\boldsymbol{y}_k\\ &\quad + \frac{\eta}{2}\|\boldsymbol{x}_k\|^2_{\boldsymbol{K}} + \frac{\eta}{2}\|\boldsymbol{g}_{k+1} - \boldsymbol{g}_k^0\|^2 + \frac{1}{2}\eta^2\alpha^2\|\boldsymbol{x}_k\|^2_{\boldsymbol{L}^2}\end{aligned}$$

$$
\begin{aligned}
&+\frac{1}{2}\eta^2\beta^2\|\boldsymbol{y}_k\|^2_{\boldsymbol{K}}+\frac{1}{2}\eta^2\alpha^2\|\boldsymbol{x}_k\|^2_{\boldsymbol{L}^2}\\
&+\frac{1}{2}\eta^2\|\boldsymbol{g}_{k+1}-\boldsymbol{g}^0_k\|^2\\
&+\eta^2\beta^2\|\boldsymbol{y}_k\|^2_{\boldsymbol{K}}+\eta^2\|\boldsymbol{g}_{k+1}-\boldsymbol{g}^0_k\|^2\\
\le\ &\frac{1}{2}\|\boldsymbol{x}_k\|^2_{\boldsymbol{K}}-\|\boldsymbol{x}_k\|^2_{\eta\alpha\rho_2(L)\boldsymbol{K}-\frac{\eta}{2}\boldsymbol{K}-\frac{3\eta^2\alpha^2}{2}\rho(L^2)\boldsymbol{K}}\\
&-\eta\beta\boldsymbol{x}^\top_{k+1}\boldsymbol{K}\boldsymbol{y}_k+\frac{1}{2}\|\boldsymbol{x}_{k+1}-\boldsymbol{x}_k\|^2_{\boldsymbol{K}}+\|\boldsymbol{y}_k\|^2_{2\eta^2\beta^2\boldsymbol{K}}\\
&+\|\boldsymbol{x}_{k+1}\|^2_{\eta(1+3\eta)L_f^2\boldsymbol{K}}+\eta^3(1+3\eta)L_f^2\|\bar{\boldsymbol{g}}_{k+1}\|^2
\end{aligned}\tag{52}
$$

where the first equality holds due to (45a); the second equality holds due to (2); the first inequality holds due to the Cauchy–Schwarz inequality and $\rho(\boldsymbol{K})=1$; and the last inequality holds due to (3) and (49).

We have

$$
\begin{aligned}
V_{2,k+1}=\ &\frac{1}{2}\|\boldsymbol{y}_k+\eta\beta\boldsymbol{L}\boldsymbol{x}_{k+1}+\frac{1}{\beta}(\boldsymbol{g}^0_{k+1}-\boldsymbol{g}^0_k)\|^2_{\boldsymbol{Q}+\frac{\alpha}{\beta}\boldsymbol{K}}\\
=\ &V_{2,k}+\eta\boldsymbol{x}^\top_{k+1}(\beta\boldsymbol{K}+\alpha\boldsymbol{L})\boldsymbol{y}_k\\
&+\|\boldsymbol{x}_{k+1}\|^2_{\frac{\eta^2\beta}{2}(\beta\boldsymbol{L}+\alpha\boldsymbol{L}^2)}+\frac{1}{2\beta^2}\|\boldsymbol{g}^0_{k+1}-\boldsymbol{g}^0_k\|^2_{\boldsymbol{Q}+\frac{\alpha}{\beta}\boldsymbol{K}}\\
&+\frac{1}{\beta}(\boldsymbol{y}_k+\eta\beta\boldsymbol{L}\boldsymbol{x}_{k+1})^\top(\boldsymbol{Q}+\frac{\alpha}{\beta}\boldsymbol{K})(\boldsymbol{g}^0_{k+1}-\boldsymbol{g}^0_k)\\
\le\ &V_{2,k}+\eta\boldsymbol{x}^\top_{k+1}(\beta\boldsymbol{K}+\alpha\boldsymbol{L})\boldsymbol{y}_k+\|\boldsymbol{x}_{k+1}\|^2_{\eta^2\beta(\beta\boldsymbol{L}+\alpha\boldsymbol{L}^2)}\\
&+\|\boldsymbol{y}_k\|^2_{\frac{\eta}{2\beta}(\boldsymbol{Q}+\frac{\alpha}{\beta}\boldsymbol{K})}+\left(\frac{1}{\beta^2}+\frac{1}{2\eta\beta}\right)\|\boldsymbol{g}^0_{k+1}-\boldsymbol{g}^0_k\|^2_{\boldsymbol{Q}+\frac{\alpha}{\beta}\boldsymbol{K}}\\
\le\ &V_{2,k}+\eta\boldsymbol{x}^\top_{k+1}(\beta\boldsymbol{K}+\alpha\boldsymbol{L})\boldsymbol{y}_k\\
&+\|\boldsymbol{x}_{k+1}\|^2_{\eta^2\beta(\beta\boldsymbol{L}+\alpha\boldsymbol{L}^2)}+\|\boldsymbol{y}_k\|^2_{\frac{\eta}{2\beta}(\boldsymbol{Q}+\frac{\alpha}{\beta}\boldsymbol{K})}\\
&+\left(\frac{1}{\beta^2}+\frac{1}{2\eta\beta}\right)\left(\frac{1}{\rho_2(L)}+\frac{\alpha}{\beta}\right)\|\boldsymbol{g}^0_{k+1}-\boldsymbol{g}^0_k\|^2\\
\le\ &V_{2,k}+\eta\boldsymbol{x}^\top_{k+1}(\beta\boldsymbol{K}+\alpha\boldsymbol{L})\boldsymbol{y}_k+\|\boldsymbol{y}_k\|^2_{\frac{\eta}{2\beta}(\frac{1}{\rho_2(L)}+\frac{\alpha}{\beta})\boldsymbol{K}}\\
&+\|\boldsymbol{x}_{k+1}\|^2_{\eta^2\beta(\beta\rho(L)+\alpha\rho(L^2))\boldsymbol{K}}\\
&+\eta\left(\frac{\eta}{\beta^2}+\frac{1}{2\beta}\right)\left(\frac{1}{\rho_2(L)}+\frac{\alpha}{\beta}\right)L_f^2\|\bar{\boldsymbol{g}}_{k+1}\|^2
\end{aligned}\tag{53}
$$

where the first equality holds due to (45b); the second equality holds due to (2) and (4); the first inequality holds due to the Cauchy–Schwarz inequality, (2), and (4); the second inequality holds due to $\rho(\boldsymbol{Q}+\frac{\alpha}{\beta}\boldsymbol{K})\le\rho(\boldsymbol{Q})+\frac{\alpha}{\beta}\rho(\boldsymbol{K})$, (5), $\rho(\boldsymbol{K})=1$; and the last inequality holds due to (3), (5), and (48).

We have

$$
\begin{aligned}
V_{3,k+1}=\ &\boldsymbol{x}^\top_{k+1}\boldsymbol{K}(\boldsymbol{v}_{k+1}+\frac{1}{\beta}\boldsymbol{g}^0_{k+1})\\
=\ &(\boldsymbol{x}_k-\eta(\alpha\boldsymbol{L}\boldsymbol{x}_k+\beta\boldsymbol{v}_k+\boldsymbol{g}^0_k+\boldsymbol{g}_{k+1}-\boldsymbol{g}^0_k))^\top\boldsymbol{K}\boldsymbol{y}_k\\
&+\boldsymbol{x}^\top_{k+1}\boldsymbol{K}(\eta\beta\boldsymbol{L}\boldsymbol{x}_{k+1}+\frac{1}{\beta}(\boldsymbol{g}^0_{k+1}-\boldsymbol{g}^0_k))\\
=\ &\boldsymbol{x}^\top_k(\boldsymbol{K}-\eta\alpha\boldsymbol{L})\boldsymbol{y}_k-\eta\beta\|\boldsymbol{y}_k\|^2_{\boldsymbol{K}}-\eta(\boldsymbol{g}_{k+1}-\boldsymbol{g}^0_k)^\top\boldsymbol{K}\boldsymbol{y}_k\\
&+\eta\beta\boldsymbol{x}^\top_{k+1}\boldsymbol{L}\boldsymbol{x}_{k+1}+\frac{1}{\beta}\boldsymbol{x}^\top_{k+1}\boldsymbol{K}(\boldsymbol{g}^0_{k+1}-\boldsymbol{g}^0_k)\\
\le\ &\boldsymbol{x}^\top_k(\boldsymbol{K}-\eta\alpha\boldsymbol{L})\boldsymbol{y}_k-\eta\beta\|\boldsymbol{y}_k\|^2_{\boldsymbol{K}}+\frac{\eta}{2}\|\boldsymbol{g}_{k+1}-\boldsymbol{g}^0_k\|^2\\
&+\frac{\eta}{2}\|\boldsymbol{y}_k\|^2_{\boldsymbol{K}}+\eta\beta\boldsymbol{x}^\top_{k+1}\boldsymbol{L}\boldsymbol{x}_{k+1}\\
&+\frac{\eta}{2}\|\boldsymbol{x}_{k+1}\|^2_{\boldsymbol{K}}+\frac{1}{2\eta\beta^2}\|\boldsymbol{g}^0_{k+1}-\boldsymbol{g}^0_k\|^2\\
\le\ &\boldsymbol{x}^\top_k\boldsymbol{K}\boldsymbol{y}_k-\eta\alpha(\boldsymbol{x}_k-\boldsymbol{x}_{k+1}+\boldsymbol{x}_{k+1})^\top\boldsymbol{L}\boldsymbol{y}_k\\
&-\|\boldsymbol{y}_k\|^2_{\eta(\beta-\frac{1}{2})\boldsymbol{K}}+\eta^3L_f^2\|\bar{\boldsymbol{g}}_{k+1}\|^2+\eta L_f^2\|x_{k+1}\|^2_{\boldsymbol{K}}\\
&+\|\boldsymbol{x}_{k+1}\|^2_{\frac{\eta}{2}\boldsymbol{K}+\eta\beta\boldsymbol{L}}+\frac{\eta L_f^2}{2\beta^2}\|\bar{\boldsymbol{g}}_{k+1}\|^2\\
\le\ &\boldsymbol{x}^\top_k\boldsymbol{K}\boldsymbol{y}_k-\eta\alpha\boldsymbol{x}^\top_{k+1}\boldsymbol{L}\boldsymbol{y}_k-\|\boldsymbol{y}_k\|^2_{\eta(\beta-\frac{1}{2})\boldsymbol{K}-\frac{\eta^2\alpha^2}{2}\boldsymbol{K}}\\
&+\frac{\rho(L^2)}{2}\|\boldsymbol{x}_{k+1}-\boldsymbol{x}_k\|^2_{\boldsymbol{K}}+(\eta^3+\frac{\eta}{2\beta^2})L_f^2\|\bar{\boldsymbol{g}}_{k+1}\|^2\\
&+\|\boldsymbol{x}_{k+1}\|^2_{\eta(\frac{1}{2}+L_f^2)\boldsymbol{K}+\eta\beta\rho(L)\boldsymbol{K}}
\end{aligned}\tag{54}
$$

where the second equality holds due to (45); the third equality holds due to (2); the first inequality holds due to the Cauchy–Schwarz inequality, (2), and $\rho(\boldsymbol{K})=1$; the second inequality holds due to (48) and (49); and the last inequality holds due to (3).

We have

$$
\begin{aligned}
V_{4,k+1}=\ &n(f(\bar{x}_{k+1})-f^*)=\tilde{f}(\bar{\boldsymbol{x}}_{k+1})-nf^*\\
=\ &\tilde{f}(\bar{\boldsymbol{x}}_k)-nf^*+\tilde{f}(\bar{\boldsymbol{x}}_{k+1})-\tilde{f}(\bar{\boldsymbol{x}}_k)\\
\le\ &\tilde{f}(\bar{\boldsymbol{x}}_k)-nf^*-\eta\bar{\boldsymbol{g}}^\top_{k+1}\boldsymbol{g}^0_k+\frac{\eta^2L_f}{2}\|\bar{\boldsymbol{g}}_{k+1}\|^2\\
=\ &n(f(\bar{x}_k)-f^*)-\frac{\eta}{2}\bar{\boldsymbol{g}}^\top_{k+1}(\bar{\boldsymbol{g}}_{k+1}+\bar{\boldsymbol{g}}^0_k-\bar{\boldsymbol{g}}_{k+1})\\
&-\frac{\eta}{2}(\bar{\boldsymbol{g}}_{k+1}-\bar{\boldsymbol{g}}^0_k+\bar{\boldsymbol{g}}^0_k)^\top\bar{\boldsymbol{g}}^0_k+\frac{\eta^2L_f}{2}\|\bar{\boldsymbol{g}}_{k+1}\|^2\\
\le\ &n(f(\bar{x}_k)-f^*)-\frac{\eta}{4}\|\bar{\boldsymbol{g}}_{k+1}\|^2+\frac{\eta}{4}\|\bar{\boldsymbol{g}}^0_k-\bar{\boldsymbol{g}}_{k+1}\|^2\\
&-\frac{\eta}{4}\|\bar{\boldsymbol{g}}^0_k\|^2+\frac{\eta}{4}\|\bar{\boldsymbol{g}}^0_k-\bar{\boldsymbol{g}}_{k+1}\|^2+\frac{\eta^2L_f}{2}\|\bar{\boldsymbol{g}}_{k+1}\|^2\\
\le\ &n(f(\bar{x}_k)-f^*)-\frac{\eta}{4}(1-2\eta L_f-4\eta^2L_f^2)\|\bar{\boldsymbol{g}}_{k+1}\|^2\\
&+\|\boldsymbol{x}_{k+1}\|^2_{\eta L_f^2\boldsymbol{K}}-\frac{\eta}{4}\|\bar{\boldsymbol{g}}^0_k\|^2
\end{aligned}\tag{55}
$$

where the first inequality holds since $\tilde{f}$ is smooth, (7) and (46); the third equality holds due to $\bar{\boldsymbol{g}}^\top_{k+1}\boldsymbol{g}^0_k=\boldsymbol{g}^\top_{k+1}\boldsymbol{H}\boldsymbol{g}^0_k=\boldsymbol{g}^\top_{k+1}\boldsymbol{H}\boldsymbol{H}\boldsymbol{g}^0_k=\bar{\boldsymbol{g}}^\top_{k+1}\bar{\boldsymbol{g}}^0_k$; and the last inequality holds due to (50).

Finally, from (51)–(55), we have (14).

References

[1] P. Bianchi and J. Jakubowicz, "Convergence of a multi-agent projected stochastic gradient algorithm for non-convex optimization," *IEEE Trans. Autom. Control*, vol. 58, no. 2, pp. 391–405, Feb. 2012.

[2] P. A. Forero, A. Cano, and G. B. Giannakis, "Distributed clustering using wireless sensor networks," *IEEE J. Sel. Topics Signal Process.*, vol. 5, no. 4, pp. 707–724, Aug. 2011.

[3] S. Patterson, Y. C. Eldar, and I. Keidar, "Distributed compressed sensing for static and time-varying networks," *IEEE Trans. Signal Process.*, vol. 62, no. 19, pp. 4931–4946, Oct. 2014.

[4] H.-T. Wai, T.-H. Chang, and A. Scaglione, "A consensus-based decentralized algorithm for non-convex optimization with application to dictionary learning," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, 2015, pp. 3546–3550.

[5] L. Bottou, F. E. Curtis, and J. Nocedal, "Optimization methods for large-scale machine learning," *SIAM Rev.*, vol. 60, no. 2, pp. 223–311, 2018.

[6] M. Zhu and S. Martínez, "An approximate dual subgradient algorithm for multi-agent non-convex optimization," *IEEE Trans. Autom. Control*, vol. 58, no. 6, pp. 1534–1539, Jun. 2013.

[7] P. D. Lorenzo and G. Scutari, "NEXT: In-network nonconvex optimization," *IEEE Trans. Signal Inf. Process. Netw.*, vol. 2, no. 2, pp. 120–136, Jun. 2016.

[8] N. Chatzipanagiotis and M. M. Zavlanos, "On the convergence of a distributed augmented Lagrangian method for nonconvex optimization," *IEEE Trans. Autom. Control*, vol. 62, no. 9, pp. 4405–4420, Sep. 2017.

[9] P. Xu, F. Roosta, and M. W. Mahoney, "Newton-type methods for non-convex optimization under inexact Hessian information," *Math. Program.*, vol. 184, no. 1, pp. 35–70, Nov. 2020.

[10] M. Hong, M. Razaviyayn, and J. Lee, "Gradient primal-dual algorithm converges to second-order stationary solution for nonconvex distributed optimization over networks," in *Proc. Int. Conf. Mach. Learn.*, 2018, pp. 2009–2018.

[11] J. Zeng and W. Yin, "On nonconvex decentralized gradient descent," *IEEE Trans. Signal Process.*, vol. 66, no. 11, pp. 2834–2848, Jun. 2018.

[12] D. Hajinezhad and M. Hong, "Perturbed proximal primal–dual algorithm for nonconvex nonsmooth optimization," *Math. Program.*, vol. 176, no. 1/2, pp. 207–245, 2019.

[13] B. Swenson, R. Murray, H. V. Poor, and S. Kar, "Distributed gradient descent: Nonconvergence to saddle points and the stable-manifold theorem," in *Proc. Annu. Allerton Conf. Commun. Control Comput.*, 2019, pp. 595–601.

[14] J. George, T. Yang, H. Bai, and P. Gurram, "Distributed stochastic gradient method for non-convex problems with applications in supervised learning," in *Proc. IEEE Conf. Decis. Control*, 2019, pp. 5538–5543.

[15] H. Sun and M. Hong, "Distributed non-convex first-order optimization and information processing: Lower complexity bounds and rate optimal algorithms," *IEEE Trans. Signal Process.*, vol. 67, no. 22, pp. 5912–5928, Nov. 2019.

[16] X. Yi, S. Zhang, T. Yang, T. Chai, and K. H. Johansson, "Linear convergence of first-and zeroth-order primal-dual algorithms for distributed nonconvex optimization," *IEEE Trans. Autom. Control*, early access, Sep. 10, 2021, doi: 10.1109/TAC.2021.3108501.

[17] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Found. Trends Mach. Learn.*, vol. 3, no. 1, pp. 1–122, 2011.

[18] R. Zhang and J. Kwok, "Asynchronous distributed ADMM for consensus optimization," in *Proc. Int. Conf. Mach. Learn.*, 2014, pp. 1701–1709.

[19] W. Deng and W. Yin, "On the global and linear convergence of the generalized alternating direction method of multipliers," *J. Sci. Comput.*, vol. 66, no. 3, pp. 889–916, 2016.

[20] E. Wei and A. Ozdaglar, "Distributed alternating direction method of multipliers," in *Proc. IEEE Conf. Decis. Control*, 2012, pp. 5445–5450.

[21] A. Makhdoumi and A. Ozdaglar, "Broadcast-based distributed alternating direction method of multipliers," in *Proc. Annu. Allerton Conf. Commun. Control Comput.*, 2014, pp. 270–277.

[22] W. Shi, Q. Ling, K. Yuan, G. Wu, and W. Yin, "On the linear convergence of the ADMM in decentralized consensus optimization," *IEEE Trans. Signal Process.*, vol. 62, no. 7, pp. 1750–1761, Apr. 2014.

[23] Q. Ling, W. Shi, G. Wu, and A. Ribeiro, "DLM: Decentralized linearized alternating direction method of multipliers," *IEEE Trans. Signal Process.*, vol. 63, no. 15, pp. 4051–4064, Aug. 2015.

[24] F. Iutzeler, P. Bianchi, P. Ciblat, and W. Hachem, "Explicit convergence rate of a distributed alternating direction method of multipliers," *IEEE Trans. Autom. Control*, vol. 61, no. 4, pp. 892–904, Apr. 2016.

[25] T.-H. Chang, "A proximal dual consensus ADMM method for multi-agent constrained optimization," *IEEE Trans. Signal Process.*, vol. 64, no. 14, pp. 3719–3734, Jul. 2016.

[26] A. Mokhtari, W. Shi, Q. Ling, and A. Ribeiro, "DQM: Decentralized quadratically approximated alternating direction method of multipliers," *IEEE Trans. Signal Process.*, vol. 64, no. 19, pp. 5158–5173, Oct. 2016.

[27] N. S. Aybat, Z. Wang, T. Lin, and S. Ma, "Distributed linearized alternating direction method of multipliers for composite convex consensus optimization," *IEEE Trans. Autom. Control*, vol. 63, no. 1, pp. 5–20, Jan. 2018.

[28] A. Makhdoumi and A. Ozdaglar, "Convergence rate of distributed ADMM over networks," *IEEE Trans. Autom. Control*, vol. 62, no. 10, pp. 5082–5095, Oct. 2017.

[29] M. Maros and J. Jaldén, "On the Q-linear convergence of distributed generalized ADMM under non-strongly convex function components," *IEEE Trans. Signal Inf. Process. Netw.*, vol. 5, no. 3, pp. 442–453, Sep. 2019.

[30] Y. Liu, W. Xu, G. Wu, Z. Tian, and Q. Ling, "Communication-censored ADMM for decentralized consensus optimization," *IEEE Trans. Signal Process.*, vol. 67, no. 10, pp. 2565–2579, May 2019.

[31] A. Falsone, I. Notarnicola, G. Notarstefano, and M. Prandini, "Tracking-ADMM for distributed constraint-coupled optimization," *Automatica*, vol. 117, 2020, Art. no. 108962.

[32] R. Carli and M. Dotoli, "Distributed alternating direction method of multipliers for linearly constrained optimization over a network," *IEEE Control Syst. Lett.*, vol. 4, no. 1, pp. 247–252, Jan. 2020.

[33] M. Hong, Z.-Q. Luo, and M. Razaviyayn, "Convergence analysis of alternating direction method of multipliers for a family of nonconvex problems," *SIAM J. Optim.*, vol. 26, no. 1, pp. 337–364, 2016.

[34] T.-H. Chang, M. Hong, W.-C. Liao, and X. Wang, "Asynchronous distributed ADMM for large-scale optimization–Part I: Algorithm and convergence analysis," *IEEE Trans. Signal Process.*, vol. 64, no. 12, pp. 3118–3130, Jun. 2016.

[35] M. Hong, "A distributed, asynchronous, and incremental algorithm for nonconvex optimization: An ADMM approach," *IEEE Trans. Control Netw. Syst.*, vol. 5, no. 3, pp. 935–945, Sep. 2018.

[36] M. Hong, D. Hajinezhad, and M.-M. Zhao, "Prox-PDA: The proximal primal-dual algorithm for fast distributed nonconvex optimization and learning over networks," in *Proc. Int. Conf. Mach. Learn.*, 2017, pp. 1529–1538.

[37] X. Yi, L. Yao, T. Yang, J. George, and K. H. Johansson, "Distributed optimization for second-order multi-agent systems with dynamic event-triggered communication," in *Proc. IEEE Conf. Decis. Control*, 2018, pp. 3397–3402.

[38] Y. Nesterov, *Lectures on Convex Optimization*, 2nd ed. Berlin, Germany: Springer, 2018.

[39] H. Karimi, J. Nutini, and M. Schmidt, "Linear convergence of gradient and proximal-gradient methods under the Polyak-Łojasiewicz condition," in *Proc. Joint Eur. Conf. Mach. Learn. Knowl. Discov. Databases*, 2016, pp. 795–811.

[40] Z. Li and J. Li, "A simple proximal stochastic gradient method for nonsmooth nonconvex optimization," in *Proc. Adv. Neural Inf. Process. Syst.*, 2018, pp. 5569–5579.

[41] K. Zhong, Z. Song, P. Jain, P. L. Bartlett, and I. S. Dhillon, "Recovery guarantees for one-hidden-layer neural networks," in *Proc. Int. Conf. Mach. Learn.*, 2017, pp. 4140–4149.

[42] H. Fu, Y. Chi, and Y. Liang, "Local geometry of one-hidden-layer neural networks for logistic regression," 2018, *arXiv:1802.06463*.

[43] J. Zhang and K. You, "Fully asynchronous distributed optimization with linear convergence in directed networks," 2019, *arXiv:1901.08215v3*.

[44] R. Xin, U. A. Khan, and S. Kar, "An improved convergence analysis for decentralized online stochastic non-convex optimization," *IEEE Trans. Signal Process.*, vol. 69, pp. 1842–1858, Mar. 2021.

[45] M. Fazel, R. Ge, S. Kakade, and M. Mesbahi, "Global convergence of policy gradient methods for the linear quadratic regulator," in *Proc. Int. Conf. Mach. Learn.*, 2018, pp. 1467–1476.

[46] M. Schuresko and J. Cortés, "Distributed motion constraints for algebraic connectivity of robotic networks," *J. Intell. Robotic Syst.*, vol. 56, no. 1–2, pp. 99–126, 2009.

[47] T. M. D. Tran and A. Y. Kibangou, "Consensus-based distributed estimation of Laplacian eigenvalues of undirected graphs," in *Proc. Eur. Control Conf.*, 2013, pp. 227–232.

[48] T.-H. Chang, M. Hong, and X. Wang, "Multi-agent distributed optimization via inexact consensus ADMM," *IEEE Trans. Signal Process.*, vol. 63, no. 2, pp. 482–497, Jun. 2015.

[49] T.-H. Chang, M. Hong, H.-T. Wai, X. Zhang, and S. Lu, "Distributed learning in the nonconvex world: From batch data to streaming and beyond," *IEEE Signal Process. Mag.*, vol. 37, no. 3, pp. 26–38, May 2020.

[50] Z. Wang, J. Zhang, T.-H. Chang, J. Li, and Z.-Q. Luo, "Distributed stochastic consensus optimization with momentum for nonconvex nonsmooth problems," *IEEE Trans. Signal Process.*, vol. 69, pp. 4486–4501, Jul. 2021.

[51] W. Shi, Q. Ling, G. Wu, and W. Yin, "EXTRA: An exact first-order algorithm for decentralized consensus optimization," *SIAM J. Optim.*, vol. 25, no. 2, pp. 944–966, 2015.

[52] E. J. Candes, T. Strohmer, and V. Voroninski, "PhaseLift: Exact and stable signal recovery from magnitude measurements via convex programming," *Commun. Pure Appl. Math.*, vol. 66, no. 8, pp. 1241–1274, 2013.

[53] Y. Tang, J. Zhang, and N. Li, "Distributed zero-order algorithms for nonconvex multiagent optimization," *IEEE Trans. Control Netw. Syst.*, vol. 8, no. 1, pp. 269–281, Mar. 2021.

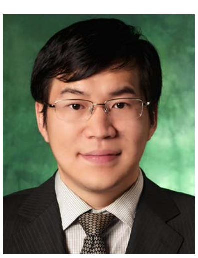
[54] G. Qu and N. Li, "Harnessing smoothness to accelerate distributed optimization," *IEEE Trans. Control Netw. Syst.*, vol. 5, no. 3, pp. 1245–1260, Sep. 2018.

Xinlei Yi received the B.S. and M.S. degrees in mathematics from the China University of Geoscience, Wuhan, China, and Fudan University, Shanghai, China, in 2011 and 2014, respectively, and the Ph.D. degree in electrical engineering from the School of Electrical Engineering and Computer Science, KTH Royal Institute of Technology, Stockholm, Sweden, in 2020.

He is currently a Postdoc with the KTH Royal Institute of Technology. His current research interests include online optimization, distributed optimization, and event-triggered control.

Shengjun Zhang received the B.Eng. degree in automation of honors program from China Agricultural University, Beijing, China, in 2014, the M.S. degree in electrical engineering from New York University, New York, NY, USA, in 2017, and the Ph.D. degree in electrical engineering from the Department of Electrical Engineering, College of Engineering, University of North Texas, Denton, TX, USA, in 2022.

He is currently a Postdoctoral Researcher with the Huazhong University of Science and Technology, Wuhan, China. His current research interests include optimization and applications in artificial intelligence.



Tao Yang (Senior Member, IEEE) received the Ph.D. degree in electrical engineering from Washington State University, Pullman, WA, USA, in 2012.

He is currently a Professor with the State Key Laboratory of Synthetical Automation for Process Industries, Northeastern University, Boston, MA, USA. He was an Assistant Professor with the Department of Electrical Engineering, University of North Texas, Denton, TX, USA, from 2016 to 2019. Between 2012 and 2014, he was an ACCESS Postdoctoral Researcher with the ACCESS Linnaeus Centre, Royal Institute of Technology, Stockholm, Sweden. He then joined the Pacific Northwest National Laboratory, Richland, WA, USA, as a Postdoctoral Researcher, and was promoted to Scientist/Engineer II in 2015. His research interests include industrial artificial intelligence, integrated optimization and control, distributed control and optimization with applications to process industries, cyber–physical systems, and networked control systems.

Dr. Yang is an Associate Editor for IEEE TRANSACTIONS ON CONTROL SYSTEMS TECHNOLOGY, IEEE TRANSACTIONS ON NEURAL NETWORKS AND LEARNING SYSTEMS, and *IEEE/CAA Journal of Automatica Sinica*. He is a member of several technical committees of the IEEE Control Systems Society and IFAC. He was a recipient of Ralph E. Powe Junior Faculty Enhancement Award and Best Student Paper award (as an advisor) of the *14th IEEE International Conference on Control & Automation* in 2018.

Tianyou Chai (Life Fellow, IEEE) received the Ph.D. degree in control theory and engineering from Northeastern University, Shenyang, China, in 1985.

He became a Professor with Northeastern University in 1988. He is the Founder and Director of the Center of Automation, which became a National Engineering and Technology Research Center and a State Key Laboratory. He has served as Director with the Department of Information Science, National Natural Science Foundation of China, Beijing, China, from 2010 to 2018. His current research interests include modeling, control, optimization, and integrated automation of complex industrial processes. He has authored or coauthored 297 peer-reviewed international journal papers. He has developed control technologies with applications to various industrial processes.

Dr. Chai's paper titled "Hybrid intelligent control for optimal operation of shaft furnace roasting process" was selected as one of three best papers for the Control Engineering Practice Paper Prize for 2011–2013. For his contributions, he has won five prestigious awards of National Natural Science, National Science and Technology Progress and National Technological Innovation, the 2007 Industry Award for Excellence in Transitional Control Research from IEEE Multiple-Conference on Systems and Control, and the 2017 Wook Hyun Kwon Education Award from the Asian Control Association. He is a member of Chinese Academy of Engineering and an IFAC Fellow.

Karl Henrik Johansson (Fellow, IEEE) received the M.Sc. degree in electrical engineering and the Ph.D. degree in automatic control from Lund University, Lund, Sweden, in 1992 and 1997, respectively.

He is currently a Professor with the School of Electrical Engineering and Computer Science, KTH Royal Institute of Technology, Stockholm, Sweden. He has held visiting positions at UC Berkeley, Berkeley, CA, USA; Caltech, Pasadena, CA, USA; NTU, Singapore; HKUST Institute of Advanced Studies, Hong Kong; and NTNU, Trondheim, Norway. His research interests include networked control systems, cyber–physical systems, and applications in transportation, energy, and automation.

Dr. Johansson has served on the IEEE Control Systems Society Board of Governors, the IFAC Executive Board, and the European Control Association Council. He is the recipient of several best paper awards and other distinctions from IEEE and ACM. He has been awarded Distinguished Professor with the Swedish Research Council and Wallenberg Scholar with the Knut and Alice Wallenberg Foundation. He is the recipient of the Future Research Leader Award from the Swedish Foundation for Strategic Research and the triennial Young Author Prize from IFAC. He is a Fellow of the Royal Swedish Academy of Engineering Sciences and an IEEE Control Systems Society Distinguished Lecturer.