Distributed Nonconvex Optimization With Event-Triggered Communication

Lei Xu, Xinlei Yi, Yang Shi, *Fellow, IEEE*, Karl H. Johansson, *Fellow, IEEE*, Tianyou Chai, *Life Fellow, IEEE*, and Tao Yang, *Senior Member, IEEE*

***Abstract*—This article considers distributed nonconvex optimization for minimizing the sum of local cost functions by using local information exchange. In order to avoid continuous communication among agents and reduce communication overheads, we develop a distributed algorithm with a dynamic exponentially decaying event-triggered scheme. We show that the proposed algorithm is free of Zeno behavior (i.e., finite number of triggers in any finite time interval) by contradiction and asymptotically converges to a stationary point if the local cost functions are smooth. Moreover, we show that the proposed algorithm exponentially converges to the global optimal point if, in addition, the global cost function satisfies the Polyak–Łojasiewicz condition, which is weaker than the standard strong convexity condition, and the global minimizer is not necessarily unique. The theoretical results are illustrated by a numerical simulation example.**

***Index Terms*—Distributed nonconvex algorithm, event-triggered communication, exponential convergence, Polyak–Łojasiewicz (P–Ł) condition, Zeno behavior.**

I. Introduction

Distributed optimization can be traced back to [1] and [2]. Recent years have witnessed an increasing interest in distributed optimization, due to their applications in power systems, sensor networks, and machine learning. Much attention has been devoted to developing discrete-time distributed optimization algorithms, e.g., survey papers [3], [4] and references therein. Since many practical systems, such as robots and unmanned vehicles, operate in continuous time [5], [6], various continuous-time distributed optimization algorithms have also been proposed. For example, Wang and Elia [7] developed a distributed algorithm based on the proportional–integral (PI) control strategy and established its asymptotic convergence for undirected graphs. The extension to directed weight-balanced graphs was considered in [8]. Kia et al. [9] proposed an alternative distributed PI algorithm with less communication and showed that the proposed algorithm exponentially converges to a neighborhood of the global minimizer. Varagnolo et al. [10] proposed a distributed zero-gradient-sum (ZGS) algorithm and established its exponential convergence if local cost functions are strongly convex and have locally Lipschitz Hessians. Distributed Newton–Raphson algorithms were developed in [11] and [12]. All the aforementioned studies focused on the case where the local cost functions are strongly convex. The extension to the nonstrongly convex case was considered in [13] and [14]. In many applications, however, such as traffic flow management problems, wind farm planning problems, and resource allocation problems, the cost functions are usually nonconvex; hence, the nonconvex case was studied in [15], [16], and [17].

However, distributed algorithms proposed in these studies require continuous information exchange among agents, which may be impractical in physical applications. In order to avoid continuous communication and reduce communication overheads, by using the idea of event-triggered communication and control [18], [19], [20], [21], [22], various distributed event-triggered optimization algorithms have been developed. A key challenge for developing distributed event-triggered algorithms is to design triggering laws to ensure that the algorithms are free of Zeno behavior, i.e., infinite number of triggers in a finite time interval [23]. For example, Kia et al. [9] proposed a distributed PI algorithm with an event-triggered mechanism. The distributed ZGS algorithm proposed in [10] was equipped with static and dynamic event-triggered mechanisms in [24] and [25], respectively. Liu et al. [26] proposed a distributed algorithm with event-triggered quantized communications.

Note that most of the existing distributed event-triggered algorithms focused on the convex optimization. George and Gurram [27] proposed a discrete-time distributed event-triggered algorithm for addressing nonconvex optimization challenges commonly encountered in distributed deep learning. Considering that most of the practical systems operate in continuous time, this motivates us to propose a continuous-time distributed event-triggered algorithm for solving nonconvex optimization problems. The contributions of this article are as follows.

1) We propose a distributed event-triggered algorithm for solving nonconvex optimization over undirected connected graphs. For smooth local cost functions, we show that the proposed algorithm does not have Zeno behavior and asymptotically converges to a stationary point.
2) If, in addition, the global cost function satisfies the Polyak–Łojasiewicz (P–Ł) condition, which is weaker than the strong convexity condition, and the global minimizer is not necessarily

Manuscript received 13 June 2023; revised 12 November 2023; accepted 18 November 2023. Date of publication 5 December 2023; date of current version 29 March 2024. This work was supported in part by the National Natural Science Foundation of China under Grant 62133003, Grant 61991403, and Grant 61991400, in part by the Knut and Alice Wallenberg Foundation, in part by the Swedish Foundation for Strategic Research, and in part by the Swedish Research Council. Recommended by Associate Editor Y. Wang. *(Corresponding author: Tao Yang.)*

Lei Xu is with the State Key Laboratory of Synthetical Automation for Process Industries, Northeastern University, Shenyang 110819, China, and also with the Department of Mechanical Engineering, University of Victoria, Victoria, BC V8W 2Y2, Canada (e-mail: 2010345@stu.neu.edu.cn).

Xinlei Yi is with the Lab for Information and Decision Systems, Massachusetts Institute of Technology, Cambridge MA 02139 USA (e-mail: xinleiyi@mit.edu).

Yang Shi is with the Department of Mechanical Engineering, University of Victoria, Victoria, BC V8W 2Y2, Canada (e-mail: yshi@uvic.ca).

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

Karl H. Johansson is with the Division of Decision and Control Systems, School of Electrical Engineering and Computer Science, KTH Royal Institute of Technology, 100 44 Stockholm, Sweden (e-mail: kallej@kth.se).

Digital Object Identifier 10.1109/TAC.2023.3339439

unique, we show that the proposed algorithm exponentially converges to a global optimal point.

To the best of our knowledge, this is among the first continuous-time distributed event-triggered algorithms to solve the distributed nonconvex optimization problem. It should be highlighted that compared with convex optimization, nonconvex optimization is more challenging since the useful properties of the convex condition, for example, the set of optimal points of a convex function is convex, and the (sub)gradient of a convex function is a monotonic mapping, cannot be used. The innovation in this article's algorithm analysis is the design of a suitable Lyapunov function to prove the convergence of the proposed distributed nonconvex optimization algorithm with event-triggered communication.

Compared with [7], [8], [10], and [13], which proposed distributed convex optimization algorithms with continuous communications, this article not only considers the more general nonconvex optimization but also proposes an event-triggered mechanism to reduce communication burden. More specifically, algorithms proposed in [7], [8], and [10] require the global cost function to be strongly convex, and the algorithm in [13] requires the global cost function to satisfy the restricted secant inequality condition, while the global cost function in our work can be nonconvex. Compared with [9] and [24], which proposed distributed event-triggered convex optimization algorithms for the strongly convex case, this article establishes exponential convergence under weaker conditions, i.e., the global cost function satisfies the P–Ł condition, which does not require the cost function to be convex. In contrast to [27], where both the local cost function and its gradient were required to be Lipschitz continuous, this article necessitates only the Lipschitz continuity of the gradients of local cost functions. In addition, we also consider the case where the global cost function satisfies the P–Ł condition.

The rest of this article is organized as follows. Section II presents the problem formulation and motivation. Section III proposes a distributed event-triggered algorithm and analyzes its convergence results. Section IV presents a numerical simulation example. Finally, Section V concludes this article. To improve the readability, all the proofs are given in the Appendixes.

Notation: Let $\mathbf{1}_n$ (or $\mathbf{0}_n$) be the $n \times 1$ vector with all ones (or zeros), and $\mathbf{I}_n$ be the n-dimensional identity matrix. $\|\cdot\|$ is the Euclidean vector norm or spectral matrix norm, and null$(\cdot)$ refers to the null space. For any positive-semidefinite matrix $\mathcal{C}$, with appropriate dimensions vectors x and v, we define $\mathcal{Q}_{\mathcal{C}}(x) = x^T\mathcal{C}x$, $\langle x, v\rangle_{\mathcal{C}} = x^T\mathcal{C}v$, and $\rho(\mathcal{C})$ and $\underline{\rho}(\mathcal{C})$ are the spectral radius and the minimum positive eigenvalue of matrix $\mathcal{C}$, respectively. Let $\text{diag}[a_1, \dots, a_n]$ denote a diagonal matrix with the ith diagonal element being a_i. For a differentiable function f, ∇f and $\nabla^2 f$ refer to the gradient and Hessian of f, respectively. $A \otimes B$ represents the Kronecker product of matrices A and B.

II. Problem Formulation and Motivation

Consider a group of n agents distributed on fixed undirected graph $\mathcal{G} = (\mathcal{V}, \mathcal{E}, \mathcal{A})$, where $\mathcal{V} = \{1, 2, \dots, n\}$ is the agent set and $\mathcal{E} \subseteq \mathcal{V} \times \mathcal{V}$ is the edge set. $(v_j, v_i) \in \mathcal{E}$ indicates that the agents v_j and v_i can communicate with each other. $\mathcal{A} = [a_{ij}] \in \mathbb{R}^{n\times n}$ is the adjacency matrix, where $a_{ij} > 0$ if $(v_j, v_i) \in \mathcal{E}$; otherwise, $a_{ij} = 0$. Let $\mathcal{N}_i = \{v_j \in \mathcal{V} : a_{ij} > 0\}$ and $d_i = \sum_{j=1}^n a_{ij}$ denote the neighbor set and weighted degree of agent i, respectively. The degree matrix is defined as $\mathcal{D} = \text{diag}[d_1, \dots, d_n]$, and the graph Laplacian matrix is $\mathcal{L} = \mathcal{D} - \mathcal{A}$. A path from agents i_1 to i_k is a sequence of agents $i_1, \dots, i_k$ such that $(i_j, i_{j+1}) \in \mathcal{E}$ for $j = 1, \dots, k-1$. An undirected graph is said to be connected if there exists a path between any pair of distinct agents.

Assume that each agent has a private local cost function $f_i : \mathbb{R}^p \to \mathbb{R}$, the optimal set $\mathbb{X}^\star = \text{argmin}_{x\in\mathbb{R}^p} f(x)$ is nonempty, and $f^\star = \min_{x\in\mathbb{R}^p} f(x) > -\infty$. The objective is to find an optimizer $x^\star$ to minimize the average of all local cost functions, i.e.,

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

Various continuous-time distributed optimization algorithms have been developed [7], [8], [9], [10], [11], [12], [13], [14]. Noting that communication resources are limited, distributed event-triggered algorithms have been proposed to avoid continuous communication and reduce the communication burden [9], [24], [25], [26], [28]. However, these algorithms require the local cost functions to be convex. In many applications, the cost functions are usually nonconvex. This motivates us to develop a distributed event-triggered algorithm for the nonconvex case.

Throughout this article, we make the following assumptions.

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

Assumption 2: Each local cost function $f_i(x)$ is twice continuously differentiable and smooth with constant $L_f > 0$, i.e.,

$$\|\nabla f_i(x) - \nabla f_i(y)\| \le L_f\|x - y\| \quad \forall x, y \in \mathbb{R}^p. \tag{2}$$

Remark 1: Assumptions 1 and 2 are common in the literature (see, e.g., [3] and [4]). Note that the convexity of the local cost functions is not assumed. Moreover, it follows from [29, Lemma 1.2.2] that Assumption 2 implies

$$\|\nabla^2 f_i(x)\| \le L_f \text{ and } \|\nabla^2 f(x)\| \le L_f \quad \forall x \in \mathbb{R}^p. \tag{3}$$

III. Main Results

In this section, we first propose a distributed event-triggered algorithm to solve the distributed nonconvex optimization problem (1). We then analyze its convergence properties.

Our proposed algorithm is as follows:

$$\dot{x}_i(t) = -\alpha\sum_{j=1}^n L_{ij}x_j(t^j_{k_j(t)}) - \beta v_i(t) - \nabla f_i(x_i(t)) \tag{4a}$$

$$\dot{v}_i(t) = \beta\sum_{j=1}^n L_{ij}x_j(t^j_{k_j(t)}) \quad \forall x_i(0) \in \mathbb{R}^p$$

$$\sum_{j=1}^n v_j(0) = \mathbf{0}_p, \quad i \in \mathcal{V} \tag{4b}$$

where $\alpha > 0$ and $\beta > 0$ are gain parameters, and the sequence $\{t^j_k\}_{k=1}^\infty$, $\forall j \in \mathcal{V}$ is the triggering times for agent j to be determined later and $t^j_{k_j(t)} = \max\{t^j_k : t^j_k \le t\}$. Without loss of generality, we assume $t^j_1 = 0 \ \forall j \in \mathcal{V}$. The algorithm is motivated by the PI control strategy. More specifically, in (4a), the term $-\nabla f_i(x_i(t))$ ensures that each agent follows its local gradient descent, and the term $\sum_{j=1}^n L_{ij}x_j(t^j_{k_j(t)})$ ensures that consensus is achieved among agents. However, if the dynamics just contains these two terms, the agents' states would not converge since the local gradients are not the same in general. Thus, to correct the error, the additional integral feedback term $v_i(t)$ whose dynamics is governed by (4b) is introduced.

We present the following convergence result.

Theorem 1: Let Assumptions 1 and 2 hold. If each agent $i \in \mathcal{V}$ runs the distributed event-triggered algorithm (4) and determines its triggering time sequence by

$$t^i_{k+1} = \max_{t\ge t^i_k}\left\{t : \|x_i(t) - x_i(t^i_k)\| \le a_i e^{-b_i t}\right\},$$

$$k = 1, 2, \ldots \quad (5)$$

where $a_i > 0$ and $b_i > 0$ are designed parameters, then: 1) there is no Zeno behavior and 2) if $\alpha \in [3\beta + \kappa_1, \kappa_2\beta]$, $\beta > \max\{\frac{\kappa_1}{\kappa_2 - 3}, \kappa_3\}$, then the algorithm asymptotically converges to a stationary point, i.e.,

$$\lim_{t\to\infty} \left(\frac{1}{n} \sum_{i=1}^{n} \|x_i(t) - \bar{x}(t)\| + \|\nabla f(\bar{x}(t))\| \right) = 0 \quad (6)$$

where $\bar{x}(t) = \frac{1}{n}\sum_{i=1}^{n} x_i(t)$ and

$$\kappa_1 = \left[\frac{1}{\kappa_3}\left(4L_f^4 + 1\right) + 4L_f^2 + 1\right] \frac{1}{\underline{\rho}(L)}$$

$$\kappa_2 = -\frac{1}{\underline{\rho}(L)} + \sqrt{\kappa_3^2 - \kappa_3}$$

$$\kappa_3 > \max\left\{ \frac{1}{2}\left[1 + \left(1 + \left(6 + \frac{2}{\underline{\rho}(L)}\right)^2\right)^{\frac{1}{2}}\right], 4L_f^2 \right\}.$$

Proof: The proof is given in Appendix B. ■

Next, we consider the case when the following additional assumption is satisfied.

Assumption 3: The global cost function $f(x)$ satisfies the P–Ł condition with constant $\nu > 0$, i.e.,

$$\frac{1}{2}\|\nabla f(x)\|^2 \geq \nu(f(x) - f^\star) \quad \forall x \in \mathbb{R}^p. \quad (7)$$

Remark 2: Note that in [7], [8], and [10], the global cost function is required to be strongly convex. In [13], the global cost function needs to satisfy the restricted secant inequality condition, and the global optimal set needs to be a singleton. In this work, we assume that the global cost function satisfies the P–Ł condition, and the global minimizer is not necessarily unique. Note that the P–Ł condition does not imply the convexity of the global cost function; however, it implies invexity [30], i.e., all stationary points are global optimal points. Therefore, the global minimizer may be nonunique. It has been shown that the P–Ł condition is satisfied in some useful applications. For example, Li and Li [31] demonstrated that the loss functions in some applications in deep learning satisfy the P–Ł condition in the region near a local minimum, and Fazel et al. [32] demonstrated that the linear–quadratic regulator problem in reinforcement learning is nonconvex and satisfies the P–Ł condition.

Before presenting the convergence result when the additional Assumption 3 is satisfied, we need the following lemma.

Lemma 1 (see [33, Th. 2]): Suppose that the function f satisfies the P–Ł condition. Let $\mathcal{P}_{\mathbb{X}^\star}(x)$ be the projection x onto the set $\mathbb{X}^\star$, i.e., $\mathcal{P}_{\mathbb{X}^\star}(x) = \arg\min_{y\in\mathbb{X}^\star} \|x - y\|^2$. Suppose that $\mathcal{P}_{\mathbb{X}^\star}(x)\ \forall x \in \mathbb{R}^p$ is well defined. Then

$$f(x) - f^\star \geq 2\nu\|\mathcal{P}_{\mathbb{X}^\star}(x) - x\|^2 \quad \forall x \in \mathbb{R}^p. \quad (8)$$

From [29, Th. 2.2.11], we know that $\mathcal{P}_{\mathbb{X}^\star}(x)$ is well defined if $\mathbb{X}^\star$ is closed and convex.

We are now ready to present the convergence result.

Theorem 2: Let Assumptions 1–3 hold. If each agent $i \in \mathcal{V}$ runs the distributed event-triggered algorithm (4) with the same α and β as chosen in Theorem 1 and determines its triggering time sequence by (5), then

$$\frac{1}{n}\sum_{i=1}^{n}\|x_i(t) - \bar{x}(t)\|^2 + f(\bar{x}(t)) - f^\star \leq c_1 e^{-\frac{\epsilon_1}{\epsilon_2}t} \quad (9)$$

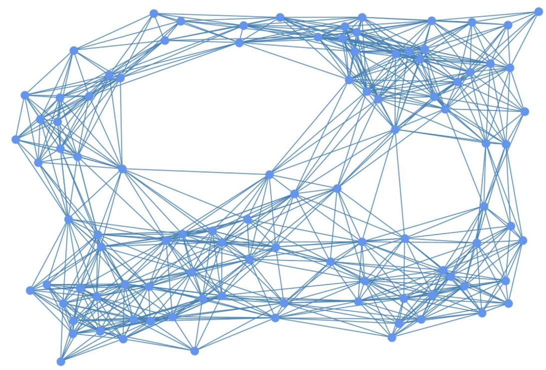

Fig. 1. Random connected network of 100 agents.

where $\epsilon_1 > 0$, $\epsilon_2 > 0$, and $c_1 > 0$ are constants given in Appendix A. Moreover, if the projection operator $\mathcal{P}_{\mathbb{X}^\star}(\cdot)$ is well defined, then

$$\sum_{i=1}^{n}\|x_i(t) - \mathcal{P}_{\mathbb{X}^\star}(\bar{x}(t))\|^2 \leq \left(1 + \frac{1}{2\nu}\right) c_1 e^{-\frac{\epsilon_1}{\epsilon_2}t}. \quad (10)$$

Proof: The proof is given in Appendix C. ■

Remark 3: Compared with [9] and [24], this article establishes exponential convergence under weaker conditions. More specifically, the authors of [9] and [24] proposed distributed event-triggered algorithms and established exponential convergence for the strongly convex case, while our proposed algorithm exponentially converges to a global optimal point if the global cost function satisfies the P–Ł condition, which does not require the cost function to be convex.

IV. Simulation

In this section, we demonstrate the effectiveness of the proposed distributed event-triggered nonconvex optimization algorithm. Consider an undirected network consisting of 100 agents, and the graph is randomly generated, as shown in Fig. 1. The local cost functions associated with agents are given as

$$f_j(x) = 0.2\sqrt{x^4 + 3} + 0.7\cos^2 x$$
$$f_{10+j}(x) = 2\sin x - 0.1(x^2 + 2)^{\frac{1}{3}}$$
$$f_{20+j}(x) = \frac{0.3x^2}{\sqrt{x^2 + 1}}$$
$$f_{30+j}(x) = -0.1\sqrt{x^4 + 3} - \sin x$$
$$f_{40+j}(x) = \frac{-0.2x^2}{\sqrt{x^2 + 1}} + 2\sin^2 x$$
$$f_{50+j}(x) = -0.1\sqrt{x^4 + 3} - \frac{0.1x^2}{\sqrt{x^2 + 1}}$$
$$f_{60+j}(x) = -\sin x - 1$$
$$f_{70+j}(x) = x^2 + 0.3\cos^2 x$$
$$f_{80+j}(x) = 2\sin^2 x + 0.2(x^2 + 2)^{\frac{1}{3}}$$
$$f_{90+j}(x) = -0.1(x^2 + 2)^{\frac{1}{3}}$$

where $j = 1, \ldots, 10$. It is easy to check that Assumptions 1 and 2 are satisfied. Moreover, it can be found that the global cost function is $\frac{1}{10}(x^2 + 3\sin^2(x))$, which is nonconvex but satisfies the P–Ł condition, as shown in [33]. Consider the proposed event-triggered

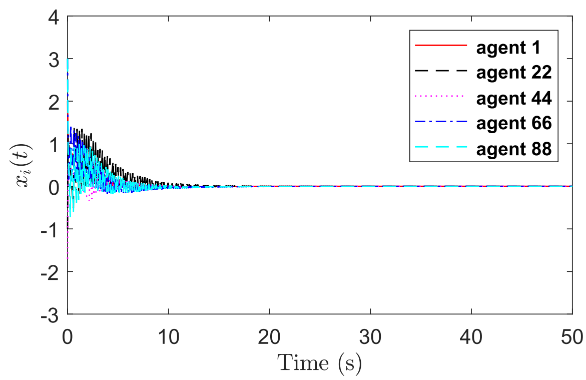

Fig. 2. State evolutions of agents 1, 22, 44, 66, and 88.

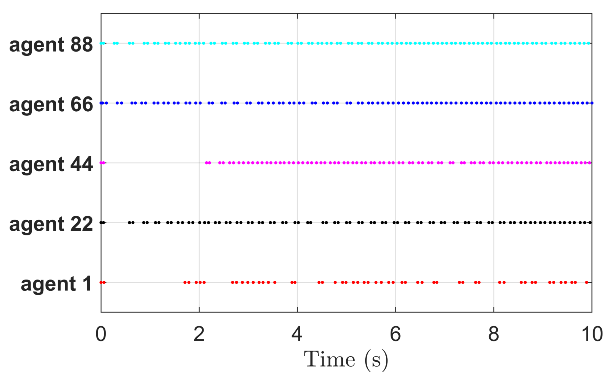

Fig. 3. Triggering times of agents 1, 22, 44, 66, and 88.

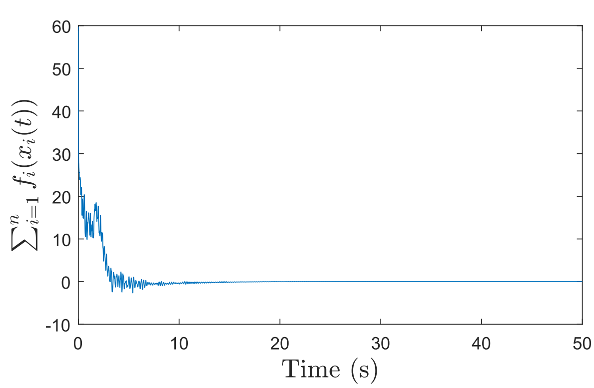

Fig. 4. Evolution of the function $\sum_{i=1}^n f_i(x_i(t))$.

algorithm (4). We randomly choose the initial values $x_i(0)$ and $v_i(0)$ for all $i \in \mathcal{V}$, such that (4b) is satisfied. Also, we randomly choose the design parameters a_i and b_i in the triggering law (5). The state evolutions of $x_i(t)$ for $i \in \{1, 22, 44, 66, 88\}$ are plotted in Fig. 2. It is clear to see that all the states converge to the global minimizer $x^\star = 0$. In order to better demonstrate that the proposed algorithm (4) can effectively avoid continuous communication, we select the time period [0,10], and the triggering times of these agents are given in Fig. 3, which shows that the proposed event-triggered algorithm is free of Zeno behavior. Fig. 4 shows that the function $\sum_{i=1}^n f_i(x_i(t))$ converges to the global optimum $f^\star = 0$.

V. Conclusion

In this article, we studied distributed nonconvex optimization over an undirected connected network. In order to avoid continuous communication, we developed a distributed event-triggered algorithm. For the case where local cost functions are smooth, we showed that the proposed algorithm is free of Zeno behavior and asymptotically converges to a stationary point. Moreover, if the global cost function satisfies the P–Ł condition, we showed that the proposed algorithm exponentially converges to a global optimal point. One future direction is to consider the directed graph and the self-triggered mechanism.

Appendix

A. Proof Preparation

1) In this part, we introduce some useful constants, which are used to simplify the proof.
Denote

$$\epsilon_1 = \min\{\epsilon_5,\, 2\epsilon_6\nu,\, \epsilon_7,\, 1\},\ \epsilon_2 = \max\left\{\frac{1}{2}(c_3+1),\, \frac{1}{b}\right\}$$

$$\epsilon_3 = \min\left\{\frac{1}{4},\, \frac{1}{b}\right\},\ \epsilon_4 = \min\{\epsilon_5,\, \epsilon_6,\, \epsilon_7\}$$

$$\epsilon_5 = \frac{1}{2}(\alpha - 3\beta)\underline{\rho}(L) - \frac{1}{\beta}\left(2L_f^4 + \frac{1}{2}\right) - \frac{3}{2}L_f^2 - \frac{1}{2}$$

$$\epsilon_6 = \frac{1}{2} - \frac{2L_f^2}{\beta},\ \epsilon_7 = \frac{\beta}{2} - \frac{1}{2} - \frac{c_2c_3}{2\beta}$$

$$\epsilon_8 = \frac{1}{2}(\rho(L)(\alpha+\beta)+\beta),\ c_1 = \frac{W(0)}{n\epsilon_3},\ c_2 = \frac{1}{\rho(L)} + \frac{\alpha}{\beta}$$

$$c_3 = \frac{1}{\underline{\rho}(L)} + \frac{\alpha}{\beta},\ a = \max\{a_1, \ldots, a_n\},\ b = \min\{b_1, \ldots, b_n\}$$

where $W(0)$ is given in (23).
From $\alpha \geq 3\beta + \kappa_1$, $\beta > \kappa_3$, $\kappa_1 = [\frac{1}{\kappa_3}(4L_f^4+1) + 4L_f^2 + 1]\frac{1}{\underline{\rho}(L)}$ and $\kappa_3 > \frac{1}{2}\left[1 + \left(1 + \left(6 + \frac{2}{\underline{\rho}(L)}\right)^2\right)^{\frac{1}{2}}\right]$, we have

$$\epsilon_5 > \frac{\kappa_1}{2}\underline{\rho}(L) - \frac{1}{\kappa_3}\left(2L_f^4 + \frac{1}{2}\right) - \frac{3}{2}L_f^2 - \frac{1}{2} > 0. \quad (11)$$

From $\beta > \kappa_3$ and $\kappa_3 > 4L_f^2$, we have

$$\epsilon_6 > 0. \quad (12)$$

From $\alpha \leq \kappa_2\beta$, $\beta > \kappa_3$, we have

$$\epsilon_7 > \frac{\kappa_3}{2} - \frac{1}{2} - \frac{1}{2\kappa_3}\left(\frac{1}{\underline{\rho}(L)} + \kappa_2\right)^2 = 0. \quad (13)$$

From (11)–(13), we know that $\epsilon_1 > 0$ and $\epsilon_4 > 0$.

2) In this part, we introduce some useful properties of the Laplacian matrix L.
Note that L and $K_n = \mathbf{I}_n - \frac{1}{n}\mathbf{1}_n\mathbf{1}_n^T$ are positive semidefinite, $\text{null}(L) = \text{null}(K_n) = \{\mathbf{1}_n\}$, $L \leq \rho(L)\mathbf{I}_n$. Moreover, $K_nL = LK_n = L$, and

$$0 < \underline{\rho}(L)K_n \leq L \leq \rho(L)K_n. \quad (14)$$

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

$R\Lambda_1^{-1}R^T L = LR\Lambda_1^{-1}R^T = K_n$, and

$$\frac{1}{\rho(L)}K_n \leq R\Lambda_1^{-1}R^T \leq \frac{1}{\underline{\rho}(L)}K_n \tag{15}$$

where $\Lambda_1 = \mathrm{diag}(\lambda_2, \ldots, \lambda_n)$ with $0 < \lambda_2 \leq \cdots \leq \lambda_n$ are the eigenvalues of the Laplacian matrix L.

3) In this part, we introduce some useful notations and results. Denote $\boldsymbol{L} = L \otimes \mathbf{I}_p$, $\boldsymbol{K} = K_n \otimes \mathbf{I}_p$, $\boldsymbol{H} = \frac{1}{n}(\mathbf{1}_n\mathbf{1}_n^T \otimes \mathbf{I}_p)$, $Q_n = R\Lambda_1^{-1}R^T$, $\boldsymbol{Q} = Q_n \otimes \mathbf{I}_p$, $\boldsymbol{x}(t) = [x_1^T(t), \ldots, x_n^T(t)]^T$, $\boldsymbol{v}(t) = [v_1^T(t), \ldots, v_n^T(t)]^T$, $\bar{\boldsymbol{x}}(t) = \mathbf{1}_n \otimes \bar{x}(t)$, $\hat{x}_j(t) = x_j(t^j_{k_j(t)})$, $\hat{\boldsymbol{x}}(t) = [\hat{x}_1^T(t), \ldots, \hat{x}_n^T(t)]^T$, $e_j^x(t) = \hat{x}_j(t) - x_j(t)$, $\boldsymbol{e}^x(t) = [(e_1^x(t))^T, \ldots, (e_n^x(t))^T]^T$, $\tilde{f}(\boldsymbol{x}) = \sum_{i=1}^n f_i(x_i)$, $\boldsymbol{g}(t) = \nabla\tilde{f}(\boldsymbol{x}(t))$, $\bar{\boldsymbol{g}}(t) = \boldsymbol{H}\boldsymbol{g}(t)$, $\boldsymbol{g}^0(t) = \nabla\tilde{f}(\bar{\boldsymbol{x}}(t))$, and $\bar{\boldsymbol{g}}^0(t) = \boldsymbol{H}\boldsymbol{g}^0(t) = \mathbf{1}_n \otimes \nabla f(\bar{x}(t))$. Then, the algorithm (4) can be rewritten in a compact form

$$\dot{\boldsymbol{x}}(t) = -\alpha\boldsymbol{L}\hat{\boldsymbol{x}}(t) - \beta\boldsymbol{v}(t) - \nabla\tilde{f}(\boldsymbol{x}(t)) \tag{16a}$$

$$\dot{\boldsymbol{v}}(t) = \beta\boldsymbol{L}\hat{\boldsymbol{x}}(t),\ \forall\boldsymbol{x}(0) \in \mathbb{R}^{np},\ (\mathbf{1}_{np})^T\boldsymbol{v}(0) = \mathbf{0}_{np}. \tag{16b}$$

Noting that $\rho(\boldsymbol{H}) = 1$ and $\nabla\tilde{f}$ is Lipschitz continuous with constant $L_f > 0$ as assumed in Assumption 2, we have

$$\begin{aligned}\|\bar{\boldsymbol{g}}^0(t) - \bar{\boldsymbol{g}}(t)\|^2 &= \|\boldsymbol{H}(\boldsymbol{g}^0(t) - \boldsymbol{g}(t))\|^2\\ &\leq \|\rho(\boldsymbol{H})(\boldsymbol{g}^0(t) - \boldsymbol{g}(t))\|^2 = \|\boldsymbol{g}^0(t) - \boldsymbol{g}(t)\|^2\\ &\leq L_f^2\|\bar{\boldsymbol{x}}(t) - \boldsymbol{x}(t)\|^2 = L_f^2\mathcal{Q}_{\boldsymbol{K}}(\boldsymbol{x}(t)).\end{aligned} \tag{17}$$

Denote $\bar{v}(t) = \frac{1}{n}(\mathbf{1}_n^T \otimes \mathbf{I}_p)\boldsymbol{v}(t)$. From (16b) and $\sum_{i=1}^n L_{ij} = \sum_{j=1}^n L_{ij} = 0$, it is known that $\dot{\bar{v}}(t) = \mathbf{0}_p$. Then, from $\sum_{j=1}^n v_j(0) = \mathbf{0}_p$, we know that $\bar{v}(t) \equiv \mathbf{0}_p$. Then, from (16a) and $\sum_{i=1}^n L_{ij} = \sum_{j=1}^n L_{ij} = 0$, we have

$$\dot{\bar{x}}(t) = -\frac{1}{n}\sum_{i=1}^n \nabla f_i(x_i(t)). \tag{18}$$

4) In this part, we show that the algorithm parameters α and β are well defined. From $\kappa_3 > \frac{1}{2}\left[1 + \left(1 + \left(6 + \frac{2}{\underline{\rho}(L)}\right)^2\right)^{\frac{1}{2}}\right]$, we have

$$\kappa_2 = -\frac{1}{\underline{\rho}(L)} + \sqrt{\kappa_3^2 - \kappa_3} > 3.$$

Thus, β is well defined. Then, from $\beta > \frac{\kappa_1}{\kappa_2 - 3}$, we have $3\beta + \kappa_1 < \beta\kappa_2$. Thus, α is well defined.

B. Proof of Theorem 1

1) In this part, we show that there is no Zeno behavior by contradiction. Suppose that Zeno behavior exists. Then, there exists an agent $i \in \mathcal{V}$, such that $\lim_{k\to\infty} t_k^i = T_0$, where $T_0 > 0$ is a constant. Note that $x_i(t)$ and $v_i(t)$ are continuous. Therefore, there exist constants $p_1 > 0$ and $p_2 > 0$, such that $\|x_i(t)\| \leq p_1$, $\|v_i(t)\| \leq p_2$ for all $i \in \mathcal{V}$ and for all $t \in [0, T_0]$.
From Assumption 2, we know that $f(x)$ is continuously differentiable. Also noting that $\|x_i(t)\| \leq p_1, \forall i \in \mathcal{V}, \forall t \in [0, T_0]$, there exists a constant $p_3 > 0$ such that $\|\nabla\tilde{f}(\boldsymbol{x})\| \leq p_3, \forall t \in [0, T_0]$.
Let $C_0 = 2\alpha\max_{i\in\mathcal{V}}\{L_{ii}\}p_1 + \beta p_2 + p_3$. It then follows from algorithm (4) that

$$\begin{aligned}\|\dot{x}_i(t)\| &\leq \|\alpha\sum_{j=1}^n L_{ij}x_j(t^j_{k_j(t)})\| + \|\beta v_i(t)\| + \|\nabla f_i(x_i(t))\|\\ &\leq C_0 \quad \forall i \in \mathcal{V}\ \ \forall t \in [0, T_0].\end{aligned} \tag{19}$$

Given the triggering law (5), we know that

$$\|e_i^x(t)\| = \|x_i(t) - x_i(t_k^i)\| \leq a_i e^{-b_i t}. \tag{20}$$

It then follows from (19) that one sufficient condition to ensure (20) is

$$C_0(t - t_k^i) \leq a_0 e^{-b_0 t}, \quad t \in [t_k^i, t_{k+1}^i) \tag{21}$$

where $a_0 = \min\{a_1, \ldots, a_n\}$ and $b_0 = \max\{b_1, \ldots, b_n\}$.

Denote $\epsilon = \frac{a_0 e^{-b_0 T_0}}{2C_0}$. It then follows from the property of limits that there exists a positive integer $N(\epsilon)$ such that

$$t_k^i \in [T_0 - \epsilon, T_0] \quad \forall k \geq N(\epsilon). \tag{22}$$

Suppose that we have determined the $N(\epsilon)$th triggering time of agent i, which is denoted by $t^i_{N(\epsilon)}$. Let $t^i_{N(\epsilon)+1}$ and $\tilde{t}^i_{N(\epsilon)+1}$ denote the next triggering time determined by the triggering law (5) and the inequality (21), respectively. Then

$$\begin{aligned}t^i_{N(\epsilon)+1} \geq \tilde{t}^i_{N(\epsilon)+1} &= t^i_{N(\epsilon)} + \frac{a_0 e^{-b_0\tilde{t}^i_{N(\epsilon)+1}}}{C_0}\\ &\geq t^i_{N(\epsilon)} + \frac{a_0 e^{-b_0 T_0}}{C_0} = t^i_{N(\epsilon)} + 2\epsilon\end{aligned}$$

where the first two inequalities hold due to $\tilde{t}^i_{N(\epsilon)+1} \leq t^i_{N(\epsilon)+1} \leq T_0$. Thus

$$t^i_{N(\epsilon)+1} - t^i_{N(\epsilon)} \geq 2\epsilon$$

which contradicts (22). Hence, the event-triggered algorithm (4) is free of Zeno behavior.

2) In this part, we use the Lyapunov stability analysis to show that every $x_i(t)$ asymptotically converges to a stationary point.

1) In this part, we show the following function is nonincreasing and, thus, can be used as the Lyapunov candidate

$$W(t) = F(t) + \frac{a^2\epsilon_8}{b}\|\boldsymbol{z}(t)\|^2 \tag{23}$$

where $F(t) = \sum_{i=1}^4 F_i(t)$, $F_1(t) = \frac{1}{2}\mathcal{Q}_{\boldsymbol{K}}(\boldsymbol{x}(t))$, $F_2(t) = \frac{1}{2}\mathcal{Q}_{\boldsymbol{Q}+\frac{\alpha}{\beta}\boldsymbol{K}}\left(\boldsymbol{v}(t) + \frac{1}{\beta}\boldsymbol{g}^0(t)\right)$, $F_3(t) = \langle\boldsymbol{x}(t), \boldsymbol{v}(t) + \frac{1}{\beta}\boldsymbol{g}^0(t)\rangle_{\boldsymbol{K}}$, $F_4(t) = n(f(\bar{x}(t)) - f^\star)$, and $\boldsymbol{z}(t) = [e^{-bt}, \ldots, e^{-bt}]^T \in \mathbb{R}^n$.
For $F_1(t)$, we have that

$$\begin{aligned}\dot{F}_1(t) &= \boldsymbol{x}^T(t)\boldsymbol{K}(-\alpha\boldsymbol{L}\hat{\boldsymbol{x}}(t) - \beta\boldsymbol{v}(t) - \nabla\tilde{f}(\boldsymbol{x}(t)))\\ &= -\alpha\mathcal{Q}_{\boldsymbol{L}}(\boldsymbol{x}(t)) - \beta\Big\langle\boldsymbol{x}(t), \boldsymbol{v}(t) + \frac{1}{\beta}\boldsymbol{g}(t)\\ &\quad - \frac{1}{\beta}\boldsymbol{g}^0(t) + \frac{1}{\beta}\boldsymbol{g}^0(t)\Big\rangle_{\boldsymbol{K}} - \alpha\langle\boldsymbol{x}(t), \boldsymbol{e}^x(t)\rangle_{\boldsymbol{L}}\\ &= -\alpha\mathcal{Q}_{\boldsymbol{L}}(\boldsymbol{x}(t)) - \beta\Big\langle\boldsymbol{x}(t), \boldsymbol{v}(t) + \frac{1}{\beta}\boldsymbol{g}^0(t)\Big\rangle_{\boldsymbol{K}}\\ &\quad + \langle\boldsymbol{x}(t), \boldsymbol{g}^0(t) - \boldsymbol{g}(t)\rangle_{\boldsymbol{K}} - \alpha\langle\boldsymbol{x}(t), \boldsymbol{e}^x(t)\rangle_{\boldsymbol{L}}\\ &\leq -\alpha\mathcal{Q}_{\boldsymbol{L}}(\boldsymbol{x}(t)) - \beta\Big\langle\boldsymbol{x}(t), \boldsymbol{v}(t) + \frac{1}{\beta}\boldsymbol{g}^0(t)\Big\rangle_{\boldsymbol{K}}\\ &\quad + \frac{1}{2}\mathcal{Q}_{\boldsymbol{K}}(\boldsymbol{x}(t)) + \frac{L_f^2}{2}\mathcal{Q}_{\boldsymbol{K}}(\boldsymbol{x}(t)) + \frac{\alpha}{2}\mathcal{Q}_{\boldsymbol{L}}(\boldsymbol{x}(t))\\ &\quad + \frac{\alpha}{2}\mathcal{Q}_{\boldsymbol{L}}(\boldsymbol{e}^x(t))\end{aligned} \tag{24}$$

where the first equality holds due to (16a); the last inequality holds due to the Cauchy–Schwarz inequality and (2).

For $F_2(t)$, we have

$$\begin{aligned}
\dot{F}_2(t) =& \left(\boldsymbol{v}(t) + \frac{1}{\beta}\boldsymbol{g}^0(t)\right)^T \left(\boldsymbol{Q} + \frac{\alpha}{\beta}\boldsymbol{K}\right) \\
&\times \left(\beta \boldsymbol{L}\hat{\boldsymbol{x}}(t) - \frac{1}{\beta}\nabla^2 \tilde{f}(\bar{\boldsymbol{x}}(t))(\bar{\boldsymbol{g}}(t) - \bar{\boldsymbol{g}}^0(t) + \bar{\boldsymbol{g}}^0(t))\right) \\
=& \left\langle \boldsymbol{v}(t) + \frac{1}{\beta}\boldsymbol{g}^0(t), \beta \boldsymbol{L}\boldsymbol{x}(t) \right\rangle_{\boldsymbol{Q}+\frac{\alpha}{\beta}\boldsymbol{K}} \\
&- \left\langle \frac{1}{\beta}\left(\boldsymbol{v}(t) + \frac{1}{\beta}\boldsymbol{g}^0(t)\right), \nabla^2 \tilde{f}(\bar{\boldsymbol{x}}(t)) \right. \\
&\times \left. (\bar{\boldsymbol{g}}(t) - \bar{\boldsymbol{g}}^0(t)) \right\rangle_{\boldsymbol{Q}+\frac{\alpha}{\beta}\boldsymbol{K}} - \left\langle \frac{1}{\beta}\left(\boldsymbol{v}(t) + \frac{1}{\beta}\boldsymbol{g}^0(t)\right), \right. \\
&\left. \nabla^2 \tilde{f}(\bar{\boldsymbol{x}}(t))\bar{\boldsymbol{g}}^0(t) \right\rangle_{\boldsymbol{Q}+\frac{\alpha}{\beta}\boldsymbol{K}} \\
&+ \langle \boldsymbol{v}(t) + \frac{1}{\beta}\boldsymbol{g}^0(t), \beta \boldsymbol{L}\boldsymbol{e}^x(t) \rangle_{\boldsymbol{Q}+\frac{\alpha}{\beta}\boldsymbol{K}} \\
\leq& \left\langle \boldsymbol{v}(t) + \frac{1}{\beta}\boldsymbol{g}^0(t), \beta \boldsymbol{L}\boldsymbol{x}(t) \right\rangle_{\boldsymbol{Q}+\frac{\alpha}{\beta}\boldsymbol{K}} \\
&+ \frac{c_2}{2\beta}\mathcal{Q}_{\boldsymbol{Q}+\frac{\alpha}{\beta}\boldsymbol{K}}\left(\boldsymbol{v}(t) + \frac{1}{\beta}\boldsymbol{g}^0(t)\right) \\
&+ \frac{1}{\beta c_2}\mathcal{Q}_{\boldsymbol{Q}+\frac{\alpha}{\beta}\boldsymbol{K}}(\nabla^2 \tilde{f}(\bar{\boldsymbol{x}}(t))(\bar{\boldsymbol{g}}(t) - \bar{\boldsymbol{g}}^0(t))) \\
&+ \frac{1}{\beta c_2}\mathcal{Q}_{\boldsymbol{Q}+\frac{\alpha}{\beta}\boldsymbol{K}}(\nabla^2 \tilde{f}(\bar{\boldsymbol{x}}(t))(\bar{\boldsymbol{g}}^0(t))) \\
&+ \left\langle \boldsymbol{v}(t) + \frac{1}{\beta}\boldsymbol{g}^0(t), \alpha \boldsymbol{L}\boldsymbol{e}^x(t) \right\rangle_{\boldsymbol{K}} + \frac{\beta}{2}\mathcal{Q}_{\boldsymbol{K}}(\boldsymbol{e}^x(t)) \\
&+ \frac{\beta}{2}\mathcal{Q}_{\boldsymbol{K}}\left(\boldsymbol{v}(t) + \frac{1}{\beta}\boldsymbol{g}^0(t)\right) \\
\leq& \left\langle \boldsymbol{v}(t) + \frac{1}{\beta}\boldsymbol{g}^0(t), \beta \boldsymbol{L}\boldsymbol{x}(t) \right\rangle_{\boldsymbol{Q}+\frac{\alpha}{\beta}\boldsymbol{K}} \\
&+ \frac{c_2}{2\beta}\mathcal{Q}_{\boldsymbol{Q}+\frac{\alpha}{\beta}\boldsymbol{K}}\left(\boldsymbol{v}(t) + \frac{1}{\beta}\boldsymbol{g}^0(t)\right) + \frac{L_f^2}{\beta}\|\bar{\boldsymbol{g}}(t) - \bar{\boldsymbol{g}}^0(t)\|^2 \\
&+ \frac{L_f^2}{\beta}\|\bar{\boldsymbol{g}}^0(t)\|^2 + \left\langle \boldsymbol{v}(t) + \frac{1}{\beta}\boldsymbol{g}^0(t), \alpha \boldsymbol{L}\boldsymbol{e}^x(t) \right\rangle_{\boldsymbol{K}} \\
&+ \frac{\beta}{2}\mathcal{Q}_{\boldsymbol{K}}(\boldsymbol{e}^x(t)) + \frac{\beta}{2}\mathcal{Q}_{\boldsymbol{K}}\left(\boldsymbol{v}(t) + \frac{1}{\beta}\boldsymbol{g}^0(t)\right) \\
\leq& \left\langle \boldsymbol{v}(t) + \frac{1}{\beta}\boldsymbol{g}^0(t), \beta \boldsymbol{L}\boldsymbol{x}(t) \right\rangle_{\boldsymbol{Q}+\frac{\alpha}{\beta}\boldsymbol{K}} \\
&+ \frac{c_2}{2\beta}\mathcal{Q}_{\boldsymbol{Q}+\frac{\alpha}{\beta}\boldsymbol{K}}\left(\boldsymbol{v}(t) + \frac{1}{\beta}\boldsymbol{g}^0(t)\right) + \frac{L_f^4}{\beta}\mathcal{Q}_{\boldsymbol{K}}(\boldsymbol{x}(t)) \\
&+ \frac{L_f^2}{\beta}\|\bar{\boldsymbol{g}}^0(t)\|^2 + \left\langle \boldsymbol{v}(t) + \frac{1}{\beta}\boldsymbol{g}^0(t), \alpha \boldsymbol{L}\boldsymbol{e}^x(t) \right\rangle_{\boldsymbol{K}} \\
&+ \frac{\beta}{2}\mathcal{Q}_{\boldsymbol{K}}(\boldsymbol{e}^x(t)) + \frac{\beta}{2}\mathcal{Q}_{\boldsymbol{K}}\left(\boldsymbol{v}(t) + \frac{1}{\beta}\boldsymbol{g}^0(t)\right) \qquad (25)
\end{aligned}$$

where the first equality holds due to (16b) and (18); the first inequality holds due to the Cauchy–Schwarz inequality; the second inequality holds due to (3) and $\rho\left(\boldsymbol{Q} + \frac{\alpha}{\beta}\boldsymbol{K}\right) \leq \rho(\boldsymbol{Q}) + \rho\left(\frac{\alpha}{\beta}\boldsymbol{K}\right)$, $\rho(\boldsymbol{K}) = 1$; the last inequality holds due to the Cauchy–Schwarz inequality and (17).

For $F_3(t)$, we have

$$\begin{aligned}
\dot{F}_3(t) =& \left(\boldsymbol{v}(t) + \frac{1}{\beta}\boldsymbol{g}^0(t)\right)^T \boldsymbol{K} \\
&\times (-\alpha \boldsymbol{L}\hat{\boldsymbol{x}}(t) - \beta \boldsymbol{v}(t) - \nabla \tilde{f}(\boldsymbol{x}(t)) + \boldsymbol{g}^0(t) - \boldsymbol{g}^0(t)) \\
&+ \boldsymbol{x}^T(t)\boldsymbol{K}\left(\beta \boldsymbol{L}\hat{\boldsymbol{x}}(t) - \frac{1}{\beta}\nabla^2 \tilde{f}(\bar{\boldsymbol{x}}(t))\bar{\boldsymbol{g}}(t)\right) \\
=& -\alpha\left\langle \boldsymbol{v}(t) + \frac{1}{\beta}\boldsymbol{g}^0(t), \boldsymbol{x}(t) \right\rangle_{\boldsymbol{L}} - \beta\mathcal{Q}_{\boldsymbol{K}}\left(\boldsymbol{v}(t) + \frac{1}{\beta}\boldsymbol{g}^0(t)\right) \\
&+ \left\langle \boldsymbol{v}(t) + \frac{1}{\beta}\boldsymbol{g}^0(t), \boldsymbol{g}^0(t) - \boldsymbol{g}(t) \right\rangle_{\boldsymbol{K}} + \beta\mathcal{Q}_{\boldsymbol{L}}(\boldsymbol{x}(t)) \\
&- \frac{1}{\beta}\langle \boldsymbol{x}(t), \nabla^2 \tilde{f}(\bar{\boldsymbol{x}}(t))(\bar{\boldsymbol{g}}(t) - \bar{\boldsymbol{g}}^0(t) + \bar{\boldsymbol{g}}^0(t))\rangle_{\boldsymbol{K}} \\
&- \left\langle \boldsymbol{v}(t) + \frac{1}{\beta}\boldsymbol{g}^0(t), \alpha \boldsymbol{L}\boldsymbol{e}^x(t) \right\rangle_{\boldsymbol{K}} + \langle \boldsymbol{x}(t), \beta \boldsymbol{L}\boldsymbol{e}^x(t)\rangle_{\boldsymbol{K}} \\
\leq& -\alpha\left\langle \boldsymbol{v}(t) + \frac{1}{\beta}\boldsymbol{g}^0(t), \boldsymbol{x}(t) \right\rangle_{\boldsymbol{L}} - \beta\mathcal{Q}_{\boldsymbol{K}}\left(\boldsymbol{v}(t) + \frac{1}{\beta}\boldsymbol{g}^0(t)\right) \\
&+ \frac{1}{2}\mathcal{Q}_{\boldsymbol{K}}\left(\boldsymbol{v}(t) + \frac{1}{\beta}\boldsymbol{g}^0(t)\right) + \frac{L_f^2}{2}\mathcal{Q}_{\boldsymbol{K}}(\boldsymbol{x}(t)) + \beta\mathcal{Q}_{\boldsymbol{L}}(\boldsymbol{x}(t)) \\
&+ \frac{1}{2\beta}\mathcal{Q}_{\boldsymbol{K}}(\boldsymbol{x}(t)) + \frac{L_f^4}{\beta}\mathcal{Q}_{\boldsymbol{K}}(\boldsymbol{x}(t)) + \frac{L_f^2}{\beta}\mathcal{Q}_{\boldsymbol{K}}(\bar{\boldsymbol{g}}^0(t)) \\
&- \left\langle \boldsymbol{v}(t) + \frac{1}{\beta}\boldsymbol{g}^0(t), \alpha \boldsymbol{L}\boldsymbol{e}^x(t) \right\rangle_{\boldsymbol{K}} + \frac{\beta}{2}\mathcal{Q}_{\boldsymbol{L}}(\boldsymbol{x}(t)) \\
&+ \frac{\beta}{2}\mathcal{Q}_{\boldsymbol{L}}(\boldsymbol{e}^x(t)) \qquad (26)
\end{aligned}$$

where the first equality holds due to algorithm (16); the inequality holds due to (2), the Cauchy–Schwarz inequality, $\rho\left(\boldsymbol{Q} + \frac{\alpha}{\beta}\boldsymbol{K}\right) \leq \rho(\boldsymbol{Q}) + \rho\left(\frac{\alpha}{\beta}\boldsymbol{K}\right)$, $\rho(\boldsymbol{K}) = 1$, and (17).

For $F_4(t)$, we have

$$\begin{aligned}
\dot{F}_4(t) =& n(\nabla f(\bar{x}(t)))^T\left(-\frac{1}{n}\sum_{i=1}^n \nabla f_i(x_i(t))\right) \\
=& -(\nabla f(\bar{x}(t)))^T \sum_{i=1}^n (\nabla f_i(x_i(t)) - \nabla f_i(\bar{x}(t)) \\
&+ \nabla f_i(\bar{x}(t))) \\
=& -\frac{1}{2n}\left(\sum_{i=1}^n \nabla f_i(\bar{x}(t))\right)^T \left(\sum_{i=1}^n \nabla f_i(\bar{x}(t))\right) \\
&+ (\nabla f(\bar{x}(t)))^T \sum_{i=1}^n (\nabla f_i(\bar{x}(t)) - \nabla f_i(x_i(t))) \\
&- \frac{n}{2}\|\nabla f(\bar{x}(t))\|^2 \\
\leq& -\frac{1}{2}\|\bar{\boldsymbol{g}}^0(t)\|^2 + \|\nabla f(\bar{x}(t))\| \sum_{i=1}^n \|\nabla f_i(\bar{x}(t)) \\
&- \nabla f_i(x_i(t))\| - \frac{n}{2}\|\nabla f(\bar{x}(t))\|^2
\end{aligned}$$

$$\le -\frac{1}{2}\|\bar{\boldsymbol{g}}^0(t)\|^2 + L_f\|\nabla f(\bar{x}(t))\|\sum_{i=1}^n\|\bar{x}(t)-x_i(t)\| - \frac{n}{2}\|\nabla f(\bar{x}(t))\|^2$$

$$\le -\frac{1}{2}\|\bar{\boldsymbol{g}}^0(t)\|^2 + \frac{n}{2}\|\nabla f(\bar{x}(t))\|^2 + \frac{L_f^2}{2n}\left(\sum_{i=1}^n\|\bar{x}(t)-x_i(t)\|\right)^2 - \frac{n}{2}\|\nabla f(\bar{x}(t))\|^2$$

$$= -\frac{1}{2}\|\bar{\boldsymbol{g}}^0(t)\|^2 + \frac{L_f^2}{2n}\left(\sum_{i=1}^n\|\bar{x}(t)-x_i(t)\|\right)^2$$

$$\le -\frac{1}{2}\|\bar{\boldsymbol{g}}^0(t)\|^2 + \frac{L_f^2}{2}\sum_{i=1}^n\|\bar{x}(t)-x_i(t)\|^2$$

$$= -\frac{1}{2}\|\bar{\boldsymbol{g}}^0(t)\|^2 + \frac{L_f^2}{2}\mathcal{Q}_{\boldsymbol{K}}(\boldsymbol{x}(t)) \tag{27}$$

where the first equality holds due to (18); the first, second, and last inequalities hold due to the Cauchy–Schwarz inequality; the second inequality holds due to (3); and the last equality holds due to $\sum_{i=1}^n\|\bar{x}(t)-x_i(t)\|^2 = \|\bar{\boldsymbol{x}}(t)-\boldsymbol{x}(t)\|^2 = \mathcal{Q}_{\boldsymbol{K}}(\boldsymbol{x}(t))$.

Then, based on (24)–(27), we have

$$\dot{F}(t) \le -\mathcal{Q}_{\frac{1}{2}(\alpha-3\beta)\boldsymbol{L}-\left[\frac{1}{\beta}\left(2L_f^4+\frac{1}{2}\right)+\frac{3}{2}L_f^2+\frac{1}{2}\right]\boldsymbol{K}}(\boldsymbol{x}(t)) - \mathcal{Q}_{\left(\frac{\beta}{2}-\frac{1}{2}-\frac{\alpha c_2}{2\beta^2}\right)\boldsymbol{K}-\frac{c_2}{2\beta}\boldsymbol{Q}}\left(\boldsymbol{v}(t)+\frac{1}{\beta}\boldsymbol{g}^0(t)\right) - \left(\frac{1}{2}-\frac{2L_f^2}{\beta}\right)\|\bar{\boldsymbol{g}}^0(t)\|^2 + \mathcal{Q}(\boldsymbol{e}^x(t))_{\frac{1}{2}[(\alpha+\beta)\boldsymbol{L}+\beta\boldsymbol{K}]}$$

$$\le -\epsilon_5\mathcal{Q}_{\boldsymbol{K}}(\boldsymbol{x}(t)) - \epsilon_6\|\bar{\boldsymbol{g}}^0(t)\|^2 - \epsilon_7\mathcal{Q}_{\boldsymbol{K}}\left(\boldsymbol{v}(t)+\frac{1}{\beta}\boldsymbol{g}^0(t)\right) + \epsilon_8\|\boldsymbol{e}^x(t)\|^2 \tag{28}$$

$$\le -\epsilon_4(\mathcal{Q}_{\boldsymbol{K}}(\boldsymbol{x}(t)) + \mathcal{Q}_{\boldsymbol{K}}\left(\boldsymbol{v}(t)+\frac{1}{\beta}\boldsymbol{g}^0(t)\right) + \|\bar{\boldsymbol{g}}^0(t)\|^2) + \epsilon_8\|\boldsymbol{e}^x(t)\|^2 \tag{29}$$

where the second inequality holds due to (14) and (15).

Then, from (20), it can be obtained that

$$\|\boldsymbol{e}^x(t)\|^2 \le na^2e^{-2bt}. \tag{30}$$

It then follows from (29) and (30) that

$$\dot{W}(t) = \dot{F}(t) - 2a^2\epsilon_8\|\boldsymbol{z}(t)\|^2 \le -\epsilon_4\left(\mathcal{Q}_{\boldsymbol{K}}(\boldsymbol{x}(t)) + \mathcal{Q}_{\boldsymbol{K}}\left(\boldsymbol{v}(t)+\frac{1}{\beta}\boldsymbol{g}^0(t)\right) + \|\bar{\boldsymbol{g}}^0(t)\|^2\right) - a^2\epsilon_8\|\boldsymbol{z}(t)\|^2. \tag{31}$$

Thus, from (31) and $\epsilon_4 > 0$, we know that $W(t)$ is nonincreasing.

2) In this part, we use the Barbălat's lemma [34, Th. 5] to prove (6). From the Cauchy–Schwarz inequality, we have

$$W(t) \ge \epsilon_3(\mathcal{Q}_{\boldsymbol{K}}(\boldsymbol{x}(t)) + \mathcal{Q}_{\boldsymbol{K}}\left(\boldsymbol{v}(t)+\frac{1}{\beta}\boldsymbol{g}^0(t)\right) + n(f(\bar{x}(t))-f^\star) + a^2\epsilon_8\|\boldsymbol{z}(t)\|^2) \ge 0. \tag{32}$$

Then, from the fact that $W(t)$ is nonincreasing and (32), we have

$$0 \le W(t) \le W(0). \tag{33}$$

Similar to the way to get (24), we have

$$\left(\frac{d\|\boldsymbol{x}(t)-\bar{\boldsymbol{x}}(t)\|}{dt}\right)^2 = \left(\frac{(\boldsymbol{x}(t)-\bar{\boldsymbol{x}}(t))^T(\dot{\boldsymbol{x}}(t)-\dot{\bar{\boldsymbol{x}}}(t))}{\|\boldsymbol{x}(t)-\bar{\boldsymbol{x}}(t)\|}\right)^2$$

$$\le \|\dot{\boldsymbol{x}}(t)\|_{\boldsymbol{K}}^2 = \|-\alpha\boldsymbol{L}\hat{\boldsymbol{x}}(t) - \beta\boldsymbol{v}(t) - \nabla\tilde{f}(\boldsymbol{x}(t))\|_{\boldsymbol{K}}^2$$

$$= \alpha^2\mathcal{Q}_{\boldsymbol{L}^2}(\boldsymbol{x}(t)) + 2\alpha^2\langle\boldsymbol{x}(t),\boldsymbol{e}^x(t)\rangle_{\boldsymbol{L}^2} + \alpha^2\mathcal{Q}_{\boldsymbol{L}^2}(\boldsymbol{e}^x(t))$$

$$+ 2\beta\left\langle\boldsymbol{v}(t)+\frac{1}{\beta}\boldsymbol{g}^0(t)+\frac{1}{\beta}\boldsymbol{g}(t)-\frac{1}{\beta}\boldsymbol{g}^0(t), \alpha\boldsymbol{L}\boldsymbol{x}(t)\right\rangle_{\boldsymbol{K}}$$

$$+ 2\beta\left\langle\boldsymbol{v}(t)+\frac{1}{\beta}\boldsymbol{g}^0(t)+\frac{1}{\beta}\boldsymbol{g}(t)-\frac{1}{\beta}\boldsymbol{g}^0(t), \alpha\boldsymbol{L}\boldsymbol{e}^x(t)\right\rangle_{\boldsymbol{K}}$$

$$+ \beta^2\mathcal{Q}_{\boldsymbol{K}}\left(\boldsymbol{v}(t)+\frac{1}{\beta}\boldsymbol{g}^0(t)+\frac{1}{\beta}\boldsymbol{g}(t)-\frac{1}{\beta}\boldsymbol{g}^0(t)\right)$$

$$\le c_4\left(\mathcal{Q}_{\boldsymbol{K}}(\boldsymbol{x}(t)) + \mathcal{Q}_{\boldsymbol{K}}\left(\boldsymbol{v}(t)+\frac{1}{\beta}\boldsymbol{g}^0(t)\right) + \|\boldsymbol{e}^x(t)\|^2\right) \tag{34}$$

where $c_4 = \max\{4(\alpha^2\rho^2(L)+L_f^2), 4\beta^2\}$; the last inequality holds due to the Cauchy–Schwarz inequality, (14), (17), and $\rho(\boldsymbol{K}) = 1$.

Similarly, we have

$$\left(\frac{d\|\bar{\boldsymbol{g}}^0(t)\|}{dt}\right)^2 = \left(\frac{(\bar{\boldsymbol{g}}^0(t))^T\dot{\bar{\boldsymbol{g}}}^0(t)}{\|\bar{\boldsymbol{g}}^0(t)\|}\right)^2$$

$$\le \|\dot{\bar{\boldsymbol{g}}}^0(t)\|^2 = n\|\nabla^2 f(\bar{x}(t))\dot{\bar{x}}(t)\|^2$$

$$\le nL_f^2\|\dot{\bar{x}}(t)\|^2 = L_f^2\|\bar{\boldsymbol{g}}(t)\|^2 = L_f^2\|\bar{\boldsymbol{g}}(t)-\bar{\boldsymbol{g}}^0(t)+\bar{\boldsymbol{g}}^0(t)\|^2$$

$$\le 2L_f^2\|\bar{\boldsymbol{g}}(t)-\bar{\boldsymbol{g}}^0(t)\|^2 + 2L_f^2\|\bar{\boldsymbol{g}}^0(t)\|^2$$

$$\le 2L_f^4\mathcal{Q}_{\boldsymbol{K}}(\boldsymbol{x}(t)) + 2L_f^2\|\bar{\boldsymbol{g}}^0(t)\|^2 \tag{35}$$

where the second, third, and the last inequalities hold due to (3), (17), and (18), respectively.

From (31) and (33), we know that $\mathcal{Q}_{\boldsymbol{K}}(\boldsymbol{x}(t)) + \mathcal{Q}_{\boldsymbol{K}}\left(\boldsymbol{v}(t)+\frac{1}{\beta}\boldsymbol{g}^0(t)\right) + \|\bar{\boldsymbol{g}}^0(t)\|^2$ is integrable. Then, from (34) and (35), we know that $d\|\boldsymbol{x}(t)-\bar{\boldsymbol{x}}(t)\|/dt$ and $d\|\bar{\boldsymbol{g}}^0(t)\|/dt$ are square integrable. Finally, from the Barbălat's lemma, we have

$$\lim_{t\to\infty}(\|\boldsymbol{x}(t)-\bar{\boldsymbol{x}}(t)\| + \|\bar{\boldsymbol{g}}^0(t)\|) = 0$$

i.e., (6) holds.

C. Proof of Theorem 2

1) From (7) and (28), we have

$$\dot{F}(t) \le -\epsilon_5\mathcal{Q}_{\boldsymbol{K}}(\boldsymbol{x}(t)) - 2\epsilon_6\nu n(f(\bar{x}(t))-f^\star) - \epsilon_7\mathcal{Q}_{\boldsymbol{K}}\left(\boldsymbol{v}(t)+\frac{1}{\beta}\boldsymbol{g}^0(t)\right) + \epsilon_8\|\boldsymbol{e}^x(t)\|^2. \tag{36}$$

From (36), it can be obtained that

$$\dot{W}(t) = \dot{F}(t) - 2a^2\epsilon_8\|\boldsymbol{z}(t)\|^2 \le -\epsilon_1(\mathcal{Q}_{\boldsymbol{K}}(\boldsymbol{x}(t)) + \mathcal{Q}_{\boldsymbol{K}}\left(\boldsymbol{v}(t)+\frac{1}{\beta}\boldsymbol{g}^0(t)\right) + n(f(\bar{x}(t))-f^\star) + a^2\epsilon_8\|\boldsymbol{z}(t)\|^2). \tag{37}$$

From the Cauchy–Schwarz inequality, we have

$$W(t) \leq \epsilon_2(\mathcal{Q}_{\boldsymbol{K}}(\boldsymbol{x}(t)) + \mathcal{Q}_{\boldsymbol{K}}\left(\boldsymbol{v}(t) + \frac{1}{\beta}\boldsymbol{g}^0(t)\right) + n(f(\bar{x}(t)) - f^\star) + a^2\epsilon_8\|\boldsymbol{z}(t)\|^2). \quad (38)$$

Then, from (37) and (38), we have

$$\dot{W}(t) \leq -\frac{\epsilon_1}{\epsilon_2}W(t). \quad (39)$$

Finally, from (32) and (39), we have

$$\frac{1}{n}\sum_{i=1}^{n}\|x_i(t) - \bar{x}(t)\|^2 + f(\bar{x}(t)) - f^\star \leq \frac{W(t)}{n\epsilon_3} \leq c_1 e^{-\frac{\epsilon_1}{\epsilon_2}t}.$$

Thus, (9) is obtained.

2) If the projection operator $\mathcal{P}_{\mathbb{X}^\star}(\cdot)$ is well defined, then from the Cauchy–Schwarz inequality and (8), we know

$$\begin{aligned}
&\|\boldsymbol{x}(t) - \mathbf{1}_n \otimes \mathcal{P}_{\mathbb{X}^\star}(\bar{x}(t))\|^2 \\
&\leq \left(1 + \frac{1}{2\nu}\right)\|\boldsymbol{x}(t) - \bar{\boldsymbol{x}}(t)\|^2 + (1+2\nu)n\|\bar{x}(t) - \mathcal{P}_{\mathbb{X}^\star}(\bar{x}(t))\|^2 \\
&\leq \left(1 + \frac{1}{2\nu}\right)\|\boldsymbol{x}(t) - \bar{\boldsymbol{x}}(t)\|^2 + (1+2\nu)\frac{n}{2\nu}(f(\bar{x}(t)) - f^\star) \\
&= \left(1 + \frac{1}{2\nu}\right)(\mathcal{Q}_{\boldsymbol{K}}(\boldsymbol{x}(t)) + n(f(\bar{x}(t)) - f^\star)). \quad (40)
\end{aligned}$$

Finally, (9) and (40) yield (10).

REFERENCES

[1] J. Tsitsiklis, D. Bertsekas, and M. Athans, "Distributed asynchronous deterministic and stochastic gradient optimization algorithms," *IEEE Trans. Autom. Control*, vol. AC-31, no. 9, pp. 803–812, Sep. 1986.

[2] D. P. Bertsekas and J. N. Tsitsiklis, *Parallel and Distributed Computation: Numerical Methods*. Englewood Cliffs, NJ, USA: Prentice-Hall, 1989.

[3] A. Nedić and J. Liu, "Distributed optimization for control," *Annu. Rev. Control, Robot., Auton. Syst.*, vol. 1, pp. 77–103, 2018.

[4] T. Yang et al., "A survey of distributed optimization," *Annu. Rev. Control*, vol. 47, pp. 278–305, 2019.

[5] J. R. Marden, S. D. Ruben, and L. Y. Pao, "A model-free approach to wind farm control using game theoretic methods," *IEEE Trans. Control Syst. Technol.*, vol. 21, no. 4, pp. 1207–1214, Jul. 2013.

[6] B. Charrow, N. Michael, and V. Kumar, "Cooperative multi-robot estimation and control for radio source localization," *Int. J. Robot. Res.*, vol. 33, no. 4, pp. 569–580, 2014.

[7] J. Wang and N. Elia, "Control approach to distributed optimization," in *Proc. Annu. Allerton Conf. Commun., Control, Comput.*, 2010, pp. 557–561.

[8] B. Gharesifard and J. Cortés, "Distributed continuous-time convex optimization on weight-balanced digraphs," *IEEE Trans. Autom. Control*, vol. 59, no. 3, pp. 781–786, Mar. 2014.

[9] S. S. Kia, J. Cortés, and S. Martínez, "Distributed convex optimization via continuous-time coordination algorithms with discrete-time communication," *Automatica*, vol. 55, pp. 254–264, 2015.

[10] J. Lu and C. Y. Tang, "Zero-gradient-sum algorithms for distributed convex optimization: The continuous-time case," *IEEE Trans. Autom. Control*, vol. 57, no. 9, pp. 2348–2354, Sep. 2012.

[11] D. Varagnolo, F. Zanella, A. Cenedese, G. Pillonetto, and L. Schenato, "Newton-Raphson consensus for distributed convex optimization," *IEEE Trans. Autom. Control*, vol. 61, no. 4, pp. 994–1009, Apr. 2016.

[12] H. Moradian and S. S. Kia, "A distributed continuous-time modified Newton–Raphson algorithm," *Automatica*, vol. 136, 2021, Art. no. 109886.

[13] X. Yi, S. Zhang, T. Yang, T. Chai, and K. Johansson, "Exponential convergence for distributed optimization under the restricted secant inequality condition," in *Proc. IFAC World Congr.*, 2020, pp. 12–17.

[14] L. Shu, L. Wang, and G. Yin, "Exponential convergence of distributed primal–dual convex optimization algorithm without strong convexity," *Automatica*, vol. 105, pp. 298–306, 2019.

[15] 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*, vol. 67, no. 8, pp. 4194–4201, Aug. 2022.

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

[17] C. Zhang and Y. Wang, "Sensor network event localization via nonconvex nonsmooth ADMM and augmented lagrangian methods," *IEEE Trans. Control Netw. Syst.*, vol. 6, no. 4, pp. 1473–1485, Dec. 2019.

[18] K. J. Åström and B. M. Bernhardsson, "Comparison of Riemann and Lebesgue sampling for first order stochastic systems," in *Proc. IEEE Conf. Decis. Control*, 2002, pp. 2011–2016.

[19] P. Tabuada, "Event-triggered real-time scheduling of stabilizing control tasks," *IEEE Trans. Autom. Control*, vol. 52, no. 9, pp. 1680–1685, Sep. 2007.

[20] D. V. Dimarogonas, E. Frazzoli, and K. H. Johansson, "Distributed event-triggered control for multi-agent systems," *IEEE Trans. Autom. Control*, vol. 57, no. 5, pp. 1291–1297, May 2012.

[21] X. Yi, K. Liu, D. V. Dimarogonas, and K. H. Johansson, "Dynamic event-triggered and self-triggered control for multi-agent systems," *IEEE Trans. Autom. Control*, vol. 64, no. 8, pp. 3300–3307, Aug. 2019.

[22] W. Heemels, K. H. Johansson, and P. Tabuada, "An introduction to event-triggered and self-triggered control," in *Proc. IEEE Conf. Decis. Control*, 2012, pp. 3270–3285.

[23] K. H. Johansson, M. Egerstedt, J. Lygeros, and S. Sastry, "On the regularization of Zeno hybrid automata," *Syst. Control Lett.*, vol. 38, no. 3, pp. 141–150, 1999.

[24] W. Chen and W. Ren, "Event-triggered zero-gradient-sum distributed consensus optimization over directed networks," *Automatica*, vol. 65, pp. 90–97, 2016.

[25] W. Du, X. Yi, J. George, K. H. Johansson, and T. Yang, "Distributed optimization with dynamic event-triggered mechanisms," in *Proc. IEEE Conf. Decis. Control*, 2018, pp. 969–974.

[26] S. Liu, L. Xie, and D. E. Quevedo, "Event-triggered quantized communication-based distributed convex optimization," *IEEE Trans. Control Netw. Syst.*, vol. 5, no. 1, pp. 167–178, Mar. 2018.

[27] J. George and P. Gurram, "Distributed stochastic gradient descent with event-triggered communication," in *Proc. AAAI Conf. Artif. Intell.*, vol. 34, no. 5, 2020, pp. 7169–7178.

[28] 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.

[29] Y. Nesterov, *Lectures on Convex Optimization*, 2nd ed. New York, NY, USA: Springer, 2018.

[30] F. Dinuzzo, S. O. Cheng, P. V. Gehler, and G. Pillonetto, "Learning output kernels with block coordinate descent," in *Proc. Int. Conf. Mach. Learn.*, 2011, pp. 49–56.

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

[32] 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.

[33] 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.

[34] B. Farkas and S.-A. Wegner, "Variations on Barbălat's lemma," *Amer. Math. Monthly*, vol. 123, no. 8, pp. 825–830, 2016.