# Event-Triggered Proportional-Integral Algorithms for Distributed Optimization

Wen Du, Xinlei Yi, Shengjun Zhang, Jemin George, and Tao Yang

*Abstract*— **In this paper, we consider the distributed optimization problem, whose objective is to minimize a global cost function formed by a sum of local private cost functions, by using local information exchange. In order to avoid continuous communication among agents and reduce communication overheads, we develop event-triggered distributed optimization algorithms for undirected connected graphs based on the proportional-integral control strategy. We show that the proposed algorithms are free of Zeno behavior, and asymptotically converge to one of global minimizers, if the local cost functions are convex and differentiable. Moreover, we show that the proposed algorithms exponentially converge to the unique global minimizer if in addition, the local cost functions have locally Lipschitz gradients, and the global cost function is restricted strongly convex with respect to the global minimizer. The theoretical results are illustrated by numerical simulations.**

## I. Introduction

For a network of agents, each of which has a local convex objective function, the goal of the distributed optimization problem is to minimize the global cost function formed by a sum of the local cost functions. The distributed optimization problem has a long history which can be traced back to [1], [2]. Such a problem has gained renewed interests in recent years due to its wide applications on power system, machine learning, and sensor network, just to name a few. Various algorithms have been developed and can be divided into two categories depending on whether the algorithm is discrete-time or continuous-time.

Most existing discrete-time algorithms are based on the consensus and distributed (sub)gradient descent (DGD) method [3]–[8]. Although the DGD algorithm can deal with non-smooth convex functions and has been extended in several directions to handle more realistic scenarios, the convergence rate is rather slow due to the diminishing step-sizes. With a fixed step-size, the DGD algorithm converges fast, but only to a neighborhood of an optimal point [9], [10]. Recent studies focused on developing accelerated algorithms with fixed step-sizes by using historical information. In particular, algorithms proposed in [11], [12] are based on the proportional-integral (PI) control strategy, while algorithms proposed in [11], [13]–[19] are based on the combination of the distributed inexact gradient method and the gradient tracking technique via distributed dynamic average [20].

The existing continuous-time algorithms can be classified into two classes depending on whether the algorithm uses the first-order gradient information or the second-order Hessian information. Most gradient-based algorithms proposed in [21]–[23] are based on the PI control strategy, where each agent uses an auxiliary state (integral feedback) to correct the error caused by different local gradients. The algorithms in the second class use the second-order Hessian information, see, e.g., [24]–[26].

In order to avoid continuous communication and reduce communication overheads, the idea of event-triggered communication and control, originally proposed for the single system [27]–[29] and later extended to distributed consensus [30]–[36], has been applied to distributed optimization. A few distributed optimization algorithms with event-triggered mechanisms have been developed [23], [37]–[41]. In particular, the authors of [23] developed an event-triggered algorithm which is free of Zeno behavior [42], i.e., an infinite number of triggered events in a finite period of time, and established its exponential convergence to a neighborhood of the global minimizer if the underlying undirected graph is connected and the local cost functions are strongly convex and differentiable with locally Lipschitz gradients. Motivated by the zero-gradient-sum (ZGS) algorithm proposed in [24], the authors of [38] equipped the ZGS algorithm with a periodical time-triggered communication scheme, while the authors of [39] developed a dynamic event-triggered communication mechanism for the ZGS algorithm. The exponential convergences have been established for undirected connected graphs or weight-balanced strongly connected directed graphs if the local cost functions are strongly convex and twice differentiable with locally Lipschitz Hessians.

*Statement of Contributions:* In this paper, we develop distributed event-triggered algorithms based on the PI control strategy. Under the condition that the undirected graph is connected, and the local cost functions are convex and differentiable, we show that the proposed event-triggered algorithms are free of Zeno behavior and asymptotically converge to one of the global minimizers. Moreover, we establish the exponential convergence if in addition, the local cost functions have locally Lipschitz gradients, and the global cost function is restricted strongly convex with respect to the global minimizer.

Compared to the event-triggered algorithm [23], which only converges to a neighborhood of the global minimizer for undirected graphs if the local cost functions are strongly con-

Wen Du, Shengjun Zhang, and Tao Yang are with the Department of Electrical Engineering, University of North Texas, Denton, TX 76203 USA (e-mail: WenDu@my.unt.edu, ShengjunZhang@my.unt.edu, Tao.Yang@unt.edu).

Xinlei Yi is with the ACCESS Linnaeus Center, School of Electrical Engineering and Computer Science, Royal Institute of Technology, 100 44 Stockholm, Sweden (e-mail: xinleiy@kth.se).

Jemin George is with the U.S. Army Research Laboratory, Adelphi, MD 20783, USA (email: jemin.george.civ@mail.mil).

vex and differentiable with locally Lipschitz gradients, our proposed algorithms exponentially converge to the unique global minimizer exactly under less restricted conditions. Compared with the algorithms proposed in [40], our proposed algorithms are simple in the sense that some special designed gain parameters are needed when performing the algorithms in [40]. Compared to event-triggered ZGS algorithms proposed in [38], [39], which need the second-order Hessian information, our proposed algorithms are first-order gradient based, which are easily to be implemented. Moreover, the ZGS algorithms require the special initialization, while our proposed algorithms allow an arbitrary initialization on the estimates of decision variables.

The rest of this paper is organized as follows. In Section II, we introduce some preliminaries. In Section III, we formulate the distributed optimization problem and motivate our study. The main results are stated in Sections IV. Simulations are given in Section V. Finally, concluding remarks are offered in Section VI.

*Notations:* Given a matrix $A$, $A^\top$ denotes its transpose. A symmetric matrix $A$ is positive (negative) semi-definite if and only if all its eigenvalues are non-negative (non-positive). Given two symmetric matrices $M, N$, $M \leq N$ means that $M-N$ is negative semi-definite. The notation $A \otimes B$ denotes the Kronecker product between matrices $A$ and $B$. $\rho(\cdot)$ stands for the spectral radius for matrices and $\rho_2(\cdot)$ indicates the minimum positive eigenvalue for nonnegative matrices. $I_n$ denotes the identity matrix of dimension $n \times n$. $\mathbf{1}_n$ denotes the column vector of $n$-dimension with each entry being 1. $\|\cdot\|$ denotes either the Euclidean norm for vectors or the induced 2-norm for matrices. Given a vector $[a_1, \ldots, a_N]^\top \in \mathbb{R}^N$, $\mathrm{diag}([a_1, \ldots, a_N])$ is a diagonal matrix with the $i$-th diagonal element being $a_i$. For column vectors $x_1, x_2, \ldots, x_N$, the stacked column vector of $x_1, x_2, \ldots, x_N$ is denoted by $[x_1, x_2, \ldots, x_N]$.

## II. Preliminaries

In this section, we introduce some basic knowledge of graph theory [43]. Consider an undirected graph $\mathcal{G} = (\mathcal{V}, \mathcal{E}, \mathcal{A})$ with $N$ agents, where $\mathcal{V} = \{1, 2, \ldots, N\}$ represents the set of agents, $\mathcal{E} \subset \{(i,j) : i, j, \in \mathcal{V}, i \neq j\}$ represents the set of links, and the weighted adjacency matrix is $\mathcal{A} = [a_{ij}] \in \mathbb{R}^{N \times N}$, where $a_{ij} > 0$ if and only if $(j,i) \in \mathcal{E}$, and $a_{ij} = 0$ otherwise. In this paper, we also assume that there's no self-loops, i.e., $a_{ii} = 0$ for all $i \in \mathcal{V}$. The neighbor set of agent $i$ is defined as $\mathcal{N}_i = \{j \in \mathcal{V} | a_{ij} > 0\}$. A path from node $i_1$ to node $i_k$ is a sequence of nodes $i_1, \ldots, i_k$, such that $(i_j, i_{j+1}) \in \mathcal{E}$ for $j = 1, \ldots, k-1$ in the undirected graph. An undirected graph is said to be connected if there exists a path between any pair of distinct nodes.

For an undirected weighted graph $\mathcal{G}$, the weighted Laplacian matrix $L = [L_{ij}] \in \mathbb{R}^{N \times N}$ is defined as $L_{ii} = \sum_{j=1}^N a_{ij}$ and $L_{ij} = -a_{ij}$ for $i \neq j$. The row sums of the Laplacian matrix are zero. If the undirected weighted graph $\mathcal{G}$ is connected, the Laplacian matrix $L$ has a simple eigenvalue at zero with corresponding right eigenvector $\mathbf{1}$, and all other eigenvalues are greater than zero.

## III. Problem Formulation and motivation

Consider a network of $N$ agents, each of which has a local private convex function $f_i : \mathbb{R}^n \to \mathbb{R}$. All agents collaborate together to find an optimizer $x^*$ that minimizes the global objective $f(x) = \sum_{i=1}^N f_i(x)$, i.e.,

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

The communication among agents is described by an undirected weighted graph $\mathcal{G} = (\mathcal{V}, \mathcal{E}, \mathcal{A})$, where $\mathcal{V} = \{1, 2, \ldots, N\}$ is the agent set, $\mathcal{E} \subseteq \mathcal{V} \times \mathcal{V}$ is the edge set, and $\mathcal{A}$ is the weighted adjacency matrix.

As reviewed in the introduction section, in order to avoid continuous communication among agents, a few distributed event-triggered algorithms have been developed. However, most existing algorithms either require special initializations and/or only converges to a neighborhood of the global minimizer. These motivates our study of this paper. More specifically, we aim to develop event-triggered algorithms with arbitrary initializations and converge to the global minimizer exactly.

We make the following assumptions about local and global cost functions:

**Assumption 1.** *For each $i \in \mathcal{V}$, the local cost function $f_i(x)$ is continuously differentiable and convex. Moreover, the global minimizer to $\sum_{i=1}^N f_i(x)$ is bounded.*

**Assumption 2.** *The global cost function $f(x) = \sum_{i=1}^N f_i(x)$ is restricted strongly convex with respect to its global minimizer $x^*$ with convexity parameter $m_f > 0$, i.e., for all $x \in \mathbb{R}^n$,*

$$\sum_{i=1}^N (\nabla f_i(x) - \nabla f_i(x^*))^\top (x - x^*) \geq m_f \|x - x^*\|^2.$$

**Assumption 3.** *For each $i \in \mathcal{V}$, the local cost function $f_i(x)$ has a locally Lipschitz gradient, i.e., for any compact set $D \subseteq \mathbb{R}^n$, there exists a constant $M_i(D) > 0$, such that $\|\nabla f_i(x) - \nabla f_i(y)\| \leq M_i(D)\|x - y\|$, $\forall x, y \in D$.*

Note that under Assumption 1, the global minimizer $x^*$ to the distributed optimization problem (1) may be nonunique. Moreover, if Assumption 2 also holds, it is easy to show that the global minimizer $x^*$ is unique. This assumption is less restricted compared to the assumption that the local cost functions are strongly convex, commonly assumed in the literature. For detailed discussions, please see [11], [40]. Finally, we note that Assumption 3 is also widely used in the existing literature.

## IV. Distributed event-triggered algorithms

In this section, we propose distributed event-triggered algorithms and establish the convergence results.

Our first proposed event-triggered algorithm is based on the distributed PI algorithm [21], [22] and is described as:

$$\dot{q}_i(t) = \sum_{j=1}^{N} L_{ij} x_j(t^j_{k_j(t)}),\ \forall q_i(0), \tag{2a}$$

$$\begin{aligned}\dot{x}_i(t) = &-\sum_{j=1}^{N} L_{ij} x_j(t^j_{k_j(t)}) - \sum_{j=1}^{N} L_{ij} q_j(t^j_{k_j(t)}) \\ &- \nabla f_i(x_i(t)),\ \forall x_i(0),\ t \in [t^i_k, t^i_{k+1}),\end{aligned} \tag{2b}$$

where the sequence $\{t^j_k\}_{k=1}^{\infty}, \forall j \in \mathcal{V}$ is the triggering times to be determined later and $t^j_{k_j(t)} = \max\{t^j_k : t^j_k \le t\}$. We assume $t^j_1 = 0,\ \forall j \in \mathcal{V}$. For simplicity, let $\hat{x}_j(t) = x_j(t^j_{k_j(t)})$, $\hat{q}_j(t) = q_j(t^j_{k_j(t)})$, $e^x_j(t) = \hat{x}_j(t) - x_j(t)$, $e^q_j(t) = \hat{q}_j(t) - q_j(t)$, $\boldsymbol{x}(t) = [x_1(t), \dots, x_N(t)]^\top$, and $\boldsymbol{q}(t) = [q_1(t), \dots, q_N(t)]^\top$.

The key question is to develop triggering laws such that the proposed algorithms are free of Zeno behavior and also converge to the global minimizer. Such a result is presented in the following theorem.

**Theorem 1.** *Suppose that the undirected graph is connected. Let Assumption 1 holds. If each agent $i \in \mathcal{V}$ runs the distributed event-triggered algorithm in (2), and determines its triggering time sequence by*

$$\begin{aligned}t^i_{k+1} = \max_{t \ge t^i_k} \Big\{ t : \|x_i(t) - x_i(t^i_k)\| \le a_i e^{-b_i t} \ \text{ and } \\ \|q_i(t) - q_i(t^i_k)\| \le c_i e^{-d_i t} \Big\}, k = 1, 2, \dots,\end{aligned} \tag{3}$$

*where all $a_i, b_i, c_i, d_i > 0$ are design parameters, then (i) there is no Zeno behavior; (ii) every $x_i(t)$, $i \in \mathcal{V}$ asymptotically converges to the global minimizer $x^*$; (iii) if Assumptions 2–3 also hold, then every $x_i(t)$, $i \in \mathcal{V}$ exponentially converges to the unique global minimizer $x^*$ with a rate no less than $\frac{\epsilon_2}{2\epsilon_3}$, where $\epsilon_2$ and $\epsilon_3$ are two positive constants and are given in the proof.*

**Proof:** The proof is given in Appendix B. ∎

If we let $a_i = c_i = 0$, $i \in \mathcal{V}$ in the triggering law (3), then $t^j_{k_j(t)} = t$, i.e., the event-triggered algorithm (2) becomes the following algorithm:

$$\dot{q}_i(t) = \sum_{j=1}^{N} L_{ij} x_j(t),\ \forall q_i(0), \tag{4a}$$

$$\dot{x}_i(t) = -\sum_{j=1}^{N} L_{ij} x_j(t) - \sum_{j=1}^{N} L_{ij} q_j(t) - \nabla f_i(x_i(t)),\ \forall x_i(0), \tag{4b}$$

which is exactly the distributed PI algorithm proposed in [21], [22]. In this case, from the proof of Theorem 1, we know that the second and the third statements still hold, which leads to the following result.

**Corollary 1.** *Suppose the undirected graph is connected. Let Assumption 1 holds. If each agent $i \in \mathcal{V}$ runs the distributed event-triggered algorithm in (4), then (i) every $x_i(t)$, $i \in \mathcal{V}$ asymptotically converges to the global minimizer $x^*$; (ii) if Assumptions 2–3 also hold, then every $x_i(t)$ exponentially converges to the unique global minimizer $x^*$.*

**Remark 1.** *Note that in [21], [22], the asymptotic convergence is established even for more general non-differentiable cost functions. Also in [22, Theorem 5.4], the asymptotic convergence is established for the case where the local cost functions are convex and differentiable with globally Lipschitz gradients and the directed graph is weight-balanced and strongly connected. Here we focus on undirected connected graphs. Corollary 1 establishes the exponential convergence under an additional condition that the global cost function is restricted strongly convex with respect to the global minimizer.*

Our second proposed event-triggered algorithm is based on the algorithm proposed in [23] and is described as follows:

$$\dot{q}_i(t) = \sum_{j=1}^{N} L_{ij} x_j(t^j_{k_j(t)}),\ \sum_{i=1}^{N} q_i(0) = 0, \tag{5a}$$

$$\begin{aligned}\dot{x}_i(t) = &-\sum_{j=1}^{N} L_{ij} x_j(t^j_{k_j(t)}) - q_i(t) \\ &- \nabla f_i(x_i(t)),\ \forall x_i(0,)\ t \in [t^i_k, t^i_{k+1}).\end{aligned} \tag{5b}$$

Similar results as shown in Theorem 1 are given in the following theorem. We omit the details due to the space limitation.

**Theorem 2.** *Suppose that the undirected graph is connected. Let Assumption 1 holds. If each agent $i \in \mathcal{V}$ runs the distributed event-triggered algorithm in (5), and determines its triggering time sequence by*

$$t^i_{k+1} = \max_{t \ge t^i_k} \Big\{ t :\ \|x_i(t) - x_i(t^i_k)\| \le a_i e^{-b_i t} \Big\},\ k = 1, 2, \dots \tag{6}$$

*where both $a_i$ and $b_i > 0$ are design parameters, then (i) there is no Zeno behavior; (ii) every $x_i(t)$, $i \in \mathcal{V}$ asymptotically converges to the global minimizer $x^*$; (iii) if Assumptions 2–3 also hold, then every $x_i(t)$ exponentially converges to the unique global minimizer $x^*$.*

**Remark 2.** *The proposed algorithms (2) and (5) are motivated by the distributed PI algorithms proposed in [21], [22] and [23], respectively. However, here we consider event-triggered communication schemes inspired by the general time-dependent triggering laws proposed in [31].*

*Different from the algorithm (2) where both $x_i(0)$ and $q_i(0)$ can be arbitrarily chosen, in the algorithm (5), it is required that $\sum_{i=1}^N q_i(0) = 0$ although $x_i(0)$ can be arbitrarily chosen. In other words, the algorithm (2) is robust to the initial condition $q_i(0)$. However, the algorithm (2) requires additional communication of $q_j$ in (2a), compared to the algorithm (5). This additional communication also makes the triggering laws (3) more involved compared to that of the triggering law (6).*

**Remark 3.** *Theorems 1 and 2 show that the proposed event-trigger algorithms exponentially converge to the unique* *global minimizer if the local cost functions are convex and differentiable with locally Lipschitz gradients and the global cost function is restricted strongly convex. Note that not all local cost functions need to be so or strongly convex, which is a less restricted condition, compared with the condition that all local cost functions are strongly convex required in [23, Theorem 13]. Moreover, our proposed algorithms exponentially converge to the exact global minimizer, while the algorithm proposed in [23] only converges to a neighborhood of the global minimizer. Compared to existing event-triggered ZGS algorithms proposed in [38], [39], which require the special initialization, our proposed algorithms allow an arbitrary initialization in* $x_i(0)$.

If we let $a_i = 0$, $i \in \mathcal{V}$ in the triggering law (6), then $t^j_{k_j(t)} = t$, i.e., the event-triggered algorithm (5) becomes:

$$\dot{q}_i(t) = \sum_{j=1}^{N} L_{ij} x_j(t), \ \sum_{i=1}^{N} q_i(0) = 0, \tag{7a}$$

$$\dot{x}_i(t) = -\sum_{j=1}^{N} L_{ij} x_j(t) - q_i(t) - \nabla f_i(x_i(t)), \ \forall x_i(0), \tag{7b}$$

which is exactly the distributed PI algorithm proposed in [23]. In this case, from the proof of Theorem 2, we know that the second and the third statements still hold, which leads to the following result.

**Corollary 2.** *Suppose the undirected graph is connected. Let Assumption 1 holds. If each agent* $i \in \mathcal{V}$ *runs the distributed event-triggered algorithm in (7), then (i) every* $x_i(t)$, $i \in \mathcal{V}$ *asymptotically converges to the global minimizer* $x^*$*; (ii) if Assumptions 2–3 also hold, then every* $x_i(t)$ *exponentially converges to the unique global minimizer* $x^*$.

**Remark 4.** *In [23, Theorem 6], the exponential convergence is established under the condition that local cost functions are strongly convex. Our exponential convergence result established in Corollary 2 only requires the global cost function to be restricted strongly convex with respect to its global minimizer* $x^*$*. Note that not all local cost functions need to be so or strongly convex, which is a less restricted condition, compared with that in [23].*

## V. Simulations

Consider a network with 8 agents whose topology is given in Fig. 1.

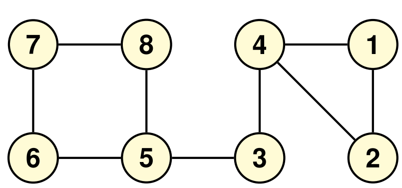


Fig. 1. Network topology.

The local cost functions $f_i$ are given by:

$$\begin{aligned} &f_1(x) = x^2 + 2x, \\ &f_2(x) = \ln(e^{-0.1x} + e^{0.3x}) + 0.1x^2, \\ &f_3(x) = f_4(x) = \tfrac{x^2}{\sqrt{x^2+1}} + 0.1x^2, \\ &f_5(x) = 2x^2, \qquad f_6(x) = x^2 + 2x, \\ &f_7(x) = x^2 + 1, \qquad f_8(x) = x^2 + x + 1. \end{aligned}$$

For the proposed algorithm in (2), we randomly choose the design parameters $a_i$, $b_i$, $c_i$, and $d_i$ in the triggering law (3). We choose the sample length as 0.01. Figs. 2 and 3 show the simulation results. Fig. 2 shows the state evolutions for each agent, from which we clearly see that all agents converge to the global minimizer $x^* = -0.3$. Fig. 3 shows the corresponding triggering time for each agent. Agents 1, 2, and 4 are triggered more often compared to other agents. Also, they are triggered more often during the time interval $[7, 15]$ compared to the time interval $[0, 7]$. We have also computed the number of triggering times during time interval $[0, 15]$. In particular, agents 1-8 are triggered 179, 143, 23, 256, 96, 25, 55, 44 times, respectively. As a result, the event-triggered algorithm (2) avoids about 73% sampling in this simulation.

For the proposed algorithm in (5), we randomly choose the design parameters $a_i$ and $b_i$ in the triggering law (6). We choose the sample length as 0.01. The simulation results are shown in Figs. 4 and 5. Fig. 4 plots the state evolutions for each agent and shows that all agents converge to the global minimizer $x^* = -0.3$. Fig. 5 shows the corresponding triggering time for each agent. Compared to the result for the algorithm (5) in Fig. 3, agents are triggered less often, which saves more communication and energy. We have also computed the number of triggering times during time interval $[0, 15]$. During time interval $[0, 15]$, agent 1-8 are triggered 27, 28, 19, 64, 31 20, 17, 23 times, respectively. Thus, the event-trigger algorithm (5) avoids 85% sampling.

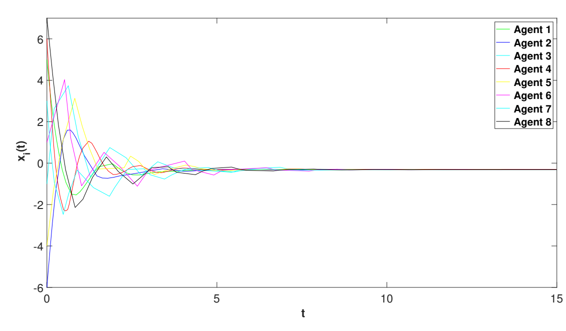


Fig. 2. State evolution for each agent of Algorithm (2).

## VI. Conclusions

In this paper, we developed distributed event-triggered PI algorithms for solving the distributed optimization problem over undirected connected graphs. For the case where local cost functions are convex and differentiable, we showed that the proposed algorithms are free of Zeno behavior and asymptotically converge to one of the global minimizers.

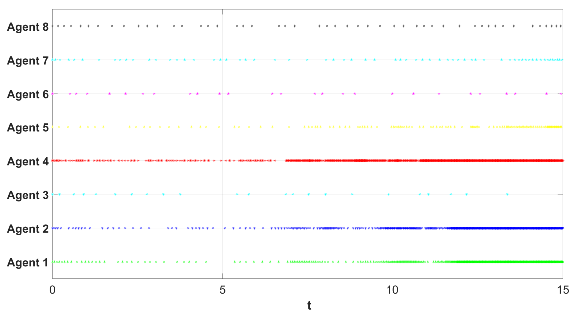


Fig. 3. Triggering time for each agent of Algorithm (2).

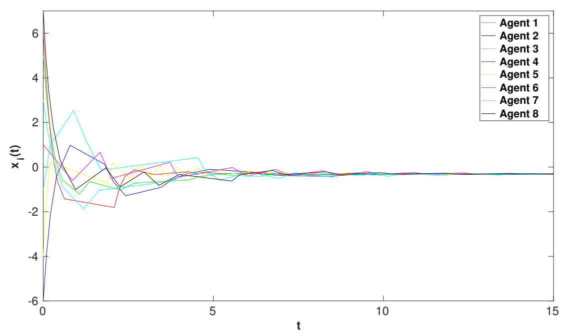


Fig. 4. State evolution for each agent of Algorithm (5).

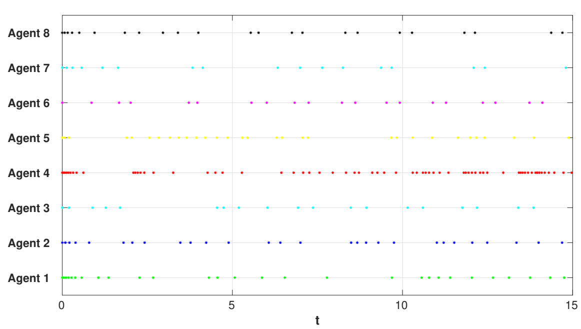


Fig. 5. Triggering time for each agent of Algorithm (5).

Moreover, if in addition, the local cost functions have locally Lipschitz gradients, and the global cost function is restricted strongly convex with respect to the global minimizer, we showed that the proposed algorithms exponentially converge to the unique global minimizer. One future direction is to develop dynamic event-triggered communication mechanisms for distributed PI algorithms.

## References


[1] J. N. Tsitsiklis, "Problems in decentralized decision making and computation," Ph.D. dissertation, MIT, Cambridge, MA, 1984.

[2] J. N. Tsitsiklis, D. P. Bertsekas, and M. Athans, "Distributed asynchronous deterministic and stochastic gradient optimization algorithms," *IEEE Transactions on Automatic Control*, vol. 31, no. 9, pp. 803–812, 1986.

[3] B. Johansson, T. Keviczky, M. Johansson, and K. H. Johansson, "Subgradient methods and consensus algorithms for solving convex optimization problems," in *Proceedings of the IEEE Conference on Decision and Control*, 2008, pp. 4185–4190.

[4] A. Nedić and A. Ozdaglar, "Distributed subgradient methods for multi-agent optimization," *IEEE Transactions on Automatic Control*, vol. 54, no. 1, pp. 48–61, 2009.

[5] M. Zhu and S. Martínez, "On distributed convex optimization under inequality and equality constraints," *IEEE Transactions on Automatic Control*, vol. 57, no. 1, pp. 151–164, 2012.

[6] K. I. Tsianos, S. Lawlor, and M. G. Rabbat, "Push-sum distributed dual averaging for convex optimization," in *Proceedings of the IEEE Conference on Decision and Control*, 2012, pp. 5453–5458.

[7] A. Nedić and A. Olshevsky, "Distributed optimization over time-varying directed graphs," *IEEE Transactions on Automatic Control*, vol. 60, no. 3, pp. 601–615, 2015.

[8] T. Yang, J. Lu, D. Wu, J. Wu, G. Shi, Z. Meng, and K. H. Johansson, "A distributed algorithm for economic dispatch over time-varying directed networks with delays," *IEEE Transactions on Industrial Electronics*, vol. 64, no. 6, pp. 5095–5106, 2017.

[9] I. Matei and J. S. Baras, "Performance evaluation of the consensus-based distributed subgradient method under random communication topologies," *IEEE Journal of Selected Topics in Signal Processing*, vol. 5, no. 4, pp. 754–771, 2011.

[10] K. Yuan, Q. Ling, and W. Yin, "On the convergence of decentralized gradient descent," *SIAM Journal on Optimization*, vol. 26, no. 3, pp. 1835–1854, 2015.

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

[12] L. Yao, Y. Yuan, S. Sundaram, and T. Yang, "Distributed finite-time optimization," in *Proceedings of the IEEE Conference on Control & Automation*, 2018, pp. 147–154.

[13] G. Qu and N. Li, "Harnessing smoothness to accelerate distributed optimization," *IEEE Transactions on Control of Network Systems*, vol. 5, no. 3, pp. 1245–1260, 2018.

[14] J. Xu, S. Zhu, Y. C. Soh, and L. Xie, "Augmented distributed gradient methods for multi-agent optimization under uncoordinated constant stepsizes," in *Proceedings of the IEEE Conference on Decision and Control*, 2015, pp. 2055–2060.

[15] Z. Li, W. Shi, and M. Yan, "A decentralized proximal-gradient method with network independent step-sizes and separated convergence rates," *arXiv preprint arXiv:1704.07807*, 2017.

[16] S. Yang, S. Tan, and J.-X. Xu, "Consensus based approach for economic dispatch problem in a smart grid," *IEEE Transactions on Power Systems*, vol. 28, no. 4, pp. 4416–4426, 2013.

[17] W. Du, L. Yao, D. Wu, X. Li, G. Liu, and T. Yang, "Accelerated distributed energy management for microgrids," in *Proceedings of the IEEE Power and Energy Society General Meeting*, 2018.

[18] S. Pu, W. Shi, J. Xu, and A. Nedić, "A push-pull gradient method for distributed optimization in networks," in *Proceedings of the IEEE Conference on Decision and Control*, 2018, pp. 3385–3390.

[19] R. Xin and U. A. Khan, "A linear algorithm for optimization over directed graphs with geometric convergence," *IEEE Control Systems Letters*, vol. 2, no. 3, pp. 325–330, 2018.

[20] M. Zhu and S. Martínez, "Discrete-time dynamic average consensus," *Automatica*, vol. 46, no. 2, pp. 322–329, 2010.

[21] J. Wang and N. Elia, "Control approach to distributed optimization," in *Proceedings of Annual Allerton Conference on Communication, Control, and Computing (Allerton)*, 2010, pp. 557–561.

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

[23] 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.

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

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

[26] E. Wei, A. Ozdaglar, and A. Jadbabaie, "A distributed Newton method for network utility maximization-I: Algorithm," *IEEE Transactions on Automatic Control*, vol. 58, no. 9, pp. 2162–2175, 2013.

[27] K. J. Aström and B. M. Bernhardsson, "Comparison of Riemann and Lebesgue sampling for first order stochastic systems," in *Proceedings of the IEEE Conference on Decision and Control*, 2002, pp. 2011–2016.
[28] P. Tabuada, "Event-triggered real-time scheduling of stabilizing control tasks," *IEEE Transactions on Automatic Control*, vol. 52, no. 9, pp. 1680–1685, 2007.
[29] A. Girard, "Dynamic triggering mechanisms for event-triggered control," *IEEE Transactions on Automatic Control*, vol. 60, no. 7, pp. 1992–1997, 2015.
[30] D. V. Dimarogonas, E. Frazzoli, and K. H. Johansson, "Distributed event-triggered control for multi-agent systems," *IEEE Transactions on Automatic Control*, vol. 57, no. 5, pp. 1291–1297, 2012.
[31] G. S. Seyboth, D. V. Dimarogonas, and K. H. Johansson, "Event-based broadcasting for multi-agent average consensus," *Automatica*, vol. 49, no. 1, pp. 245–252, 2013.
[32] X. Meng, L. Xie, Y. C. Soh, C. Nowzari, and G. J. Pappas, "Periodic event-triggered average consensus over directed graphs," in *Proceedings of the IEEE Conference on Decision and Control*, 2015, pp. 2055–2060.
[33] X. Meng, L. Xie, and Y. C. Soh, "Asynchronous periodic event-triggered consensus for multi-agent systems," *Automatica*, vol. 84, pp. 214 – 220, 2017.
[34] X. Yi, "Resource-constrained multi-agent control systems: Dynamic event-triggering, input saturation, and connectivity preservation," Licentiate Thesis, Royal Institute of Technology, Sweden, 2017.
[35] C. Nowzari, J. Cortés, and G. J. Pappas, *Event-triggered control for multi-agent average consensus*. Wiley, 2018, ch. 7, pp. 177–208.
[36] X. Yi, T. Yang, J. Wu, and K. H. Johansson, "Distributed event-triggered control for global consensus of multi-agent systems with input saturation," *Automatica*, vol. 100, pp. 1–9, 2019.
[37] S. Liu, L. Xie, and D. E. Quevedo, "Event-triggered quantized communication-based distributed convex optimization," *IEEE Transactions on Control of Network Systems*, vol. 5, no. 1, pp. 167–178, 2018.
[38] W. Chen and W. Ren, "Event-triggered zero-gradient-sum distributed consensus optimization over directed networks," *Automatica*, vol. 65, pp. 90–97, 2016.
[39] W. Du, X. Yi, J. George, K. H. Johansson, and T. Yang, "Distributed optimization with dynamic event-triggered mechanisms," in *Proceedings of the IEEE Conference on Decision and Control*, 2018, pp. 969–974.
[40] 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 *Proceedings of the IEEE Conference on Decision and Control*, 2018, pp. 3397–3402.
[41] D. Wang, V. Gupta, and W. Wang, "An event-triggered protocol for distributed optimal coordination of double-integrator multi-agent systems," *Neurocomputing*, vol. 319, pp. 34–41, 2018.
[42] K. H. Johansson, M. Egerstedt, J. Lygeros, and S. Sastry, "On the regularization of Zeno hybrid automata," *Systems & Control Letters*, vol. 38, no. 3, pp. 141 – 150, 1999.
[43] C. Godsi and G. F. Royle, *Algebraic Graph Theory*, ser. Graduate Texts in Mathematics. New York: Springer-Verlag, 2001, vol. 207.
[44] H. K. Khalil, *Nonlinear Systems*, 3rd ed. Prentice-Hall, New Jersey, 2002.

## Appendix

### A. Useful Lemmas

The following lemma extends [11, Proposition 3.6] by relaxing the globally Lipschitz gradient assumption on each local cost function with the locally Lipschitz gradient assumption, which plays an important role in the proof of the exponential convergence later.

**Lemma 1.** *Suppose that the undirected graph is connected. Let Assumptions 1–3 hold. Then for any $\varepsilon > 0$ and any compact and convex set $D \subseteq \mathbb{R}^n$ with $x^* \in D$,*

$$\begin{aligned}&(\nabla f(\boldsymbol{x}) - \nabla f(\bar{\boldsymbol{x}}))^\top (\boldsymbol{x} - \bar{\boldsymbol{x}}) + \varepsilon \boldsymbol{x}^\top (L \otimes I_n)\boldsymbol{x}\\ &\geq m(\varepsilon)\|\boldsymbol{x} - \bar{\boldsymbol{x}}\|^2,\ \forall \boldsymbol{x} \in D^N,\end{aligned} \tag{8}$$

*where $\boldsymbol{x} = [x_1, \dots, x_N]$, $f(\boldsymbol{x}) = \sum_{i=1}^N f_i(x_i)$, $\bar{\boldsymbol{x}} = \mathbf{1}_N \otimes x^*$, $m(\varepsilon) = \min\Big\{m_f - 2M(D)\iota, \frac{\varepsilon\rho_2(L)\iota^2}{1+\iota^2}\Big\}$, $M(D) = \max_{i\in\mathcal{V}}\{M_i(D)\}$, and $\iota \in (0, \frac{m_f}{2M(D)})$.*

**Proof:** For any $\boldsymbol{x} \in D^N$, decompose $\boldsymbol{x} = \boldsymbol{u} + \boldsymbol{v}$, where $\boldsymbol{u} = \mathbf{1}_N \otimes u_0$ with $u_0 = \frac{1}{N}\sum_{i=1}^N x_i \in D$ and $\boldsymbol{v} = \boldsymbol{x} - \boldsymbol{u}$. It is straightforward to see that $\boldsymbol{v}^\top(\mathbf{1}_N \otimes I_n) = \mathbf{0}_N$. The rest of the proof is the same as the proof of Proposition 3.6 in [11]. ■

The following lemma from [40] is also useful.

**Lemma 2.** *Suppose that the undirected graph is connected. Let $K_N = I_N - \frac{1}{N}\mathbf{1}_N\mathbf{1}_N^\top$. Then the Laplacian matrix $L$ is positive semi-definite, $K_N\mathbf{1}_N = 0$,*

$$K_N L = LK_N = L,\ \rho(K_n) = 1,\ 0 \leq \rho_2(L)K_n \leq L. \tag{9}$$

### B. Proof of Theorem 1

(i) In this part, we show that there is no Zeno behavior by contradiction. Suppose that Zeno behavior exists. Then there is 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 $q_i(t)$ are continuous. Therefore there exist constants $P_1 > 0$ and $P_2 > 0$, such that $\|x_i(t)\| \leq P_1$ and $\|q_i(t)\| \leq P_2$ for all $i \in \mathcal{V}$ and for all $t \in [0, T_0]$.

From Assumption 1, we know that $f(\boldsymbol{x})$ is continuously differentiable. Also noting that $\|x_i(t)\| \leq P_1,\ \forall i \in \mathcal{V},\ \forall t \in [0, T_0]$. Therefore, there exists a constant $P_3 > 0$ such that $\|\nabla f(\boldsymbol{x})\| \leq P_3,\ \forall t \in [0, T_0]$.

Let $C_1 = 2L_{ii}P_1 + 2L_{ii}P_2 + P_3$ and $C_2 = 2L_{ii}P_1$. It then follows from (2) that

$$\begin{aligned}\|\dot{q}_i(t)\| &\leq \|\sum_{j=1}^N L_{ij}x_j(t^j_{k_j(t)})\| \leq C_2,\\ \|\dot{x}_i(t)\| &\leq \|\sum_{j=1}^N L_{ij}x_j(t^j_{k_j(t)})\| + \|\sum_{j=1}^N L_{ij}q_j(t^j_{k_j(t)})\|\\ &\quad + \|\nabla f_i(x_i(t))\| \leq C_1.\end{aligned} \tag{10}$$

Define $a_0 = \min\{a_1, \dots, a_N\}$, $b_0 = \max\{b_1, \dots, b_N\}$, $c_0 = \min\{c_1, \dots, c_N\}$, $d_0 = \max\{d_1, \dots, d_N\}$, and $\epsilon_0 = \min\{\frac{a_0e^{-b_0T_0}}{2C_1}, \frac{c_0e^{-d_0T_0}}{2C_2}\} > 0$. It then follows from the property of limits that there exists a positive integer $N(\epsilon_0) \in \mathbb{Z}^+$ such that

$$t_k^i \in [T_0 - \epsilon, T_0],\ \forall k \geq N(\epsilon_0). \tag{11}$$

Given the triggering law (3), we know that

$$\begin{aligned}&\|e_i^x(t)\| = \|x_i(t) - x_i(t_k^i)\| \leq a_ie^{-b_it},\\ &\|e_i^q(t)\| = \|q_i(t) - q_i(t_k^i)\| \leq c_ie^{-d_it},\ \forall t \geq 0.\end{aligned} \tag{12}$$

Since $\hat{x}_i(t_k^i) = x_i(t_k^i)$ for any triggering time $t_k^i$, it the follows from (10) that one sufficient condition to ensure (12) is

$$C_1(t - t_k^i) \leq a_0e^{-b_0t} \text{ and } C_2(t - t_k^i) \leq c_0e^{-d_0t}. \tag{13}$$

Suppose that we have determined the $N(\epsilon)$-th triggering time of agent $i$, which is denoted by $t^i_{N(\epsilon_0)}$. Let $t^i_{N(\epsilon_0)+1}$

and $\tilde{t}^i_{N(\epsilon_0)+1}$ denote the next triggering time determine by (3) and (13), respectively. Then, $t^i_{N(\epsilon_0)+1} \geq \tilde{t}^i_{N(\epsilon_0)+1}$ and

$$\begin{aligned}\tilde{t}^i_{N(\epsilon_0)+1} =& t^i_{N(\epsilon_0)} + \min\Big\{\frac{a_0 e^{-b_0 t^i_{N(\epsilon_0)}}}{C_1}, \frac{c_0 e^{-d_0 t^i_{N(\epsilon_0)}}}{C_2}\Big\} \\ \geq& t^i_{N(\epsilon_0)} + \min\Big\{\frac{a_0 e^{-b_0 T_0}}{C_1}, \frac{c_0 e^{-d_0 T_0}}{C_2}\Big\} \\ =& t^i_{N(\epsilon_0)} + 2\epsilon_0,\end{aligned}$$

where the inequality holds since $t^i_{N(\epsilon_0)} \leq T_0$. Thus,

$$t^i_{N(\epsilon_0)+1} - t^i_{N(\epsilon_0)} \geq \tilde{t}^i_{N(\epsilon_0)+1} - t^i_{N(\epsilon_0)} \geq 2\epsilon_0,$$

which contradicts (11). Hence, the event-triggered algorithm (2) is free of Zeno behavior.

(ii) In this part, we use the Lyapunov stability analysis to show that every $x_i(t)$, $i \in \mathcal{V}$ asymptotically converges to the global minimizer $x^*$, which may be nonunique, if Assumption 1 holds. For convenience, let us denote $\hat{\boldsymbol{x}} = [\hat{x}_1, \dots, \hat{x}_N]$, $\hat{\boldsymbol{q}} = [\hat{q}_1, \dots, \hat{q}_N]$, $\boldsymbol{e}^x = [e^x_1, \dots, e^x_N]$, $\boldsymbol{e}^q = [e^q_1, \dots, e^q_N]$, and $\nabla f(\boldsymbol{x}) = [\nabla f_1(x_1), \dots, f_N(x_N)]$. Then the algorithm (2) can be rewritten in a compact form:

$$\dot{\boldsymbol{q}}(t) = (L \otimes I_n)\, \hat{\boldsymbol{x}}(t), \tag{14a}$$

$$\dot{\boldsymbol{x}}(t) = -(L \otimes I_n)\hat{\boldsymbol{x}}(t) - (L \otimes I_n)\hat{\boldsymbol{q}}(t) - \nabla f(\boldsymbol{x}(t)). \tag{14b}$$

Consider the following function:

$$V_1(\boldsymbol{x}, \boldsymbol{q}) = \frac{1}{2}\|\boldsymbol{x} - \bar{\boldsymbol{x}}\|^2 + \frac{1}{2}\|\boldsymbol{q} - \bar{\boldsymbol{q}}\|^2, \tag{15}$$

where $\bar{\boldsymbol{x}} = \mathbf{1}_N \otimes x^*$, and $\bar{\boldsymbol{q}}$ is such that $(L \otimes I_n)\bar{\boldsymbol{q}} = -\nabla f(\bar{\boldsymbol{x}})$ (the existence of such a $\bar{\boldsymbol{q}}$ follows from that [22, Proposition 3.2].)

The derivative of $V_1(\boldsymbol{x}, \boldsymbol{q})$ along the trajectories of (14) satisfies

$$\begin{aligned}\dot{V}_1 =& (\boldsymbol{x} - \bar{\boldsymbol{x}})^\top[-(L \otimes I_n)(\boldsymbol{x} + \boldsymbol{e}^x) - (L \otimes I_n)(\boldsymbol{q} + \boldsymbol{e}^q) \\ &- \nabla f(\boldsymbol{x})] + (\boldsymbol{q} - \bar{\boldsymbol{q}})^\top (L \otimes I_n)(\boldsymbol{x} + \boldsymbol{e}^x) \\ =& (\boldsymbol{x} - \bar{\boldsymbol{x}})^\top[-(L \otimes I_n)(\boldsymbol{x} + \boldsymbol{e}^x) - (L \otimes I_n)(\boldsymbol{q} - \bar{\boldsymbol{q}} + \boldsymbol{e}^q) \\ &+ \nabla f(\bar{\boldsymbol{x}}) - \nabla f(\boldsymbol{x})] + (\boldsymbol{q} - \bar{\boldsymbol{q}})^\top (L \otimes I_n)(\boldsymbol{x} + \boldsymbol{e}^x) \\ =& -\boldsymbol{x}^\top (L \otimes I_n)\boldsymbol{x} - \boldsymbol{x}^\top (L \otimes I_n)\boldsymbol{e}^x - \boldsymbol{x}^\top (L \otimes I_n)\boldsymbol{e}^q \\ &- (\boldsymbol{x} - \bar{\boldsymbol{x}})^\top [\nabla f(\boldsymbol{x}) - \nabla f(\bar{\boldsymbol{x}})] + (\boldsymbol{q} - \bar{\boldsymbol{q}})^\top (L \otimes I_n)\boldsymbol{e}^x \\ \leq& -\frac{1}{2}\boldsymbol{x}^\top (L \otimes I_n)\boldsymbol{x} + \rho(L)\|\boldsymbol{e}^x\|^2 + \rho(L)\|\boldsymbol{e}^q\|^2 \\ &+ \rho(L)\|\boldsymbol{q} - \bar{\boldsymbol{q}}\|\|\boldsymbol{e}^x\| - (\boldsymbol{x} - \bar{\boldsymbol{x}})^\top \big(\nabla f(\boldsymbol{x}) - \nabla f(\bar{\boldsymbol{x}})\big)\end{aligned} \tag{16}$$

$$\begin{aligned}\leq& -\frac{1}{2}\boldsymbol{x}^\top (L \otimes I_n)\boldsymbol{x} + \rho(L)(\|\boldsymbol{e}^x\|^2 + \|\boldsymbol{e}^q\|^2) \\ &+ \frac{b}{2}\|\boldsymbol{q} - \bar{\boldsymbol{q}}\|^2 + \frac{(\rho(L))^2}{2b}\|\boldsymbol{e}^x\|^2 \\ &- (\boldsymbol{x} - \bar{\boldsymbol{x}})^\top \big(\nabla f(\boldsymbol{x}) - \nabla f(\bar{\boldsymbol{x}})\big),\end{aligned} \tag{17}$$

where the first equality follows from $\boldsymbol{e}^x = \hat{\boldsymbol{x}} - \boldsymbol{x}$ and $\boldsymbol{e}^q = \hat{\boldsymbol{q}} - \boldsymbol{q}$; the second equality follows from $(L \otimes I_n)\bar{\boldsymbol{q}} = -\nabla \boldsymbol{f}(\bar{\boldsymbol{x}})$; and the last equality follows from $\bar{\boldsymbol{x}}^\top (L \otimes I_n) = 0$, the inequalities follows from the Young's inequality and the Cauchy-Schwarz inequality, and $b = \min\{b_1, \dots, b_N, d_1, \dots, d_N\} > 0$.

Also define $a = \max\{a_1, \dots, a_N, c_1, \dots, c_N\} > 0$. Then from (3), we have

$$\|\boldsymbol{e}^x\|^2 \leq Na^2 e^{-2bt} \text{ and } \|\boldsymbol{e}^q\|^2 \leq Na^2 e^{-2bt}. \tag{18}$$

Note that from (15), we have

$$\|\boldsymbol{q} - \bar{\boldsymbol{q}}\|^2 \leq 2V_1(\boldsymbol{x}, \boldsymbol{q}). \tag{19}$$

It then follows from (17), (18), and (19) that

$$\dot{V}_1 \leq bV_1 + 2\rho(L)Na^2 e^{-2bt} + \frac{1}{2b}(\rho(L))^2 Na^2 e^{-2bt}.$$

Therefore, we have

$$V_1(\boldsymbol{x}(t), \boldsymbol{q}(t)) \leq \gamma_1 e^{bt},$$

where $\gamma_1 = V_1(0) + \frac{2}{3b}\rho(L)Na^2 + \frac{1}{6b}(\rho(L))^2 Na^2$.

Then, from (19), we have

$$\|\boldsymbol{q} - \bar{\boldsymbol{q}}\| \leq \sqrt{2\gamma_1} e^{\frac{b}{2}t}. \tag{20}$$

It then follows from (16), (18), and (20) that

$$\begin{aligned}\dot{V}_1 \leq& -\frac{1}{2}\boldsymbol{x}^\top (L \otimes I_n)\boldsymbol{x} + \gamma_2 N e^{-\frac{b}{2}t} \\ &- (\boldsymbol{x} - \bar{\boldsymbol{x}})^\top \big(\nabla f(\boldsymbol{x}) - \nabla f(\bar{\boldsymbol{x}})\big),\end{aligned} \tag{21}$$

where $\gamma_2 = a(2a + \sqrt{\frac{2\gamma_1}{N}})\rho(L)$.

Define $z_i(t) = e^{-\frac{b}{4}t}$ for $t \geq 0$ and for all $i \in \mathcal{V}$, and $\boldsymbol{z}(t) = [z_1(t), \dots, z_N(t)]$. Consider the following Lyapunov candidate:

$$W_1(\boldsymbol{x}, \boldsymbol{q}, \boldsymbol{z}) = V_1(\boldsymbol{x}, \boldsymbol{q}) + \frac{4\gamma_2}{b}\|\boldsymbol{z}\|^2. \tag{22}$$

Then from (21), we know that the derivative of $W_1(\boldsymbol{x}, \boldsymbol{q}, \boldsymbol{z})$ along the trajectories of (14) and $z_i(t) = e^{-\frac{b}{4}t}, t \geq 0$ satisfies

$$\begin{aligned}\dot{W}_1 \leq& -\frac{1}{2}\boldsymbol{x}^\top (L \otimes I_n)\boldsymbol{x} - \gamma_2\|\boldsymbol{z}\|^2 \\ &- (\boldsymbol{x} - \bar{\boldsymbol{x}})^\top \big(\nabla f(\boldsymbol{x}) - \nabla f(\bar{\boldsymbol{x}})\big) \leq 0.\end{aligned} \tag{23}$$

Noting that $W_1(\boldsymbol{x}, \boldsymbol{q}, \boldsymbol{z})$ is radially unbounded, then by LaSalle's Invariance Principle [44], we know that $\boldsymbol{x}(t)$ asymptotically converges to $\{\mathbf{1}_N \otimes x^1 \in \mathbb{R}^{nN} : \sum_{i=1}^n (x^1 - x^*)(\nabla f_i(x^1) - \nabla f_i(x^*)) = 0\}$. Noting that $\sum_{i=1}^n \nabla f_i(x^*) = 0$, we have $\sum_{i=1}^n (x^1 - x^*)(\nabla f_i(x^1) - \nabla f_i(x^*)) = 0$ is equivalent to $\sum_{i=1}^n \nabla f_i(x^1) = 0$, i.e., $x^1$ is a global minimizer. Thus, $\boldsymbol{x}(t)$ asymptotically converges to $\mathbf{1}_N \otimes x^*$, i.e., every $x_i(t)$, $i \in \mathcal{V}$ asymptotically converges to $x^*$.

(iii) In this part, we show that the converge speed is exponential if Assumptions 2–3 also hold.

We first show that for each initial states $\boldsymbol{x}(0)$ and $\boldsymbol{q}(0)$, there exists a convex and compact set $\mathcal{C} \subseteq \mathbb{R}^n$ such that $x^* \in \mathcal{C}$ and $x_i(t) \in \mathcal{C}$, $\forall t \geq 0$, $\forall i \in \mathcal{V}$. The exact form of the set $\mathcal{C}$ depends $\boldsymbol{x}(0)$, $\boldsymbol{q}(0)$, and $x^*$ and is given next.

From (15), (22), and (23), we know that for all $t \geq 0$ and $i \in \mathcal{V}$,

$$\begin{aligned}&\|x_i(t) - x^*\|^2 \leq \|\boldsymbol{x}(t) - \bar{\boldsymbol{x}}\|^2 \leq 2W_1(\boldsymbol{x}(t), \boldsymbol{q}(t), \boldsymbol{z}(t)) \\ &\leq 2W_1(\boldsymbol{x}(0), \boldsymbol{q}(0), \boldsymbol{z}(0)).\end{aligned}$$

Thus, $\mathcal{C}=\{x\in\mathbb{R}^n:\ ||x-x^*\|^2\leq 2W_1(\boldsymbol{x}(0),\boldsymbol{q}(0),\boldsymbol{z}(0))\}$ is the convex and compact set that we aim to find.

Next, consider the following functions:

$$V_2(\boldsymbol{x},\boldsymbol{q})=(\boldsymbol{x}-\bar{\boldsymbol{x}})^\top(K_N\otimes I_n)(\boldsymbol{q}-\bar{\boldsymbol{q}}), \tag{24}$$

$$V_3(\boldsymbol{x},\boldsymbol{q})=\frac{\epsilon_1+1}{2}\|\boldsymbol{x}-\bar{\boldsymbol{x}}\|^2+\frac{\epsilon_1+2}{2}(\boldsymbol{q}-\bar{\boldsymbol{q}})^\top(K_N\otimes I_n)(\boldsymbol{q}-\bar{\boldsymbol{q}}), \tag{25}$$

where $\epsilon_1=\frac{2(M(\mathcal{C}))^2}{m_1\rho_2(L)}$, $m_1=\min\left\{m_f-2M(\mathcal{C})\iota,\frac{\rho_2(L)\iota^2}{2(1+\iota^2)}\right\}$, $M(\mathcal{C})=\max_{i\in\mathcal{V}}\{M_i(\mathcal{C})\}$, and $\iota\in(0,\frac{m_f}{2M(\mathcal{C})})$. Then, derivative of $V_2$ and $V_3$ along the trajectories of (14) satisfy

$$\begin{aligned}\dot V_2=&(\boldsymbol{q}-\bar{\boldsymbol{q}})^\top(K_N\otimes I_n)[-(L\otimes I_n)\hat{\boldsymbol{x}}-(L\otimes I_n)\hat{\boldsymbol{q}}\\&-\nabla f(\boldsymbol{x})]+(\boldsymbol{x}-\bar{\boldsymbol{x}})^\top(L\otimes I_n)\hat{\boldsymbol{x}}\\=&(\boldsymbol{q}-\bar{\boldsymbol{q}})^\top(K_N\otimes I_n)[-(L\otimes I_n)\hat{\boldsymbol{x}}-(L\otimes I_n)(\boldsymbol{q}-\bar{\boldsymbol{q}}\\&+\boldsymbol{e}^q)-(\nabla f(\boldsymbol{x})-\nabla f(\bar{\boldsymbol{x}}))]+(\boldsymbol{x}-\bar{\boldsymbol{x}})^\top(L\otimes I_n)\hat{\boldsymbol{x}}\\\leq&-\frac{3}{4}(\boldsymbol{q}-\bar{\boldsymbol{q}})^\top(L\otimes I_n)(\boldsymbol{q}-\bar{\boldsymbol{q}})+\rho(L)\|\boldsymbol{e}^q\|^2\\&-(\boldsymbol{q}-\bar{\boldsymbol{q}})^\top(L\otimes I_n)\hat{\boldsymbol{x}}(t)+(\boldsymbol{x}-\bar{\boldsymbol{x}})^\top(L\otimes I_n)\hat{\boldsymbol{x}}\\&+\frac{\rho_2(L)}{4}(\boldsymbol{q}-\bar{\boldsymbol{q}})^\top(K_N\otimes I_n)(\boldsymbol{q}-\bar{\boldsymbol{q}})\\&+\frac{1}{\rho_2(L)}\|\nabla f(\boldsymbol{x})-\nabla f(\bar{\boldsymbol{x}})\|^2\\\leq&-\frac{1}{2}(\boldsymbol{q}-\bar{\boldsymbol{q}})^\top(L\otimes I_n)(\boldsymbol{q}-\bar{\boldsymbol{q}})+\rho(L)\|\boldsymbol{e}^q\|^2\\&-(\boldsymbol{q}-\bar{\boldsymbol{q}})^\top(L\otimes I_n)\hat{\boldsymbol{x}}+(\boldsymbol{x}-\bar{\boldsymbol{x}})^\top(L\otimes I_n)\hat{\boldsymbol{x}}\\&+\frac{1}{\rho_2(L)}\|\nabla f(\boldsymbol{x})-\nabla f(\bar{\boldsymbol{x}})\|^2\\\leq&-\frac{1}{2}(\boldsymbol{q}-\bar{\boldsymbol{q}})^\top(L\otimes I_n)(\boldsymbol{q}-\bar{\boldsymbol{q}})+\rho(L)\|\boldsymbol{e}^q\|^2\\&-(\boldsymbol{q}-\bar{\boldsymbol{q}})^\top(L\otimes I_n)\hat{\boldsymbol{x}}+(\boldsymbol{x}-\bar{\boldsymbol{x}})^\top(L\otimes I_n)\hat{\boldsymbol{x}}\\&+\frac{(M(\mathcal{C}))^2}{\rho_2(L)}\|\boldsymbol{x}-\bar{\boldsymbol{x}}\|^2,\end{aligned}\tag{26}$$

where for the second equality, we have used the definition of $\boldsymbol{e}^q$, the property that $(L\otimes I_n)\bar{\boldsymbol{q}}=-\nabla\boldsymbol{f}(\bar{\boldsymbol{x}})$, and the property that $K_NL=L$ given in (9) of Lemma 2; for the first inequality, we have used the Young's inequality and the Cauchy-Schwarz inequality; for the second inequality, we have used the property that $\rho_2(L)K_N\leq L$ given in (9) of Lemma 2; and for the last inequality, we have used the fact that $f(\boldsymbol{x})$ has a locally Lipschitz gradient which follows from Assumption 3.

Similarly, we have

$$\begin{aligned}\dot V_3=&(\epsilon_1+1)(\boldsymbol{x}-\bar{\boldsymbol{x}})^\top[-(L\otimes I_n)\hat{\boldsymbol{x}}-(L\otimes I_n)\hat{\boldsymbol{q}}-\nabla f(\boldsymbol{x})]\\&(\epsilon_1+2)(\boldsymbol{q}-\bar{\boldsymbol{q}})^\top(L\otimes I_n)\hat{\boldsymbol{x}}\\=&(\boldsymbol{q}-\bar{\boldsymbol{q}})^\top(L\otimes I_n)\hat{\boldsymbol{x}}-(\boldsymbol{x}-\bar{\boldsymbol{x}})^\top(L\otimes I_n)\hat{\boldsymbol{x}}\\&-\epsilon_1(\boldsymbol{x}-\bar{\boldsymbol{x}})^\top(L\otimes I_n)(\boldsymbol{x}-\bar{\boldsymbol{x}}+\boldsymbol{e}^x)\\&-(\epsilon_1+1)(\boldsymbol{x}-\bar{\boldsymbol{x}})^\top(L\otimes I_n)\boldsymbol{e}^q\\&+(\epsilon_1+1)(\boldsymbol{q}-\bar{\boldsymbol{q}})^\top(L\otimes I_n)\boldsymbol{e}^x\\&-(\epsilon_1+1)(\boldsymbol{x}-\bar{\boldsymbol{x}})^\top[(\nabla f(\boldsymbol{x})-\nabla f(\bar{\boldsymbol{x}}))]\\\leq&(\boldsymbol{q}-\bar{\boldsymbol{q}})^\top(L\otimes I_n)\hat{\boldsymbol{x}}-(\boldsymbol{x}-\bar{\boldsymbol{x}})^\top(L\otimes I_n)\hat{\boldsymbol{x}}\\&-\frac{\epsilon_1}{2}\boldsymbol{x}^\top(L\otimes I_n)\boldsymbol{x}+\epsilon_1\rho(L)\|\boldsymbol{e}^x\|^2\\&+\frac{(\epsilon_1+1)^2\rho(L)}{\epsilon_1}\|\boldsymbol{e}^q\|^2\\&+\frac{1}{4}(\boldsymbol{q}-\bar{\boldsymbol{q}})^\top(L\otimes I_n)(\boldsymbol{q}-\bar{\boldsymbol{q}})+(\epsilon_1+1)^2\rho(L)\|\boldsymbol{e}^x\|^2\\&-(\epsilon_1+1)(\boldsymbol{x}-\bar{\boldsymbol{x}})^\top[(\nabla f(\boldsymbol{x})-\nabla f(\bar{\boldsymbol{x}}))].\end{aligned}\tag{27}$$

Since Assumptions 1–3 are satisfied, it follows from (8) that

$$\begin{aligned}&(\boldsymbol{x}-\bar{\boldsymbol{x}})^\top\big(\nabla f(\boldsymbol{x})-\nabla f(\bar{\boldsymbol{x}})\big)\\&\geq m_1\|\boldsymbol{x}-\bar{\boldsymbol{x}}\|^2-\frac{1}{2}\boldsymbol{x}^\top(L\otimes I_n)\boldsymbol{x}.\end{aligned}\tag{28}$$

Thus, from $(\boldsymbol{x}-\bar{\boldsymbol{x}})^\top[(\nabla f(\boldsymbol{x})-\nabla f(\bar{\boldsymbol{x}}))]\geq 0$, $\epsilon_1=\frac{2(M(\mathcal{C}))^2}{m_1\rho_2(L)}$, (18) and (26)–(28) we have

$$\begin{aligned}\dot V_2+\dot V_3\leq&-\frac{1}{4}(\boldsymbol{q}-\bar{\boldsymbol{q}})^\top(L\otimes I_n)(\boldsymbol{q}-\bar{\boldsymbol{q}})\\&-\epsilon_1m_1\|\boldsymbol{x}-\boldsymbol{x}^*\|^2+\gamma_3Ne^{-2bt},\end{aligned}\tag{29}$$

where $\gamma_3=a^2\rho(L)[1+\epsilon_1+(1+\epsilon_1)^2(1+\frac{1}{\epsilon_1})]$.

Define $\zeta_i(t)=e^{-bt}$ for $t\geq 0$ and for all $i\in\mathcal{V}$, and $\boldsymbol{\zeta}(t)=[\zeta_1(t),\ldots,\zeta_N(t)]$. Consider the following Lyapunov candidate:

$$W_2(\boldsymbol{x},\boldsymbol{q},\boldsymbol{\zeta})=V_2(\boldsymbol{x},\boldsymbol{q})+V_3(\boldsymbol{x},\boldsymbol{q})+\frac{\gamma_3}{b}\|\boldsymbol{\zeta}\|^2. \tag{30}$$

Then from (29), we know that the derivative of $W_1(\boldsymbol{x},\boldsymbol{q},\boldsymbol{z})$ along the trajectories of (14) and $\zeta_i(t)=e^{-bt},t\geq 0$ satisfies

$$\begin{aligned}\dot W_2\leq&-\frac{1}{4}(\boldsymbol{q}-\bar{\boldsymbol{q}})^\top(L\otimes I_n)(\boldsymbol{q}-\bar{\boldsymbol{q}})\\&-\epsilon_1m_1\|\boldsymbol{x}-\boldsymbol{x}^*\|^2-\gamma_3\|\boldsymbol{\zeta}\|^2\\\leq&-\frac{\rho_2(L)}{4}(\boldsymbol{q}-\bar{\boldsymbol{q}})^\top(K_N\otimes I_n)(\boldsymbol{q}-\bar{\boldsymbol{q}})\\&-\epsilon_1m_1\|\boldsymbol{x}-\boldsymbol{x}^*\|^2-\gamma_3\|\boldsymbol{\zeta}\|^2\\\leq&-\epsilon_2[(\boldsymbol{q}-\bar{\boldsymbol{q}})^\top(K_N\otimes I_n)(\boldsymbol{q}-\bar{\boldsymbol{q}})\\&+\|\boldsymbol{x}-\boldsymbol{x}^*\|^2+\gamma_3\|\boldsymbol{\zeta}\|^2],\end{aligned}\tag{31}$$

where $\epsilon_2=\min\{\frac{\rho_2(L)}{4},\epsilon_1m_1,1\}>0$. Let $\epsilon_3=\max\{\frac{\epsilon_1+3}{2},\frac{1}{b}\}$, then

$$W_2\leq\epsilon_3[(\boldsymbol{q}-\bar{\boldsymbol{q}})^\top(K_N\otimes I_n)(\boldsymbol{q}-\bar{\boldsymbol{q}})+\|\boldsymbol{x}-\boldsymbol{x}^*\|^2+\gamma_3\|\boldsymbol{\zeta}\|^2] \tag{32}$$

Then, (31) and (32) yield $\dot W_2\leq-\frac{\epsilon_2}{\epsilon_3}W_2$. Thus, $W_2(t)\leq W_2(0)e^{-\frac{\epsilon_2}{\epsilon_3}t}$. Noting that $\|\boldsymbol{x}-\bar{\boldsymbol{x}}\|^2\leq\frac{2}{\epsilon_1}W_2$, we know that $x_i(t)$ exponentially converges to the unique global minimizer $x^*$ with a rate no less than $\frac{\epsilon_2}{2\epsilon_3}>0$.