Distributed convex optimization via continuous-time coordination algorithms with discrete-time communication ⋆

Solmaz S. Kia [a] Jorge Cortés [b] Sonia Martínez [b]

[a] *Department of Mechanical and Aerospace Engineering, University of California, Irvine*

[b] *Department of Mechanical and Aerospace Engineering, University of California, San Diego*

Abstract

This paper proposes a novel class of distributed continuous-time coordination algorithms to solve network optimization problems whose cost function is a sum of local cost functions associated to the individual agents. We establish the exponential convergence of the proposed algorithm under (i) strongly connected and weight-balanced digraph topologies when the local costs are strongly convex with globally Lipschitz gradients, and (ii) connected graph topologies when the local costs are strongly convex with locally Lipschitz gradients. When the local cost functions are convex and the global cost function is strictly convex, we establish asymptotic convergence under connected graph topologies. We also characterize the algorithm's correctness under time-varying interaction topologies and study its privacy preservation properties. Motivated by practical considerations, we analyze the algorithm implementation with discrete-time communication. We provide an upper bound on the stepsize that guarantees exponential convergence over connected graphs for implementations with periodic communication. Building on this result, we design a provably-correct centralized event-triggered communication scheme that is free of Zeno behavior. Finally, we develop a distributed, asynchronous event-triggered communication scheme that is also free of Zeno with asymptotic convergence guarantees. Several simulations illustrate our results.

Key words: cooperative control, distributed convex optimization, weight-balanced digraphs, event-triggered communication.

1 Introduction

An important class of distributed convex optimization problems consists of the (un-)constrained network optimization of a sum of convex functions, each one representing a local cost known to an individual agent. Examples include distributed parameter estimation [Ram et al., 2010, Wan and Lemmon, 2009], statistical learning [Boyd et al., 2010], and optimal resource allocation over networks [Madan and Lall, 2006]. To find the network optimizers, we propose a coordination model where each agent runs a purely local continuous-time evolution dynamics and communicates at discrete instants with its neighbors. We are motivated by the desire of combining the conceptual ease of the analysis of continuous-time dynamics and the practical constraints imposed by real-time implementations. Our design is based on a novel continuous-time distributed algorithm whose stability can be analyzed through standard Lyapunov functions.

Literature review: In distributed convex optimization, most coordination algorithms are time-varying, consensus-based dynamics [Boyd et al., 2010, Duchi et al., 2012, Johansson et al., 2009, Nedić and Ozdaglar, 2009, Zhu and Martínez, 2012] implemented in discrete time. Recent work [Gharesifard and Cortés, 2014, Lu and Tang, 2012, Wang and Elia, 2011, Zanella et al., 2011] has introduced continuous-time dynamical solvers whose convergence properties can be analyzed via classical stability analysis. This has the added advantage of facilitating the characterization of properties such as speed of convergence, disturbance rejection, and robustness to uncertainty. Wang and Elia [2011] establish asymptotic convergence under connected graphs and Gharesifard and Cortés [2014] extend the design and analysis to strongly connected, weight-balanced digraphs. The continuous-time algorithms in [Lu and Tang, 2012, Zanella et al., 2011] require twice-differentiable, strictly convex local cost functions to make use of the inverse of their Hessian and need a careful initialization to guarantee asymptotic convergence under undirected connected graphs. The novel class of continuous-time algorithms proposed here (upon which our implementations with discrete-time communication are built) do not suffer from the limitations discussed above and have, under some regularity

⋆ Corresponding author: S.S. Kia

Email addresses: solmaz@uci.edu (Solmaz S. Kia), cortes@ucsd.edu (Jorge Cortés), soniamd@ucsd.edu (Sonia Martínez).

assumptions, exponential convergence guarantees. Our work here also touches, albeit slightly, on the concept of privacy preservation, see e.g., [Weeraddana et al., 2013, Yan et al., 2013] for recent works in the context of distributed multi-agent optimization. Regarding event-triggered control of networked systems, recent years have seen an increasing body of work that seeks to trade computation and decision making for less communication, sensing or actuator effort, see e.g. [Heemels et al., 2012, Mazo and Tabuada, 2011, Wang and Lemmon, 2011]. Closest to the problem considered here are works that study event-triggered communication laws for average consensus, see e.g., [Dimarogonas et al., 2012, Garcia et al., 2013, Nowzari and Cortés, 2014]. The strategies proposed in [Wan and Lemmon, 2009] save communication effort in discrete-time implementations by using local triggering events but are not guaranteed to avoid Zeno behavior, i.e., an infinite number of triggered events in a finite period of time. Our goal is to combine the best of both approaches by synthesizing provably-correct continuous-time distributed dynamical systems which only require communication with neighbors at discrete instants of time. We are particularly interested in the opportunistic determination of this communication times via event triggering schemes.

Statement of contributions: We propose a novel class of continuous-time, gradient-based distributed algorithms for network optimization where the global objective function is the sum of local cost functions, one per agent. We prove that these algorithms converge exponentially under strongly connected and weight-balanced agent interactions when the local cost functions are strongly convex and their gradients are globally Lipschitz. Under connected, undirected graphs, we establish exponential convergence when the local gradients are just locally Lipschitz, and asymptotic convergence when the local cost functions are simply convex and the global cost function is strictly convex. We also study convergence under networks with time-varying topologies and characterize the topological requirements on the communication graph, algorithm parameters, and initial conditions necessary for an agent to reconstruct the local gradients of local cost functions of other agents. Our technical approach builds on the identification of strict Lyapunov functions. The availability of these functions enable our ensuing design of provably-correct continuous-time implementations with discrete-time communication. In particular, for networks with connected graph topologies, we obtain an upper bound on the suitable stepsizes that guarantee exponential convergence under periodic communication. Building on this result, we design a centralized, synchronous event-triggered communication scheme with an exponential convergence guarantees and Zeno-free behavior. Finally, we develop a Zeno-free asynchronous event-triggered communication scheme whose execution only requires agents to interchange information with their neighbors and establish its exponential convergence to a neighborhood of the network optimizer. Several simulations illustrate our results.

2 Preliminaries

In this section, we introduce our notation and some basic concepts from convex functions and graph theory. Let $\mathbb{R}$ and $\mathbb{Z}_{\geq 0}$ denote, respectively, the set of real and non-negative integer numbers. We use $\Re(\cdot)$ to represent the real part of a complex number. The transpose of a matrix $\mathbf{A}$ is $\mathbf{A}^\top$. We let $\mathbf{1}_n$ (resp. $\mathbf{0}_n$) denote the vector of n ones (resp. n zeros), and denote by I_n the $n \times n$ identity matrix. We let $\mathbf{\Pi}_n = \mathsf{I}_n - \frac{1}{n}\mathbf{1}_n\mathbf{1}_n^\top$. When clear from the context, we do not specify the matrix dimensions. For $\mathbf{A} \in \mathbb{R}^{n\times m}$ and $\mathbf{B} \in \mathbb{R}^{p\times q}$, we let $\mathbf{A}\otimes\mathbf{B}$ denote their Kronecker product. For $\mathbf{u} \in \mathbb{R}^d$, $\|\mathbf{u}\| = \sqrt{\mathbf{u}^\top\mathbf{u}}$ denotes the standard Euclidean norm. For vectors $\mathbf{u}_1, \cdots, \mathbf{u}_m$, $\mathbf{u} = (\mathbf{u}_1, \cdots, \mathbf{u}_m)$ is the aggregated vector. In a networked system, we distinguish the local variables of each agent by a superscript, e.g., $\mathbf{x}^i$ is the state of agent i. If $\mathbf{p}^i \in \mathbb{R}^d$ is a variable of agent i, the aggregated $\mathbf{p}^i$'s of the network of N agents is $\mathbf{p} = (\mathbf{p}^1, \cdots, \mathbf{p}^N) \in (\mathbb{R}^d)^N$. For convenience, we use $\mathsf{r} \in \mathbb{R}^N$ and $\mathsf{R} \in \mathbb{R}^{N\times(N-1)}$,

$$\mathsf{r} = \frac{1}{\sqrt{N}}\mathbf{1}_N, \ \mathsf{r}^\top\mathsf{R} = \mathbf{0}, \ \mathsf{R}^\top\mathsf{R} = \mathsf{I}_{N-1}, \ \mathsf{R}\mathsf{R}^\top = \mathbf{\Pi}_N. \quad (1)$$

A differentiable function $f : \mathbb{R}^d \to \mathbb{R}$ is *strictly convex* over a convex set $C \subseteq \mathbb{R}^d$ iff $(\mathbf{z}-\mathbf{x})^\top(\nabla f(\mathbf{z}) - \nabla f(\mathbf{x})) > 0$ for all $\mathbf{x}, \mathbf{z} \in C$ and $\mathbf{x} \neq \mathbf{z}$, and it is *$m$-strongly convex* ($m > 0$) iff $(\mathbf{z}-\mathbf{x})^\top(\nabla f(\mathbf{z}) - \nabla f(\mathbf{x})) \geq m\|\mathbf{z}-\mathbf{x}\|^2$, for all $\mathbf{x}, \mathbf{z} \in C$. A function $\mathbf{f} : \mathbb{R}^d \to \mathbb{R}^d$ is Lipschitz with constant $M > 0$, or simply *M-Lipschitz*, over a set $C \subset \mathbb{R}^d$ iff $\|\mathbf{f}(\mathbf{x}) - \mathbf{f}(\mathbf{y})\| \leq M\,\|\mathbf{x}-\mathbf{y}\|$, for $\mathbf{x}, \mathbf{y} \in C$. For a convex function f with M-Lipschitz gradient, one has $\|\nabla f(\mathbf{z}) - \nabla f(\mathbf{x})\|^2 \leq M(\mathbf{z}-\mathbf{x})^\top(\nabla f(\mathbf{z}) - \nabla f(\mathbf{x}))$ for all $\mathbf{x}, \mathbf{z} \in C \subset \mathbb{R}^d$.

We briefly review basic concepts from algebraic graph theory following [Bullo et al., 2009]. A *digraph*, is a pair $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, where $\mathcal{V} = \{1, \ldots, N\}$ is the *node set* and $\mathcal{E} \subseteq \mathcal{V}\times\mathcal{V}$ is the *edge set*. An edge from i to j, denoted by (i, j), means that agent j can send information to agent i. For an edge $(i, j) \in \mathcal{E}$, i is called an *in-neighbor* of j and j is called an *out-neighbor* of i. A graph is *undirected* if $(i, j) \in \mathcal{E}$ anytime $(j, i) \in \mathcal{E}$. A *directed path* is a sequence of nodes connected by edges. A digraph is *strongly connected* if for every pair of nodes there is a directed path connecting them. A *weighted digraph* is a triplet $\mathcal{G} = (\mathcal{V}, \mathcal{E}, \mathsf{A})$, where $(\mathcal{V}, \mathcal{E})$ is a digraph and $\mathsf{A} \in \mathbb{R}^{N\times N}$ is a weighted *adjacency* matrix such that $\mathsf{a}_{ij} > 0$ if $(i, j) \in \mathcal{E}$ and $\mathsf{a}_{ij} = 0$, otherwise. A weighted digraph is *undirected* if $\mathsf{a}_{ij} = \mathsf{a}_{ji}$ for all $i, j \in \mathcal{V}$. We refer to a strongly connected and undirected graph as a *connected graph*. The *weighted in-* and *out-degrees* of a node i are, respectively, $\mathsf{d}^i_{\text{in}} = \Sigma_{j=1}^N \mathsf{a}_{ji}$ and $\mathsf{d}^i_{\text{out}} = \Sigma_{j=1}^N \mathsf{a}_{ij}$. A digraph is *weight-balanced* if at each node $i \in \mathcal{V}$, the weighted out-degree and weighted in-degree coincide (although they might be different across different nodes). Any undirected graph is weight-balanced. The *(out-) Laplacian* matrix is $\mathsf{L} = \mathsf{D}^{\text{out}} - \mathsf{A}$, where $\mathsf{D}^{\text{out}} = \text{Diag}(\mathsf{d}^1_{\text{out}}, \cdots, \mathsf{d}^N_{\text{out}}) \in \mathbb{R}^{N\times N}$. Note that $\mathsf{L}\mathbf{1}_N = \mathbf{0}$. A digraph is weight-balanced iff $\mathbf{1}_N^T\mathsf{L} = \mathbf{0}$ iff $\text{Sym}(\mathsf{L}) =$

$(\mathsf{L}+\mathsf{L})/2$ is positive semi-definite. Based on the structure of L, at least one of the eigenvalues of L is zero and the rest of them have nonnegative real parts. We denote the eigenvalues of L by $\lambda_1,\dots,\lambda_N$, where $\lambda_1 = 0$ and $\Re(\lambda_i) \leq \Re(\lambda_j)$, for $i<j$, and the eigenvalues of $\mathrm{Sym}(\mathsf{L})$ by $\hat{\lambda}_1,\dots,\hat{\lambda}_N$. For a strongly connected and weight-balanced digraph, zero is a simple eigenvalue of both L and $\mathrm{Sym}(\mathsf{L})$. In this case, we order the eigenvalues of $\mathrm{Sym}(\mathsf{L})$ as $\hat{\lambda}_1 = 0 < \hat{\lambda}_2 \leq \hat{\lambda}_3 \leq \cdots \leq \hat{\lambda}_N$ and we have

$$0 < \hat{\lambda}_2 \mathbf{I} \leq \mathsf{R}^\top \mathrm{Sym}(\mathsf{L}) \mathsf{R} \leq \hat{\lambda}_N \mathbf{I}. \qquad (2)$$

Notice that for connected graphs $\hat{\lambda}_i = \lambda_i$ for $i \in \{1,\cdots,N\}$. For convenience, we define $\mathbf{L} = \mathsf{L} \otimes \mathsf{I}_d$ and $\mathbf{\Pi} = \mathsf{\Pi}_N \otimes \mathsf{I}_d$ to deal with variables of dimension $d \in \mathbb{N}$.

3 Problem Definition

Consider a network of N agents interacting over a strongly connected and weight-balanced digraph $\mathcal{G}$. Each agent $i \in \mathcal{V}$ has a differentiable local cost function $f^i : \mathbb{R}^d \to \mathbb{R}$. The global cost function of the network, which we assume to be strictly convex, is $f = \Sigma_{i=1}^N f^i(\mathbf{x})$. Our aim is to design a distributed algorithm such that each agent solves

$$\min_{\mathbf{x}\in\mathbb{R}^d} f(\mathbf{x})$$

using only its own local data and exchanged information with its neighbors. We assume the above optimization problem is feasible (which, together with the strict convexity of f, implies the uniqueness of the global optimizer, which we denote $\mathbf{x}^\star \in \mathbb{R}^d$). We are also interested in characterizing the privacy preservation properties of the algorithmic solution to this distributed optimization problem. Specifically, we aim to identify conditions guaranteeing that no information about the local cost function of an agent is revealed to, or can be reconstructed by, any other agent in the network.

4 Distributed Continuous-Time Algorithm for Convex Optimization

In this section, we provide a novel continuous-time distributed coordination algorithm to solve the problem stated in Section 3 and analyze in detail its convergence properties. For $i \in \mathcal{V}$ and with $\alpha, \beta > 0$, consider

$$\dot{\mathbf{v}}^i = \alpha\beta \sum_{j=1}^N \mathsf{a}_{ij}(\mathbf{x}^i - \mathbf{x}^j), \qquad (3a)$$

$$\dot{\mathbf{x}}^i = -\alpha \nabla f^i(\mathbf{x}^i) - \beta \sum_{j=1}^N \mathsf{a}_{ij}(\mathbf{x}^i - \mathbf{x}^j) - \mathbf{v}^i, \qquad (3b)$$

In network variables $\mathbf{x},\mathbf{v} \in (\mathbb{R}^d)^N$, the algorithm reads as

$$\dot{\mathbf{v}} = \alpha\beta \mathbf{L}\mathbf{x}, \qquad (4a)$$

$$\dot{\mathbf{x}} = -\alpha \nabla \tilde{f}(\mathbf{x}) - \beta \mathbf{L}\mathbf{x} - \mathbf{v}. \qquad (4b)$$

Here, $\tilde{f} : (\mathbb{R}^d)^N \to \mathbb{R}$ is defined by $\tilde{f}(\mathbf{x}) = \Sigma_{i=1}^N f^i(\mathbf{x}^i)$. This algorithm is distributed because each agent only needs to receive information from its out-neighbors about their corresponding variables in $\mathbf{x}$. In contrast, the continuous-time coordination algorithms in [Gharesifard and Cortés, 2014, Wang and Elia, 2011] require the communication of the corresponding variables in both $\mathbf{x}$ and $\mathbf{v}$. The synthesis of algorithm (4) is inspired by the following feedback control considerations. In (4b), each agent follows a local gradient descent while trying to agree with its neighbors on their estimate of the final value. However, as the local gradients are not the same, this dynamics by itself would never converge. Therefore, to correct this error, each agent uses an integral feedback term $\mathbf{v}^i$ whose evolution is driven by the agent disagreement according to (4a).

Our analysis of the algorithm convergence is structured in two parts, depending on the directed character of the interactions. Section 4.1 deals with strongly connected and weight-balanced digraphs and Section 4.2 deals with connected graphs. In each case, we identify conditions on the agent cost functions that guarantee asymptotic convergence. Given the challenges posed by directed information flows, it is not surprising that we can establish stronger results under less restrictive assumptions for the case of undirected topologies.

4.1 Strongly Connected, Weight-Balanced Digraphs

Here, we study the convergence of the distributed optimization algorithm (3) over strongly connected and weight-balanced digraph topologies. We first consider the case where the interaction topology is fixed, and then discuss the time-varying interaction topologies. The following result identifies conditions on the local cost functions $\{f^i\}_{i=1}^N$ and the parameter β to guarantee the exponential convergence of (3) to the solution of the distributed optimization problem.

Theorem 1 (Convergence of (3) *over strongly connected and weight-balanced digraphs): Let $\mathcal{G}$ be a strongly connected and weight-balanced digraph. Assume each f^i, $i \in \mathcal{V}$, is m^i-strongly convex, differentiable, and its gradient is M^i-Lipschitz on $\mathbb{R}^d$. Given $\alpha > 0$, $\underline{m} = \min\{m^1,\dots,m^N\}$ and $\overline{M} = \max\{M^1,\dots,M^N\}$, let $\beta,\phi > 0$ satisfy $\phi + 1 > 4\overline{M}$ and*

$$\gamma = \alpha^2(\phi+1)\underline{m} + 9\beta\hat{\lambda}_2\phi\alpha - 4\alpha^2(\overline{M}\underline{m} + (\phi+1)^2) > 0, \qquad (5)$$

Then, for each $i \in \mathcal{V}$, starting from $\mathbf{x}^i(0), \mathbf{v}^i(0) \in \mathbb{R}^d$ with $\sum_{i=1}^N \mathbf{v}^i(0) = \mathbf{0}_d$, the algorithm (3) *over $\mathcal{G}$ makes $\mathbf{x}^i(t) \to \mathbf{x}^\star$ exponentially fast as $t \to \infty$ with a rate no less than*

$$\min\{\frac{7}{16}, \frac{1}{9}\gamma\}/(2\bar{\lambda}_{\mathbf{F}}). \qquad (6)$$

Here, $\bar{\lambda}_{\mathbf{F}}$ is the maximum eigenvalue of

$$\mathbf{F} = \frac{1}{2}\begin{bmatrix} \frac{1}{9}\alpha(\phi+1)\mathsf{I}_d & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \alpha(\phi+1)\mathsf{I}_{(N-1)d} & \mathsf{I}_{(N-1)d} \\ \mathbf{0} & \mathsf{I}_{(N-1)d} & \frac{1}{\alpha}\mathsf{I}_{(N-1)d} \end{bmatrix}. \qquad (7)$$

PROOF. For weight-balanced digraphs, we have $\mathbf{1}_N^\top \mathsf{L} = \mathbf{0}$. Thus, left multiplying (4a) by $\mathbf{1}_N^\top \otimes \mathsf{I}_d$ results in

$$\sum_{i=1}^N \dot{\mathbf{v}}^i = \mathbf{0} \Rightarrow \sum_{i=1}^N \mathbf{v}^i(t) = \sum_{i=1}^N \mathbf{v}^i(0) = \mathbf{0},\ \forall t \geq 0. \qquad (8)$$

Next, we obtain the equilibrium point of (3), $(\bar{\mathbf{x}}, \bar{\mathbf{v}})$, from

$$\mathbf{0} = \alpha\beta\mathbf{L}\bar{\mathbf{x}}, \tag{9a}$$
$$\mathbf{0} = -\alpha\nabla\tilde{f}(\bar{\mathbf{x}}) - \beta\mathbf{L}\bar{\mathbf{x}} - \bar{\mathbf{v}}. \tag{9b}$$

Given (9a), $\bar{\mathbf{x}}$ belongs to the null-space of $\mathbf{L}$. For strongly connected digraphs the null-space of $\mathbf{L}$ is spanned by $\mathbf{1}_N$. Thus, $\bar{\mathbf{x}} = \mathbf{1}_N \otimes \boldsymbol{\theta}$, where $\boldsymbol{\theta} \in \mathbb{R}^d$. Left multiplying (9b) by $\mathbf{1}_N^\top \otimes \mathbf{I}_d$ and using (8), we obtain $\mathbf{0} = \sum_{i=1}^N \nabla f^i(\bar{\mathbf{x}}^i)$. Then, the optimality condition $\nabla f(\mathbf{x}^\star) = \mathbf{0}_d$ along with $\nabla f(\boldsymbol{\theta}) = \sum_{i=1}^N \nabla f^i(\boldsymbol{\theta})$ imply

$$\bar{\mathbf{x}}^i = \mathbf{x}^\star, \quad i \in \mathcal{V}.$$

Substituting this value in (9b), we obtain

$$\bar{\mathbf{v}}^i = -\alpha\nabla f^i(\mathbf{x}^\star), \quad i \in \mathcal{V}. \tag{10}$$

To study the stability of (3), we transfer the equilibrium point to the origin and then apply a change of variables

$$\mathbf{u} = \mathbf{v} - \bar{\mathbf{v}}, \quad \mathbf{y} = \mathbf{x} - \bar{\mathbf{x}}, \tag{11a}$$
$$\mathbf{u} = ([\mathbf{r} \quad \mathbf{R}] \otimes \mathbf{I}_d)\mathbf{w}, \quad \mathbf{y} = ([\mathbf{r} \quad \mathbf{R}] \otimes \mathbf{I}_d)\mathbf{z}, \tag{11b}$$

where we used (1). We partition the new variables as follows: $\mathbf{w} = (\mathbf{w}_1, \mathbf{w}_{2:N})$ and $\mathbf{z} = (\mathbf{z}_1, \mathbf{z}_{2:N})$, where $\mathbf{w}_1, \mathbf{z}_1 \in \mathbb{R}^d$. In these new variables, the algorithm (3) reads as

$$\begin{aligned}
\dot{\mathbf{w}}_1 &= \mathbf{0}_d, \\
\dot{\mathbf{w}}_{2:N} &= \alpha\beta(\mathbf{R}^\top\mathbf{L}\mathbf{R} \otimes \mathbf{I}_d)\mathbf{z}_{2:N}, \\
\dot{\mathbf{z}}_1 &= -\alpha(\mathbf{r}^\top \otimes \mathbf{I}_d)\mathbf{h}, \\
\dot{\mathbf{z}}_{2:N} &= -\alpha(\mathbf{R}^\top \otimes \mathbf{I}_d)\mathbf{h} - \beta(\mathbf{R}^\top\mathbf{L}\mathbf{R} \otimes \mathbf{I}_d)\mathbf{z}_{2:N} - \mathbf{w}_{2:N},
\end{aligned} \tag{12}$$

where

$$\mathbf{h} = \nabla\tilde{f}(\mathbf{y} + \bar{\mathbf{x}}) - \nabla\tilde{f}(\bar{\mathbf{x}}). \tag{13}$$

Note that the first equation in (12) corresponds to the constant of motion (8). To study the stability in the other variables, consider the candidate Lyapunov function

$$\begin{aligned}
V(\mathbf{z}, \mathbf{w}_{2:N}) = &\frac{1}{18}\alpha(\phi+1)\mathbf{z}_1{}^\top\mathbf{z}_1 + \frac{\phi\alpha}{2}\mathbf{z}_{2:N}{}^\top\mathbf{z}_{2:N} \\
&+ \frac{1}{2\alpha}(\alpha\mathbf{z}_{2:N} + \mathbf{w}_{2:N})^\top(\alpha\mathbf{z}_{2:N} + \mathbf{w}_{2:N}),
\end{aligned} \tag{14}$$

with $\phi > 0$ as in the statement. Note that $V \leq \bar{\lambda}_{\mathbf{F}}\|\mathbf{p}\|^2$, with $\mathbf{p} = (\mathbf{z}, \mathbf{w}_{2:N})$. The Lie derivative of V along (12) is

$$\begin{aligned}
\dot{V} = &-\frac{1}{9}\alpha^2(\phi+1)\mathbf{y}^\top\mathbf{h} - \frac{7}{16}\mathbf{w}_{2:N}^\top\mathbf{w}_{2:N} \\
&- \phi\alpha\beta\mathbf{z}_{2:N}{}^\top(\mathbf{R}^\top\,\mathrm{Sym}(\mathbf{L})\mathbf{R} \otimes \mathbf{I}_d)\mathbf{z}_{2:N} \\
&+ \frac{4}{9}\alpha^2\|(\mathbf{R}^\top \otimes \mathbf{I}_d)\mathbf{h}\|^2 + \frac{4}{9}\alpha^2(1+\phi)^2\mathbf{z}_{2:N}^\top\mathbf{z}_{2:N} \\
&- \|\frac{3}{4}\mathbf{w}_{2:N} + \frac{2\alpha}{3}(\mathbf{R}^\top \otimes \mathbf{I}_d)\mathbf{h} + \frac{2\alpha}{3}(\phi+1)\mathbf{z}_{2:N})\|^2.
\end{aligned}$$

Next, we show that under $\phi + 1 > 4\overline{M}$ and (5), $\dot{V}$ is negative definite. Invoking the assumptions on the local cost functions in the statement, and using the $\overline{M}$-Lipschitzness of $\nabla\tilde{f}$ and the $\underline{m}$-strongly convexity of $\tilde{f}$ along with $\|\mathbf{R}^\top \otimes \mathbf{I}_d\| = 1$ and $\|\mathbf{z}\| = \|\mathbf{y}\|$, we have

$$\|(\mathbf{R}^\top \otimes \mathbf{I}_d)\mathbf{h}\|^2 \leq \|\mathbf{h}\|^2 \leq \overline{M}\mathbf{y}^\top\mathbf{h}, \tag{15a}$$
$$\mathbf{y}^\top\mathbf{h} \geq \underline{m}\|\mathbf{y}\|^2 = \underline{m}\|\mathbf{z}\|^2. \tag{15b}$$

Given these relations and invoking (2), we have

$$\begin{aligned}
\dot{V} \leq &-\frac{\alpha^2((\phi+1) - 4\overline{M})\underline{m}}{9}\mathbf{z}^\top\mathbf{z} - \frac{7}{16}\mathbf{w}_{2:N}^\top\mathbf{w}_{2:N} \\
&- \phi\alpha\beta\hat{\lambda}_2\mathbf{z}_{2:N}^\top\mathbf{z}_{2:N} + \frac{4\alpha^2(1+\phi)^2}{9}\mathbf{z}_{2:N}^\top\mathbf{z}_{2:N} \\
&- \|\frac{3}{4}\mathbf{w}_{2:N} + \frac{2\alpha}{3}(\mathbf{R}^\top \otimes \mathbf{I}_d)\mathbf{h} + \frac{2\alpha(\phi+1)}{3}\mathbf{z}_{2:N})\|^2.
\end{aligned}$$

Since $\mathbf{z}^\top\mathbf{z} = \mathbf{z}_1^\top\mathbf{z}_1 + \mathbf{z}_{2:N}^\top\mathbf{z}_{2:N}$, it follows that $\dot{V} < -\min\{\frac{7}{16}, \frac{1}{9}\gamma\}\|\mathbf{p}\|^2 < 0$, where γ is a shorthand notation for the expression in (5). Thus, $\mathbf{z} \to \mathbf{0}$ as $t \to \infty$, equivalently $\mathbf{x}^i \to \mathbf{x}^\star$, for all $i \in \mathcal{V}$, is exponential with rate no less than (6) (cf. [Khalil, 2002, Theorem 4.10]). □

In Theorem 1, the requirement $\sum_{i=1}^N \mathbf{v}^{\mathbf{i}}(0) = \mathbf{0}_d$ is trivially satisfied by each agent with $\mathbf{v}^i(0) = \mathbf{0}_d$. This is an advantage with respect to the continuous-time coordination algorithms proposed in [Lu and Tang, 2012], which requires the nontrivial initialization $\sum_{i=1}^N \nabla f^i(\mathbf{x}^i(0)) = \mathbf{0}_d$, and in [Zanella et al., 2011], which requires the initialization on a state communicated among neighbors and is subject to communication error.

Remark 2 *(Design parameters in* (3)*):* We provide here several observations on the role of the design parameters α and β. First, note there always exist α, β satisfying (5), e.g., any $\beta > 4(\phi+1)^2\alpha/(9\phi\hat{\lambda}_2)$. The determination of these parameters can be performed by individual agents if they know an upper bound on $\overline{M}$, a lower bound on $\underline{m}$, and have knowledge of $\hat{\lambda}_2$, either through a dedicated algorithm to compute it, see e.g., Yang et al. [2010], or use a lower bound on it, see e.g. Mohar [1991]. We have observed in simulation that (5) is only sufficient and that, in fact, the algorithm (3) converges for any positive α and β in our numerical examples. Although not evident in (6), one can expect the larger α and β are, the higher the rate of convergence of the algorithm (3) is. A coefficient $\alpha > 1$ can be interpreted as a way of increasing the strong convexity coefficient of the local cost functions. A coefficient $\beta > 1$ can be interpreted as a means of increasing the graph connectivity. The relationship between these parameters and the rate of convergence of (3) is more evident for quadratic local cost functions $f^i(\mathbf{x}) = \frac{1}{2}(\mathbf{x}^\top\mathbf{x} + \mathbf{x}^\top\mathbf{a}^i + \mathbf{b}^i)$, $i \in \mathcal{V}$. In this case, the algorithm (3) is a linear time-invariant system where the eigenvalues of the system matrix are $-\alpha$, with multiplicity of Nd, and $-\beta\lambda_i$, $i \in \mathcal{V}$ (λ_i's are the eigenvalues of $\mathbf{L}$), with multiplicity d. Therefore, (3) converges regardless of the value of α, $\beta > 0$ with an exponential rate equal to $\min\{\alpha, \beta\Re(\lambda_2)\}$. When discussing discrete-time communication, some trade-offs arise regarding the choice of the parameters, as we explain later in Section 5. •

Remark 3 *(Semiglobal convergence of* (3) *under local gradients that are locally Lipschitz):* The convergence result in Theorem 1 is semiglobal [Khalil, 2002] if the local gradients are only locally Lipschitz or, equivalently, Lipschitz on compact sets. In fact, one can see from the proof of the result that, for any compact set containing the initial conditions $\mathbf{x}^i(0) \in \mathbb{R}^d$ and $\mathbf{v}^i(0) = \mathbf{0}_d$, $i \in \mathcal{V}$, one can find $\phi > 0$ and $\beta > 0$ sufficiently large such that the compact set is contained in the region of attraction of the equilibrium point. •

Next, we study the convergence of (3) over dynamically changing topologies. Since the proof of Theorem 1 relies on a Lyapunov function with no dependency on the system parameters and its derivative is upper bounded by a quadratic negative definite function, we can readily extend the convergence result to dynamically changing networks. The proof details are omitted for brevity.

Proposition 4 (Convergence of (3) *over dynamically changing interaction topologies): Let $\mathcal{G}$ be a time-varying digraph which is strongly connected and weight-balanced at all times and whose adjacency matrix is uniformly bounded and piecewise constant. Assume the local cost function f^i, $i \in \mathcal{V}$, is m^i-strongly convex, differentiable, and its gradient is M^i-Lipschitz on $\mathbb{R}^d$. Given $\alpha > 0$, let $\beta, \phi > 0$ satisfy $\phi + 1 > 4\overline{M}$ and* (5) *with $\hat{\lambda}_2$ replaced by $(\hat{\lambda}_2)_{\min} = \min_{p\in\mathcal{P}}\{\hat{\lambda}_2(\mathbf{L}_p)\}$, where $\mathcal{P}$ is the index set of all possible realizations of $\mathcal{G}$. Then, for each $i \in \mathcal{V}$, starting from $\mathbf{x}^i(0), \mathbf{v}^i(0) \in \mathbb{R}^d$ with $\sum_{i=1}^N \mathbf{v}^i(0) = \mathbf{0}_d$, the algorithm* (3) *over $\mathcal{G}$ makes $\mathbf{x}^i(t) \to \mathbf{x}^\star$ exponentially fast as $t \to \infty$.*

Our final result of this section characterizes the topological requirements on the communication graph and the knowledge about the algorithm's parameters and initial conditions that allow a passive agent (i.e., an agent that does not interfere in the algorithm execution) to reconstruct the local gradients of other agents in the network.

Proposition 5 (Privacy preservation under (3)*): Let $\mathcal{G}$ be a strongly connected and weight balanced digraph. For $\alpha, \beta > 0$, consider any execution of the coordination algorithm* (3) *over $\mathcal{G}$ starting from $\mathbf{x}^i(0), \mathbf{v}^i(0) \in \mathbb{R}^d$ with $\sum_{i=1}^N \mathbf{v}^i(0) = \mathbf{0}_d$. Then, an agent $i \in \mathcal{V}$ can reconstruct the local gradient of another agent $j \neq i$ only if j and all its out-neighbors are out-neighbors of i, and agent i knows $\mathbf{v}^j(0)$ and a_{jk}, $k \in \mathcal{V}$ (here we assume that the agent i is aware of the identity of neighbors of agent j and it has memory to save the time history of the data it receives from its out-neighbors).*

PROOF. Consider an arbitrary time $t^\star$. Let i be an in-neighbor of agent j and all of its out-neighbors. The algorithm (3) requires each agent to communicate its component of $\mathbf{x}$ to their in-neighbors. Since agent i has memory to save information it receives from its out-neighbors for all $t \leq t^\star$, it can use the time history of $\mathbf{x}^j(t)$ to numerically reconstruct $\dot{\mathbf{x}}^j(t^\star)$. Because i is the in-neighbor of j and its out-neighbors, it can use its knowledge of a_{jk}, $k \in \mathcal{V}$ to reconstruct $\sum_{k=1}^N a_{jk}(\mathbf{x}^j(t) - \mathbf{x}^k(t))$ for all $t \leq t^\star$. Agent i can reconstruct $\mathbf{v}^j(t)$ from (3a) uniquely as it knows $\mathbf{v}^j(0)$. Then, agent i has all the elements to solve for $\nabla f^j(\mathbf{x}^j(t^\star))$ in (3b). The lack of knowledge about any of this information would prevent i from reconstructing exactly the local gradient of j. □

4.2 Connected Graphs

Here, we study the convergence of the algorithm (3) over connected graph topologies. While the results of the previous section are of course valid for these topologies, here, using the structural properties of the Laplacian matrix of undirected graphs, we establish the convergence of (3) for a larger family of local cost functions. In doing so, we are also able to analytically establish convergence for any $\alpha, \beta > 0$, as we show next.

Theorem 6 (Exponential convergence of (3) *over connected graphs): Let $\mathcal{G}$ be a connected graph. Assume the local cost function f^i, $i \in \mathcal{V}$, is m^i-strongly convex and differentiable on $\mathbb{R}^d$, and its gradient is locally Lipschitz. Then, for any $\alpha, \beta > 0$ and each $i \in \mathcal{V}$, starting from $\mathbf{x}^i(0), \mathbf{v}^i(0) \in \mathbb{R}^d$ with $\sum_{i=1}^N \mathbf{v}^i(0) = \mathbf{0}$, the algorithm* (3) *over $\mathcal{G}$ satisfies $\mathbf{x}^i(t) \to \mathbf{x}^\star$ as $t \to \infty$ exponentially fast.*

PROOF. We use the equivalent representation (12) of the algorithm (3) obtained in the proof of Theorem 1. Consider the following candidate Lyapunov function

$$\begin{aligned} V(\mathbf{z}, \mathbf{w}_{2:N}) = &\frac{1}{2}\alpha(\phi+1)\mathbf{z}_1^\top \mathbf{z}_1 + \frac{\phi\alpha}{2}\mathbf{z}_{2:N}^\top \mathbf{z}_{2:N} \\ &+ \frac{1}{2\alpha}(\alpha\mathbf{z}_{2:N} + \mathbf{w}_{2:N})^\top(\alpha\mathbf{z}_{2:N} + \mathbf{w}_{2:N}) \\ &+ \frac{1}{2\beta}(\phi+1)\mathbf{w}_{2:N}^\top\big((\mathbf{R}^\top\mathbf{L}\mathbf{R})^{-1} \otimes \mathbf{I}_d\big)\mathbf{w}_{2:N}, \end{aligned} \tag{16}$$

with $\phi \geq 1$ defined below. Given (2), V is positive definite and radially unbounded and satisfies $V \leq \bar{\lambda}_{\mathbf{E}}\|\mathbf{p}\|^2$, with $\mathbf{p} = (\mathbf{z}, \mathbf{w}_{2:N})$ and $\bar{\lambda}_{\mathbf{E}} > 0$ is the maximum eigenvalue of

$$\mathbf{E} = \frac{1}{2}\begin{bmatrix} \alpha(\phi+1)\mathbf{I}_d & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \alpha(\phi+1)\mathbf{I}_{(N-1)d} & \mathbf{I}_{(N-1)d} \\ \mathbf{0} & \mathbf{I}_{(N-1)d} & \frac{1}{\alpha}\mathbf{I} + \frac{(\phi+1)}{\beta}(\mathbf{R}^\top\mathbf{L}\mathbf{R})^{-1}\otimes\mathbf{I}_d \end{bmatrix}.$$

The Lie derivative of V along the dynamics (12) is

$$\begin{aligned} \dot{V} = &-\alpha^2(\phi+1)\mathbf{y}^\top\mathbf{h} - \phi\alpha\beta{\mathbf{z}_{2:N}}^\top(\mathbf{R}^\top\mathbf{L}\mathbf{R}\otimes\mathbf{I}_d)\mathbf{z}_{2:N} \\ &- {\mathbf{w}_{2:N}}^\top\mathbf{w}_{2:N} - \alpha{\mathbf{w}_{2:N}}^\top(\mathbf{R}^\top\otimes\mathbf{I}_d)\mathbf{h}. \end{aligned}$$

Next, we show that $\dot{V}$ is upper bounded by a negative definite function. We start by identifying a compact set whose definition is independent of ϕ and contains the set $\mathcal{S}_0 = \{(\mathbf{z}, \mathbf{w}_{2:N}) \in \mathbb{R}^{Nd} \times \mathbb{R}^{(N-1)d} \mid V(\mathbf{z}, \mathbf{w}_{2:N}) \leq V(\mathbf{z}(0), \mathbf{w}_{2:N}(0))\}$. For any given initial condition, let $\rho_0 = \frac{1}{2}\alpha\mathbf{z}_1(0)^\top\mathbf{z}_1(0) + \frac{(\alpha+1)}{2}\mathbf{z}_{2:N}(0)^\top\mathbf{z}_{2:N}(0) + (\frac{1}{2\beta\lambda_2}+\frac{1}{2\alpha}+\frac{1}{2})\mathbf{w}_{2:N}(0)^\top\mathbf{w}_{2:N}(0)$ and define $\bar{\mathcal{S}}_0 = \{\mathbf{z}\in\mathbb{R}^{Nd} \mid \frac{1}{2}\alpha\mathbf{z}_1^\top\mathbf{z}_1 + \frac{1}{4}\alpha\mathbf{z}_{2:N}^\top\mathbf{z}_{2:N} \le \rho_0\}$. Observe that this set is compact. Note that $(\mathbf{z},\mathbf{w}_{2:N})\in\mathcal{S}_0$ implies $\mathbf{z}\in\bar{\mathcal{S}}_0$ because $V(\mathbf{z}(0),\mathbf{w}_{2:N}(0)) \le (\phi+1)\rho_0$ and

$$\frac{1}{2}\alpha(\phi+1)\mathbf{z}_1^\top\mathbf{z}_1 + \frac{1}{4}\alpha(\phi+1)\mathbf{z}_{2:N}^\top\mathbf{z}_{2:N} \le \frac{1}{2}\alpha(\phi+1)\mathbf{z}_1^\top\mathbf{z}_1 + \frac{1}{2}\alpha\phi\mathbf{z}_{2:N}^\top\mathbf{z}_{2:N} \le V(\mathbf{z},\mathbf{w}_{2:N}).$$

Here, we used $\phi\ge 1$ in the first inequality. Since the change of variables (11a) and (11b) are linear, the corresponding $\mathbf{x}$ and $\mathbf{y}$ for $\mathbf{z}\in\bar{\mathcal{S}}_0$ belong to compact sets, as well. Then, the assumption on the gradients of the local cost functions implies that there exists $M_0>0$ with $\|\mathbf{h}\| \le M_0\|\mathbf{y}\| = M_0\|\mathbf{z}\|$, for all $\mathbf{z}\in\bar{\mathcal{S}}_0$. Consequently,

$$\|(\mathbf{R}^\top\otimes\mathbf{I}_d)\mathbf{h}\|^2 \le M_0^2\|\mathbf{z}\|^2, \quad \forall(\mathbf{z},\mathbf{w}_{2:N})\in\mathcal{S}_0.$$

Then we can show $-\alpha{\mathbf{w}_{2:N}}^\top(\mathbf{R}^\top\otimes\mathbf{I}_d)\mathbf{h} \le \frac{1}{2}\mathbf{w}_{2:N}^\top\mathbf{w}_{2:N} + \frac{1}{2}\alpha^2M_0^2\mathbf{z}^\top\mathbf{z}$ for $(\mathbf{z},\mathbf{w}_{2:N})\in\mathcal{S}_0$. Using this inequality, fact that the local cost functions are m^i-strongly convex (and hence (15b) holds) and invoking (2) we deduce

$$\dot{V} \le -\alpha^2(\phi+1)\underline{m}(\mathbf{z}_1^\top\mathbf{z}_1 + \mathbf{z}_{2:N}^\top\mathbf{z}_{2:N}) - \phi\alpha\beta\lambda_2\mathbf{z}_{2:N}^\top\mathbf{z}_{2:N} - \frac{1}{2}\mathbf{w}_{2:N}^\top\mathbf{w}_{2:N} + \frac{1}{2}\alpha^2M_0^2(\mathbf{z}_1^\top\mathbf{z}_1 + \mathbf{z}_{2:N}^\top\mathbf{z}_{2:N}).$$

Let $\phi+1 = \frac{1}{2\underline{m}}M_0^2 + \frac{1}{2\underline{m}\alpha^2}\delta_0$, where $\delta_0>0$ is such that $\phi\ge1$ (since M_0 does not depend on ϕ, this choice is always feasible). Then, we have $\dot{V} \le -\frac{1}{2}\min\{1,\delta_0\}\|\mathbf{p}\|^2$. As such, the Lyapunov function (16) satisfies all the conditions of [Khalil, 2002, Theorem 4.10], for the dynamics (12). Therefore, the convergence of $\mathbf{z}\to\mathbf{0}$ as $t\to\infty$, equivalently $\mathbf{x}^i\to\mathbf{x}^\star$, for all $i\in\mathcal{V}$, is exponential. □

Remark 7 *(Bound on exponential rate of convergence):* One can see from the proof of Theorem 6 that, for a given initial condition, the rate of convergence is at least $\frac{1}{4}(\min\{1,\delta_0\})/\bar{\lambda}_{\mathbf{E}} > 0$. The guaranteed rate of convergence is therefore not uniform, unless the local gradients are globally Lipschitz. In this case, one recovers the result in Theorem 1 but for arbitrary α, $\beta>0$. •

Note that in Theorem 6 the local Lipschitzness of ∇f^i is trivially held if f^i is twice differentiable. The Lyapunov function (16) identified in the proof of this result plays a key role later in our study of the algorithm implementation with discrete-time communication in Section 5. Next, we study the convergence of the algorithm (3) over connected graphs when the local cost functions are only convex. Here, the lack of strong convexity makes us rely on a LaSalle function, rather than on a Lyapunov one, to establish asymptotic convergence to the optimizer.

Theorem 8 (Asymptotic convergence of (3) *over connected graphs): Let $\mathcal{G}$ be a connected graph. Assume the local cost function f^i, $i\in\mathcal{V}$, is convex and differentiable on $\mathbb{R}^d$, and the global cost function f is strictly convex. Then, for any $\alpha,\beta>0$ and each $i\in\mathcal{V}$, starting from $\mathbf{x}^i(0),\mathbf{v}^i(0)\in\mathbb{R}^d$ with $\sum_{i=1}^N\mathbf{v}^i(0)=\mathbf{0}_d$, the algorithm* (3) *over $\mathcal{G}$ make $\mathbf{x}^i(t)\to\mathbf{x}^\star$ as $t\to\infty$.*

PROOF. We use again the equivalent representation (12) of the algorithm (3) obtained in the proof of Theorem 1. To study the stability of this system, consider the following candidate Lyapunov function

$$V(\mathbf{z},\mathbf{w}_{2:N}) = \frac{1}{2}\mathbf{z}^\top\mathbf{z} + \frac{1}{2\alpha\beta}{\mathbf{w}_{2:N}}^\top\Big((\mathbf{R}^\top\mathbf{L}\mathbf{R})^{-1}\otimes\mathbf{I}_d\Big)\mathbf{w}_{2:N}.$$

Given (2), V is positive definite and radially unbounded. The Lie derivative of V along (12) is given by

$$\dot{V} = -\alpha\mathbf{y}^\top(\nabla_{\mathrm{T}}f(\mathbf{y}+\bar{\mathbf{x}}) - \nabla_{\mathrm{T}}f(\bar{\mathbf{x}})) - \beta\mathbf{y}^\top\mathbf{L}\mathbf{y} = -\alpha\sum_{i=1}^N{\mathbf{y}^i}^\top(\nabla f^i(\mathbf{y}^i+\mathbf{x}^\star) - \nabla f^i(\mathbf{x}^\star)) - \beta\mathbf{y}^\top\mathbf{L}\mathbf{y}.$$

To obtain the second summand, we have used $\mathbf{z}_{2:N} = (\mathbf{R}^\top\otimes\mathbf{I}_d)\mathbf{y}$ and $\mathbf{R}\mathbf{R}^\top = \mathbf{I}_N - \mathbf{r}\mathbf{r}^\top$. Since the local cost functions are convex, the first summand of $\dot{V}$ is non-positive for all $\mathbf{y}$. Because the graph is connected, the second summand is non-positive with its null-space spanned by $\boldsymbol{\theta}\otimes\mathbf{1}_N$, $\boldsymbol{\theta}\in\mathbb{R}^d$. On this null-space, the first summand becomes $-\alpha\boldsymbol{\theta}^\top\sum_{i=1}^N(\nabla f^i(\boldsymbol{\theta}+\mathbf{x}^\star) - \nabla f^i(\mathbf{x}^\star))$, which can only be zero when $\boldsymbol{\theta}=\mathbf{0}$, because $\sum_{i=1}^N\nabla f^i(\mathbf{x}) = \nabla f(\mathbf{x})$ and the global cost function is strictly convex by assumption. Then, the two summands of $\dot{V}$ can be zero simultaneously only when $\mathbf{y}^i=\mathbf{0}$, for all $i\in\mathcal{V}$, which is equivalent to $\mathbf{z}=\mathbf{0}$. Thus, $\dot{V}$ is negative semi-definite, with $\dot{V}(\mathbf{z},\mathbf{w}_{2:N})=0$ happening on the set $\mathcal{S} = \{(\mathbf{z},\mathbf{w}_{2:N})\in\mathbb{R}^{Nd}\times\mathbb{R}^{(N-1)d} \mid \mathbf{z}=\mathbf{0}\}$. Note that (12) on $\mathcal{S}$ reduces to $\dot{\mathbf{w}}_{2:N}=\mathbf{0}$, $\dot{\mathbf{z}}_1=\mathbf{0}$, and $\dot{\mathbf{z}}_{2:N} = -\mathbf{w}_{2:N}$. Therefore, the only trajectory of (12) that remains in $\mathcal{S}$ is the equilibrium point $(\mathbf{z}_1=\mathbf{0},\mathbf{z}_{2:N}=\mathbf{0},\mathbf{w}_{2:N}=\mathbf{0})$. The LaSalle invariance principle (cf. [Khalil, 2002, Theorem 4.4 and Corollary 4.2]) now implies that the equilibrium is globally asymptotically stable or, in other words, $\mathbf{x}^i\to\mathbf{x}^\star$, $i\in\mathcal{V}$ globally asymptotically. □

Remark 9 *(Simplification of* (3)*):* For strictly convex local cost functions, using the LaSalle function of the proof of Theorem 8, one can show that the asymptotic converge to the optimizer, starting from the initial condition stated in Theorem 8, is also guaranteed for the following algorithm over connected graphs

$$\dot{\mathbf{v}}^i = \sum_{j=1}^N \mathsf{a}_{ij}(\mathbf{x}^i-\mathbf{x}^j), \qquad \dot{\mathbf{x}}^i = -\nabla f^i(\mathbf{x}^i) - \mathbf{v}^i.$$ •

5 Continuous-time Evolution with Discrete-Time Communication

Here, we investigate the design of continuous-time coordination algorithms with discrete-time communication to solve the distributed optimization problem of Section 3. The implementation of (3) requires continuous-time communication among the agents. While this abstraction is useful for analysis, in practical scenarios communication is only available at discrete instants of time. This observation motivates our study here. Throughout the section, we deal with communication topologies described by connected graphs. Our results build on the discussion of Section 4, particularly the identification of Lyapunov functions for stability.

We start by introducing some useful conventions. At any given time $t \in \mathbb{R}_{\geq 0}$, let $\hat{\mathbf{x}}^j$ be the last known state of agent $j \in \mathcal{V}$ transmitted to its in-neighbors. If $\{t_k^i\} \subset \mathbb{R}_{\geq 0}$ denotes the times at which agent i communicates with its in-neighbors, then one has $\hat{\mathbf{x}}^i = \mathbf{x}^i(t_k^i)$ for $t \in [t_k^i, t_{k+1}^i)$. Consider the next implementation of the algorithm (3) with discrete-time communication,

$$\dot{\mathbf{v}}^i = \alpha\beta \sum_{j=1}^N \mathsf{a}_{ij}(\hat{\mathbf{x}}^i - \hat{\mathbf{x}}^j), \tag{18a}$$

$$\dot{\mathbf{x}}^i = -\alpha\nabla f^i(\mathbf{x}^i) - \beta \sum_{j=1}^N \mathsf{a}_{ij}(\hat{\mathbf{x}}^i - \hat{\mathbf{x}}^j) - \mathbf{v}^i. \tag{18b}$$

Clearly, the evolution of (18) depends on the sequences of communication times for each agent. Here, we consider three scenarios. Section 5.1 studies periodic communication schemes where all agents communicate synchronously at Δ intervals of time, i.e., $t_k^i = t_k = \Delta k$ for all $i \in \mathcal{V}$. We provide a characterization of the periods that guarantee the asymptotic convergence of (18) to the optimizer. In general, periodic schemes might result in a wasteful use of the communication resources because of the need to account for worst-case situations in determining appropriate periods. This motivates our study in Section 5.2 of event-triggered communication schemes that tie the communication times to the network state for greater efficiency. We discuss two event-triggered communication implementations, a centralized synchronous one and a distributed asynchronous one. In both cases, we pay special attention to ruling out the presence of Zeno behavior (the existence of an infinite number of updates in a finite interval of time).

5.1 Periodic Communication

The following result provides an upper bound on the size of admissible stepsizes for the execution of (18) over connected graphs with periodic communication schemes.

Theorem 10 (Convergence of (18) *with periodic communication): Let $\mathcal{G}$ be a connected graph. Assume the local cost function f^i, $i \in \mathcal{V}$, is m^i-strongly convex, differentiable, and its gradient is M^i-Lipschitz on $\mathbb{R}^d$. Given $\alpha, \beta > 0$, consider an implementation of* (18) *with agents communicating over $\mathcal{G}$ synchronously every Δ seconds starting at $t_0 = 0$, i.e., $t_k^i = t_k = \Delta k$ for all $i \in \mathcal{V}$. Let $0 < \epsilon < 1$ and $\delta > 0$ such that*

$$\phi = \frac{1}{2\underline{m}}\overline{M}^2 + \frac{1}{2\underline{m}\alpha^2}\delta - 1 > 0, \tag{19}$$

where $\overline{M}$ and $\underline{m}$ are given in Theorem 1, and define

$$\tau = \frac{1}{\alpha\overline{M}+1}\ln\Big(1 + \frac{(\alpha\overline{M}+1)\zeta}{\alpha\overline{M}+1+\beta\lambda_N\sqrt{1+\alpha^2}(1+\zeta)}\Big), \tag{20}$$

where $\zeta^2 = \frac{2\epsilon(1-\epsilon)\lambda_2\min\{\delta,1\}}{\alpha\beta\lambda_N^2\phi + 4\alpha^2\lambda_2(1+\phi)^2}$. Then, if $\Delta \in (0,\tau)$, the algorithm evolution starting from initial conditions $\mathbf{x}^i(0), \mathbf{v}^i(0) \in \mathbb{R}^d$ with $\sum_{i=1}^N \mathbf{v}^i(0) = \mathbf{0}_d$ makes $\mathbf{x}^i(t) \to \mathbf{x}^\star$ exponentially fast as $t \to \infty$, for all $i \in \mathcal{V}$ with a rate of no less than $\frac{1}{4}\epsilon(\min\{\frac{1}{2},\delta\})/\lambda_{\mathbf{E}} > 0$.

PROOF. We start by transferring the equilibrium point to the origin using (11a) and then apply the change of variables (11b) to write (18) as

$$\dot{\mathbf{w}}_1 = \mathbf{0}_d, \tag{21a}$$

$$\dot{\mathbf{w}}_{2:N} = \alpha\beta(\mathbf{R}^\top\mathbf{L}\mathbf{R}\otimes\mathbf{I}_d)(\mathbf{z}_{2:N} + \tilde{\mathbf{z}}_{2:N}), \tag{21b}$$

$$\dot{\mathbf{z}}_1 = -\alpha(\mathbf{r}^\top\otimes\mathbf{I}_d)\mathbf{h}, \tag{21c}$$

$$\dot{\mathbf{z}}_{2:N} = -\alpha(\mathbf{R}^\top\otimes\mathbf{I}_d)\mathbf{h} - \beta(\mathbf{R}^\top\mathbf{L}\mathbf{R}\otimes\mathbf{I}_d)(\mathbf{z}_{2:N} + \tilde{\mathbf{z}}_{2:N}) - \mathbf{w}_{2:N}, \tag{21d}$$

where $\tilde{\mathbf{z}}_{2:N}(t) = \mathbf{z}_{2:N}(t_k) - \mathbf{z}_{2:N}(t)$, for $t \in [t_k, t_{k+1})$, and $\mathbf{h}$ is given by (13). To study the stability of (21b)-(21d), consider the Lyapunov function (16) with ϕ satisfying (19). Its Lie derivative can be bounded by (details similar to the proof of Theorem 6 are omitted for brevity)

$$\begin{aligned}\dot{V} \leq& -\frac{\delta(1-\epsilon)}{2}(\mathbf{z}_1^\top\mathbf{z}_1 + \mathbf{z}_{2:N}^\top\mathbf{z}_{2:N}) - \frac{\delta\epsilon}{2}(\mathbf{z}_1^\top\mathbf{z}_1 + \mathbf{z}_{2:N}^\top\mathbf{z}_{2:N}) \\ &-\frac{(1-\epsilon)}{2}\mathbf{w}_{2:N}^\top\mathbf{w}_{2:N} - \frac{\epsilon}{4}\mathbf{w}_{2:N}^\top\mathbf{w}_{2:N} - \phi\alpha\beta\lambda_2(1-\epsilon)\mathbf{z}_{2:N}^\top\mathbf{z}_{2:N} \\ &+ \frac{1}{4\epsilon\lambda_2}\phi\alpha\beta\lambda_N^2\tilde{\mathbf{z}}_{2:N}^\top\tilde{\mathbf{z}}_{2:N} + \frac{1}{\epsilon}\alpha^2(\phi+1)^2\tilde{\mathbf{z}}_{2:N}^\top\tilde{\mathbf{z}}_{2:N} \\ \leq& -\phi\alpha\beta\lambda_2(1-\epsilon)\mathbf{z}_{2:N}^\top\mathbf{z}_{2:N} - \frac{1}{2}\epsilon\min\{\delta,\frac{1}{2}\}\mathbf{p}^\top\mathbf{p} \\ &- \frac{1}{2}(1-\epsilon)\min\{\delta,1\}\big(\mathbf{p}^\top\mathbf{p} - \zeta^{-2}\tilde{\mathbf{z}}_{2:N}^\top\tilde{\mathbf{z}}_{2:N}\big),\end{aligned} \tag{22}$$

where $\mathbf{p} = (\mathbf{z}, \mathbf{w}_{2:N})$ and ϵ and ζ are given in the theorem's statement. Observe that at each communication time t_k, $\|\tilde{\mathbf{z}}_{2:N}(t_k)\| = 0$, then, it grows until next communication at time t_{k+1} when it becomes zero again. Our proof proceeds by showing that if $t_{k+1} < t_k + \tau$, where τ is given in (20), then we have the guarantee that

$$\|\tilde{\mathbf{z}}_{2:N}(t)\| < \zeta\,\|\mathbf{p}(t)\|, \quad t \in [t_k, t_{k+1}), \tag{23}$$

(note that, from (22), this guarantee ensures that $\dot{V}$ is negative definite for all $t \geq 0$). To this end, we study the dynamics of $q = \|\tilde{\mathbf{z}}_{2:N}\|/\|\mathbf{p}\|$ and find a lower bound on the time that it takes for q to evolve from zero (recall $\tilde{\mathbf{z}}_{2:N}(t_k) = \mathbf{0}$) to ζ. Notice that

$$\begin{aligned}\dot{q} &= \frac{\tilde{\mathbf{z}}_{2:N}^\top\dot{\tilde{\mathbf{z}}}_{2:N}}{\|\tilde{\mathbf{z}}_{2:N}\|\|\mathbf{p}\|} - \frac{\|\tilde{\mathbf{z}}_{2:N}\|\mathbf{p}^\top\dot{\mathbf{p}}}{\|\mathbf{p}\|^3} \leq (1+q)\frac{\|\dot{\mathbf{p}}\|}{\|\mathbf{p}\|} \leq (1+q)\times \\ &\frac{(\alpha\overline{M}+\beta\lambda_N\sqrt{1+\alpha^2})\|\mathbf{p}\| + \|\mathbf{w}_{2:N}\| + \beta\lambda_N\sqrt{1+\alpha^2}\|\tilde{\mathbf{z}}_{2:N}\|}{\|\mathbf{p}\|} \\ &\leq (\alpha\overline{M}+1)(1+q) + \beta\lambda_N\sqrt{1+\alpha^2}(1+q)^2.\end{aligned}$$

Here, we used in the first inequality $d/dt(\tilde{\mathbf{z}}_{2:N}) = -\dot{\mathbf{z}}_{2:N}$ and $\|\dot{\mathbf{z}}_{2:N}\| \leq \|\dot{\mathbf{p}}\|$ and in the second one the evolution of $\dot{\mathbf{p}}$ given in (21b)-(21d), $\|\mathsf{R}^\top\mathsf{L}\mathsf{R}\| \leq \|\mathsf{L}\| = \lambda_N$ and

$$\left\|\begin{bmatrix}-\beta\mathsf{R}^\top\mathsf{L}\mathsf{R}\\ \alpha\beta\mathsf{R}^\top\mathsf{L}\mathsf{R}\end{bmatrix}\right\| = \beta\lambda_N\sqrt{1+\alpha^2}.$$

Using the Comparison Lemma (cf. [Khalil, 2002, Lemma 3.4]), we conclude that $q(t, q_0) \leq \psi(t, \psi_0)$, where $\psi(t, \psi_0)$ is the solution of $\dot{\psi} = (\alpha\overline{M}+1)(1+\psi)+\beta\lambda_N\sqrt{1+\alpha^2}(1+\psi)^2$ satisfying $\psi(0, \psi_0) = \psi_0$. Then,

$$q(t,0) \leq \psi(t,0) = \frac{(\alpha\overline{M}+1+\beta\lambda_N\sqrt{1+\alpha^2})(\mathrm{e}^{\alpha\overline{M}t+t}-1)}{-\beta\lambda_N\sqrt{1+\alpha^2}\mathrm{e}^{\alpha\overline{M}t+t}+\alpha\overline{M}+1+\beta\lambda_N\sqrt{1+\alpha^2}}.$$

The time τ for $\psi(\tau, 0) = \zeta$ is given by (20). Then, for $\{t_{k+1} - t_k\}_{k\in\mathbb{Z}_{\geq 0}} < \tau$, we have (23), and as a result from (22) we have $\dot{V} < -\frac{1}{2}\epsilon\min\{\delta, \frac{1}{2}\}\mathbf{p}^\top\mathbf{p}$. Thus, $\mathbf{z} \to \mathbf{0}$, as $t \to \infty$, which is equivalent to $\mathbf{x}^i \to \mathbf{x}^\star$ as $t \to \infty$, exponentially fast with the rate given in the statement. □

Remark 11 *(Dependence of the communication period on the design parameters):* The value of τ in Theorem 10 depends on the graph topology, the parameters of the local cost functions, the design parameters α and β, and the variables ϵ and δ. One can use this dependency to maximize the value of τ. Note that the argument of ln(.) in (20) is a monotonically increasing function of $\zeta > 0$. Therefore, the smaller the value of β, the larger the value of τ. However, the dependency of τ on the rest of the parameters listed above is more complex. For given local cost functions, fixed network topology and fixed values of α, β, the maximum value of ζ is when $\phi + 1$ is at its minimum and $\epsilon\lambda_2\min\{1-\epsilon, \delta\}$ is at its maximum. •

5.2 Event-Triggered Communication

This section studies the design of event-triggered communication schemes for the execution of (18). In contrast to periodic schemes, event-triggered implementations tie the determination of the communication times to the current network state, resulting in a more efficient use of the resources. The proof of Theorem 10 reveals that the satisfaction of condition (23) guarantees the monotonic evolution of the Lyapunov function, which in turn ensures the correct asymptotic behavior of the algorithm. One could therefore specify when communication should occur by determining the times when this condition is not satisfied. There is, however, a serious drawback to this approach: the evaluation of the condition (23) requires the knowledge of the global minimizer $\mathbf{x}^\star$, which is of course not available. To see this, note that

$$\|\tilde{\mathbf{z}}_{2:N}\| = \|\mathbf{\Pi}(\mathbf{x}(t_k) - \mathbf{x})\|, \tag{24a}$$

$$\|\mathbf{p}\| = \sqrt{\|\mathbf{x}-\bar{\mathbf{x}}\|^2 + \|\mathbf{\Pi}(\mathbf{v}-\bar{\mathbf{v}})\|^2}, \tag{24b}$$

where we have used (1) and (8) (recall $\mathbf{\Pi} = \mathsf{\Pi}_N \otimes \mathsf{I}_d$). From (10), $\bar{\mathbf{v}}^i = -\alpha\nabla f^i(\mathbf{x}^\star)$ for $i \in \mathcal{V}$, and thus the evaluation of the triggering condition (23) requires knowledge of the global optimizer. Our forthcoming discussion shows how one can circumvent this problem. We first consider the design of centralized triggers requiring global network knowledge and then discuss triggering schemes that only rely on inter-neighbor interaction.

5.2.1 Centralized Synchronous Implementation

Here, we present a centralized event-triggered scheme to determine the sequence of synchronous communication times in (18). Our discussion builds upon the examination of the Lie derivative of the Lyapunov function used in the proof of Theorem 10 and the observations made above regarding the lack of knowledge of the global optimizer. From (24), we see that an event-triggered law should not employ $\mathbf{p}$, but rather rely on $\tilde{\mathbf{z}}_{2:N}$ and $\mathbf{z}_{2:N}$, to be independent of $\mathbf{x}^\star$. With this in mind, the examination of the upper bound (22) on $\dot{V}$ reveals that, if

$$\|\tilde{\mathbf{z}}_{2:N}(t)\|^2 = \|\mathbf{\Pi}(\mathbf{x}(t_k) - \mathbf{x}(t))\|^2 \leq \kappa\|\mathbf{z}_{2:N}(t)\|^2 = \kappa\|\mathbf{\Pi}\mathbf{x}(t)\|^2, \tag{25}$$

where κ is shorthand notation for

$$\kappa = 2\frac{\epsilon\delta\lambda_2 + 2\phi\alpha\beta\lambda_2^2\epsilon^2(1-\epsilon)}{\alpha\beta\phi\lambda_N^2 + 2\lambda_2\alpha^2(1+\phi)^2}, \tag{26}$$

(here $0 < \epsilon < 1$ and ϕ is given by (19)), then we have

$$\dot{V} \leq -\phi\alpha\beta\lambda_2(1-\epsilon)^2\mathbf{z}_{2:N}{}^\top\mathbf{z}_{2:N} - \frac{1}{2}(1-\epsilon)\mathbf{w}_{2:N}{}^\top\mathbf{w}_{2:N} - \frac{1}{2}\delta\mathbf{z}_1{}^\top\mathbf{z}_1 \leq -\frac{1}{2}\min\{\delta, 2\phi\alpha\beta\lambda_2(1-\epsilon)^2, (1-\epsilon)\}\mathbf{p}^\top\mathbf{p}. \tag{27}$$

Then, we can reproduce the proof of Theorem 10 and conclude the exponential convergence to the optimal solution. Accordingly, the sequence of synchronous communication times $\{t_k\}_{k\in\mathbb{Z}_{\geq 0}} \subset \mathbb{R}_{\geq 0}$ for (18) should be determined by (25). However, for a truly implementable law, one should rule out Zeno behavior, i.e., the sequence of times does not have any finite accumulation point. However, observing (25), one can see that Zeno behavior will arise at least near the agreement surface $\mathbf{\Pi}\mathbf{x} = \mathbf{0}_{dN}$. The next result details how we address this problem.

Theorem 12 (Convergence of (18) *with Zeno-free centralized event-triggered communication): Let $\mathcal{G}$ be a connected graph. Assume the local cost function f^i, $i \in \mathcal{V}$, is m^i-strongly convex, differentiable, and its gradient is M^i-Lipschitz on $\mathbb{R}^d$. Consider an implementation of* (18) *with agents communicating over $\mathcal{G}$ synchronously at $\{t_k\}_{k\in\mathbb{Z}_{\geq 0}} \subset \mathbb{R}_{\geq 0}$, starting at $t_0 = 0$,*

$$t_{k+1} = \operatorname{argmax}\{t \in [t_k + \tau, \infty) \mid \|\mathbf{\Pi}(\mathbf{x}(t_k) - \mathbf{x}(t))\|^2 \leq \kappa\|\mathbf{\Pi}\mathbf{x}(t)\|^2\}, \tag{28}$$

where τ and $\kappa < 1$ are defined in (20) *and* (26), *respectively. Then, for any given $\alpha, \beta > 0$ and each $i \in \mathcal{V}$, the algorithm evolution starting from initial conditions $\mathbf{x}^i(0), \mathbf{v}^i(0) \in \mathbb{R}^d$ with $\sum_{i=1}^N \mathbf{v}^i(0) = \mathbf{0}_d$ makes $\mathbf{x}^i(t) \to$*

$\mathbf{x}^\star$ *exponentially fast as* $t \to \infty$ *with a rate no less than* $\frac{1}{4}(\min\{\delta, 2\phi\alpha\beta\lambda_2(1-\epsilon)^2, (1-\epsilon), \frac{1}{2}\epsilon\})/\bar{\lambda}_{\mathbf{E}} > 0$.

PROOF. We first show $\kappa < 1$. This is an important property guaranteeing that, if agents start in agreement at a point other than the optimizer $\mathbf{x}^\star$, then the condition (25) is eventually violated, enforcing information updates. Notice that (a) $4\epsilon^2(1-\epsilon)\lambda_2^2 < \lambda_N^2$ and (b) $\epsilon\delta < \alpha^2(1+\phi)^2$ imply that the numerator in (26) is smaller than its denominator, and hence $\kappa < 1$. (a) follows from noting that the maximum of $4\epsilon^2(1-\epsilon)$ for $\epsilon \in (0,1)$ is $16/27 < 1$ and the fact that $\lambda_2 \le \lambda_N$. We prove (b) reasoning by contradiction. Assume $\delta > \alpha^2(1+\phi)^2$ or equivalently $\alpha(1+\phi) - \sqrt{\delta} < 0$. Using (19) and multiplying both sides of the inequality by $2\underline{m}\alpha$, we obtain

$$\alpha^2\overline{M}^2 + \delta - 2\alpha\underline{m}\sqrt{\delta} = (\sqrt{\delta} - \alpha\underline{m})^2 + \alpha^2(\overline{M}^2 - \underline{m}^2) < 0,$$

which, since $\overline{M} \ge \underline{m}$, is a contradiction. Having established the consistency of (25), consider now the candidate Lyapunov function V in (16) and let t_k be the last time at which a communication among all neighboring agents occurred. From the proof of Theorem 10, we know that the time derivative of V is negative, $\dot{V} < -\frac{1}{2}\epsilon\min\{\delta, \frac{1}{2}\}\mathbf{p}^\top\mathbf{p}$ as long as $t < t_k + \tau$. After this time, (27) shows that as long as (25) is satisfied, $\dot{V}$ is negative, and exponential convergence follows. □

Interestingly, given that (25) does not use the full state of the network but instead relies on the disagreement, one can interpret it as an output feedback event-triggered controller. Guaranteeing the existence of lower bounded inter-execution times for such controllers is in general a difficult problem, see e.g., [Donkers and Heemels, 2012]. Augmenting (25) with the condition $t_{k+1} \ge t_k + \tau$ results in Zeno-free executions by lower bounding the inter-event times by τ. The knowledge of this value also allows the designer to compute bounds on the maximum energy spent by the network on communication.

5.2.2 Distributed Asynchronous Implementation

We present a distributed event-triggered scheme for determining the sequence of communication times in (18). At each agent, the execution of the communication scheme depends only on local variables and the triggered states received from its neighbors. This naturally results in asynchronous communication. We also show that the resulting executions are free from Zeno behavior.

Theorem 13 (Convergence of (18) with Zeno-free distributed event-triggered communication): Let $\mathcal{G}$ be a connected graph. Assume f^i, $i \in \mathcal{V}$, is m^i-strongly convex, differentiable, and its gradient is M^i-Lipschitz on $\mathbb{R}^d$. For $\boldsymbol{\epsilon} \in \mathbb{R}^N_{>0}$, consider an implementation of (18) where agent $i \in \mathcal{V}$ communicates with its neighbors in $\mathcal{G}$ at times $\{t_k^i\}_{k\in\bar{\mathcal{Z}}^i \subseteq \mathbb{Z}_{\ge 0}} \subset \mathbb{R}_{\ge 0}$, starting at $t_0^i = 0$,

$$t_{k+1}^i = \operatorname{argmax}\{t \in [t_k^i, \infty) \mid \tag{29}$$
$$4\mathsf{d}_{out}^i\|\hat{\mathbf{x}}^i(t) - \mathbf{x}^i(t)\|^2 \le \sum_{j=1}^N \mathsf{a}_{ij}\|\hat{\mathbf{x}}^i(t) - \hat{\mathbf{x}}^j(t)\|^2 + (\epsilon^i)^2\}.$$

Given $\alpha > 0$, let $\beta, \phi > 0$ satisfy $\phi + 1 > 4\overline{M}$ and

$$\gamma' = \alpha^2(\phi+1)\underline{m} + \tfrac{9}{2}\beta\hat{\lambda}_2\phi\alpha - 4\alpha^2(\overline{M}\underline{m} + (\phi+1)^2) > 0. \tag{30}$$

Then, for each $i \in \mathcal{V}$, the evolution starting from initial conditions $\mathbf{x}^i(0), \mathbf{v}^i(0) \in \mathbb{R}^d$ with $\sum_{i=1}^N \mathbf{v}^i(0) = \mathbf{0}_d$ makes $\|\mathbf{x}^i(t) - \mathbf{x}^\star\| \le \frac{\phi\alpha\beta\bar{\lambda}_F}{4\eta\underline{\lambda}_F}\|\boldsymbol{\epsilon}\|^2$ as $t \to \infty$ exponentially fast with a rate no less than $\eta/\bar{\lambda}_F$. (Here, $\eta = \min\{\frac{7}{16}, \frac{1}{9}\gamma'\}$, and $\underline{\lambda}_F$ and $\bar{\lambda}_F$ are the minimum and maximum eigenvalues of $\mathbf{F}$ in (7)). Moreover, the inter-execution times of i are lower bounded by

$$\tau^i = \frac{1}{\alpha M^i}\ln\Big(1 + \frac{\alpha M^i\epsilon^i}{2\sqrt{\mathsf{d}_{out}^i}(\alpha M^i + 2\beta\mathsf{d}_{out}^i + 1)\theta}\Big), \tag{31}$$

where $\theta = \frac{\bar{\lambda}_F}{\underline{\lambda}_F}\sqrt{\|\mathbf{x}(0) - \bar{\mathbf{x}}\|^2 + \|\mathbf{v}(0) - \bar{\mathbf{v}}\|^2 + \frac{\phi\alpha\beta\bar{\lambda}_F}{4\eta\underline{\lambda}_F}\|\boldsymbol{\epsilon}\|^2}$.

PROOF. Given an initial condition, let $[0, T)$ be the maximal interval on which there is no accumulation point in $\{t_k\}_{k\in\bar{\mathcal{Z}}} = \cup_{i=1}^N \cup_{k\in\bar{\mathcal{Z}}^i\subseteq\mathbb{Z}_{\ge 0}} t_k^i$. Note that $T > 0$, since the number of agents is finite and, for each $i \in \mathcal{V}$, $\epsilon^i > 0$ and $\tilde{\mathbf{x}}^i(0) = \hat{\mathbf{x}}^i(0) - \mathbf{x}^i(0) = \mathbf{0}$. The dynamics (18), under the event-triggered communication scheme (29), has a unique solution in the time interval $[0, T)$. Next, we use Lyapunov analysis to show that the trajectory is bounded during $[0, T)$. Consider the function V given in (14), whose Lie derivative along (21b)-(21d) is

$$\begin{aligned}\dot{V} = &-\frac{1}{9}\alpha^2(\phi+1)\mathbf{y}^\top\mathbf{h} - \frac{7}{16}\mathbf{w}_{2:N}^\top\mathbf{w}_{2:N} \\ &- \|\frac{3}{4}\mathbf{w}_{2:N} + \frac{2\alpha}{3}(\mathbf{R}^\top\otimes\mathbf{I}_d)\mathbf{h} + \frac{2\alpha}{3}(\phi+1)\mathbf{z}_{2:N})\|^2 \\ &+ \frac{4}{9}\alpha^2\|(\mathbf{R}^\top\otimes\mathbf{I}_d)\mathbf{h}\|^2 + \frac{4}{9}\alpha^2(1+\phi)^2\mathbf{z}_{2:N}^\top\mathbf{z}_{2:N} \\ &- \frac{\phi\alpha\beta}{2}\mathbf{z}_{2:N}^\top(\mathbf{R}^\top\mathbf{L}\mathbf{R}\otimes\mathbf{I}_d)\mathbf{z}_{2:N} + \frac{\phi\alpha\beta}{2}s,\end{aligned}$$

where $s = -\mathbf{z}_{2:N}^\top(\mathbf{R}^\top\mathbf{L}\mathbf{R}\otimes\mathbf{I}_d)\mathbf{z}_{2:N} - 2\mathbf{z}_{2:N}^\top(\mathbf{R}^\top\mathbf{L}\mathbf{R}\otimes\mathbf{I}_d)\tilde{\mathbf{z}}_{2:N}$, and $\tilde{\mathbf{z}}_{2:N} = \hat{\mathbf{z}}_{2:N} - \mathbf{z}_{2:N}$. Using the assumptions on the local cost functions and following steps similar to those taken in the proof of Theorem 1 to lower bound $\mathbf{y}^\top\mathbf{h}$ and upper bound $\|(\mathbf{R}^\top\otimes\mathbf{I}_d)\mathbf{h}\|$ along with using $\lambda_2\mathbf{z}_{2:N}^\top\mathbf{z}_{2:N} \le \mathbf{z}_{2:N}^\top(\mathbf{R}^\top\mathbf{L}\mathbf{R}\otimes\mathbf{I}_d)\mathbf{z}_{2:N}$, one can show that

$$\dot{V} \le -\eta\|\mathbf{p}\|^2 + \frac{\phi\alpha\beta}{2}s,$$

where $\mathbf{p} = (\mathbf{z}, \mathbf{w}_{2:N})$. Next, we show $s \le \frac{1}{2}\|\boldsymbol{\epsilon}\|^2$ for $t \in [0, T)$. Using $\mathbf{R}\mathbf{R}^\top = \mathbf{\Pi}_N$, $\mathbf{L}\mathbf{\Pi}_N = \mathbf{\Pi}_N\mathbf{L} = \mathbf{L}$, $\mathbf{z}_{2:N} = (\mathbf{R}^\top\otimes\mathbf{I}_d)\mathbf{y} = (\mathbf{R}^\top\otimes\mathbf{I}_d)\mathbf{x}$, and $\tilde{\mathbf{z}}_{2:N} = (\mathbf{R}^\top\otimes\mathbf{I}_d)\tilde{\mathbf{y}} = (\mathbf{R}^\top\otimes\mathbf{I}_d)\tilde{\mathbf{x}}$, we get $s = -\mathbf{x}^\top\mathbf{L}\mathbf{x} - 2\mathbf{x}^\top\mathbf{L}\tilde{\mathbf{x}}$. Then,

$$\begin{aligned}s &= -(\hat{\mathbf{x}} - \tilde{\mathbf{x}})^\top\mathbf{L}(\hat{\mathbf{x}} - \tilde{\mathbf{x}}) - 2(\hat{\mathbf{x}} - \tilde{\mathbf{x}})^\top\mathbf{L}\tilde{\mathbf{x}} \\ &= \tilde{\mathbf{x}}^\top\mathbf{L}\tilde{\mathbf{x}} - \hat{\mathbf{x}}^\top\mathbf{L}\hat{\mathbf{x}}.\end{aligned}$$

Given $\mathbf{L} = \mathbf{D}_{\text{out}} - \mathbf{A}$ and $\mathbf{D}_{\text{out}} + \mathbf{A} \ge 0$, we have

$$\tilde{\mathbf{x}}^\top \mathbf{L}\tilde{\mathbf{x}} \leq 2\tilde{\mathbf{x}}^\top(\mathbf{D}_{\text{out}} \otimes \mathbf{I}_d)\tilde{\mathbf{x}} = 2\sum_{i=1}^N \mathsf{d}_{\text{out}}^i \|\tilde{\mathbf{x}}^i\|^2.$$

Therefore, we can write (recall $\tilde{\mathbf{x}}^i = \hat{\mathbf{x}}^i - \mathbf{x}^i$)

$$s = \frac{1}{2}\sum_{i=1}^N \left(4\mathsf{d}_{\text{out}}^i \|\hat{\mathbf{x}}^i - \mathbf{x}^i\|^2 - \sum_{j=1}^N \mathsf{a}_{ij}\|\hat{\mathbf{x}}^i - \hat{\mathbf{x}}^j\|^2\right),$$

which, with (29), yields $s \leq \frac{1}{2}\|\boldsymbol{\epsilon}\|^2$ for $t \in [0,T)$. Then,

$$\dot{V} \leq -\eta\|\mathbf{p}\|^2 + \frac{\phi\alpha\beta}{4}\|\boldsymbol{\epsilon}\|^2, \quad t \in [0,T),$$

Recall from the proof of Theorem 1 that $\underline{\lambda}_F\|\mathbf{p}\|^2 \leq V(\mathbf{p}) \leq \bar{\lambda}_F\|\mathbf{p}\|^2$. Then, using the Comparison Lemma (cf. [Khalil, 2002, Lemma 3.4]), we deduce that

$$\begin{aligned}\|\mathbf{p}(t)\| &\leq \frac{1}{\underline{\lambda}_F}\|V(0)\|\,\mathrm{e}^{-\frac{\eta}{\bar{\lambda}_F}t} + \frac{\phi\alpha\beta\bar{\lambda}_F\|\boldsymbol{\epsilon}\|^2}{4\eta\underline{\lambda}_F}(1-\mathrm{e}^{-\frac{\eta}{\bar{\lambda}_F}t}) \\ &\leq \frac{\bar{\lambda}_F}{\underline{\lambda}_F}\|\mathbf{p}(0)\|\mathrm{e}^{-\frac{\eta}{\bar{\lambda}_F}t} + \frac{\phi\alpha\beta\bar{\lambda}_F\|\boldsymbol{\epsilon}\|^2}{4\eta\underline{\lambda}_F}(1-\mathrm{e}^{-\frac{\eta}{\bar{\lambda}_F}t}), \quad (32)\end{aligned}$$

for $t \in [0,T)$. Notice that regardless of value of T,

$$\|\mathbf{p}(t)\| \leq \frac{\bar{\lambda}_F}{\underline{\lambda}_F}\|\mathbf{p}(0)\| + \frac{\phi\alpha\beta\bar{\lambda}_F}{4\eta\underline{\lambda}_F}\|\boldsymbol{\epsilon}\|^2, \qquad (33)$$

for $t \in [0,T)$. Notice that the right-hand side corresponds to θ. This can be seen by noting that, from (11b), we have $\|\mathbf{z}\| = \|\mathbf{x} - \bar{\mathbf{x}}\|$ and $\|\mathbf{w}\| = \|\mathbf{w}_{2:N}\| = \|\mathbf{v} - \bar{\mathbf{v}}\|$ (recall $\sum_{i=1}^N \mathbf{v}^i(0) = \mathbf{0}$ results in $\mathbf{w}_1 = \mathbf{0}$ for all $t \geq 0$). Our final objective is to show that $T = \infty$. To achieve this, we firt establis a lower bound on the inter-execution times of any agent. To do this, we determine a lower bound on the time it takes $i \in \mathcal{V}$ to have $\|\hat{\mathbf{x}}^i - \mathbf{x}^i\|$ evolve from 0 to $\epsilon^i/(2\sqrt{\mathsf{d}_{\text{out}}^i})$. Using (10) and (18b), we have

$$\begin{aligned}&\frac{d}{dt}\|\hat{\mathbf{x}}^i - \mathbf{x}^i\| = -\frac{(\hat{\mathbf{x}}^i - \mathbf{x}^i)^\top\dot{\mathbf{x}}^i}{\|\hat{\mathbf{x}}^i - \mathbf{x}^i\|} \leq \|\dot{\mathbf{x}}^i\| \\ &= \| - \alpha(\nabla f^i(\mathbf{x}^i) - \nabla f^i(\mathbf{x}^\star)) - \beta\sum_{j=1}^N \mathsf{a}_{ij}(\hat{\mathbf{x}}^i - \hat{\mathbf{x}}^j) \\ &\quad -(\mathbf{v}^i + \alpha\nabla f^i(\mathbf{x}^\star)\| \\ &\leq \alpha M^i\|\mathbf{x}^i - \mathbf{x}^\star\| + \beta\sum_{j=1}^N \mathsf{a}_{ij}\|\hat{\mathbf{x}}^i - \hat{\mathbf{x}}^j\| + \|\mathbf{v}^i - \bar{\mathbf{v}}^i\|,\end{aligned}$$

which leads to

$$\begin{aligned}&\frac{d}{dt}\|\hat{\mathbf{x}}^i - \mathbf{x}^i\| \leq \alpha M^i\|\hat{\mathbf{x}}^i - \mathbf{x}^i\| + \alpha M^i\|\hat{\mathbf{x}}^i - \mathbf{x}^\star\| + \\ &\quad \beta\sum_{j=1}^N \mathsf{a}_{ij}(\|\hat{\mathbf{x}}^i - \mathbf{x}^\star\| + \|\hat{\mathbf{x}}^j - \mathbf{x}^\star\|) + \|\mathbf{v}^i - \bar{\mathbf{v}}^i\|.\end{aligned}$$

From (33), we have $\|\mathbf{x}(t) - \bar{\mathbf{x}}\| \leq \theta$ and $\|\mathbf{v}(t) - \bar{\mathbf{v}}\| \leq \theta$. This implies $\|\mathbf{v}^i(t) - \bar{\mathbf{v}}^i\| \leq \theta$, $\|\hat{\mathbf{x}}^i - \mathbf{x}^\star\| \leq \theta$, and $\sum_{j=1}^N \mathsf{a}_{ij}(\|\hat{\mathbf{x}}^i - \mathbf{x}^\star\| + \|\hat{\mathbf{x}}^j - \mathbf{x}^\star\|) \leq 2\mathsf{d}_{\text{out}}^i\theta$, for all $i \in \mathcal{V}$. Therefore, from the inequality above, $\frac{d}{dt}\|\hat{\mathbf{x}}^i - \mathbf{x}^i\| \leq \alpha M^i\|\hat{\mathbf{x}}^i - \mathbf{x}^i\| + c^i$, where $c^i = (\alpha M^i + 2\beta\mathsf{d}_{\text{out}}^i + 1)\theta$. Using the Comparison Lemma (cf. [Khalil, 2002, Lemma 3.4]) and the fact that $\|\hat{\mathbf{x}}^i - \mathbf{x}^i(t_k^i)\| = 0$, we deduce

$$\|\hat{\mathbf{x}}^i - \mathbf{x}^i(t)\| \leq c^i(\mathrm{e}^{\alpha M^i(t-t_k^i)} - 1)/(\alpha M^i), \quad t \geq t_k^i.$$

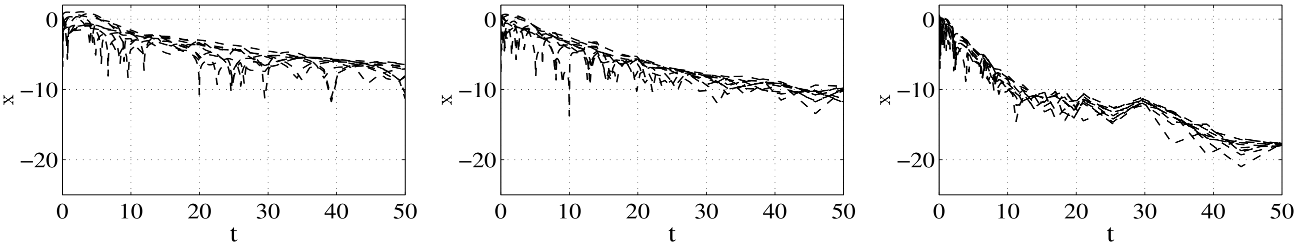
Then, the time it takes $\|\hat{\mathbf{x}}^i - \mathbf{x}^i\|$ to reach $\epsilon^i/(2\sqrt{\mathsf{d}_{\text{out}}^i})$ is lower bounded by $\tau^i > 0$ given by (31). To show $T = \infty$, we proceed by contradiction. Suppose that $T < \infty$. Then, the sequence of events $\{t_k\}_{k\in\bar{Z}}$ has an accumulation point at T. Because we have a finite number of agents, this means that there must be an agent $i \in \mathcal{V}$ for which $\{t_k^i\}_{k\in\bar{Z}^i}$ has an accumulation point at T, implying that agent i transmits infinitely often in the time interval $[T - \Delta, T)$ for any $\Delta \in (0,T]$. However, this is in contradiction with the fact that inter-event times are lower bounded by $\tau^i > 0$ on $[0,T)$. Having established $T = \infty$, note that this fact implies that under the event-triggered communication law (29), the algorithm (18) does not exhibit Zeno behavior. Furthermore, from (32), we deduce that, for each $i \in \mathcal{V}$, one has $\|\mathbf{x}^i(t) - \mathbf{x}^\star\| \leq \|\mathbf{p}(t)\| \leq \frac{\phi\alpha\beta\bar{\lambda}_F}{4\eta\underline{\lambda}_F}\|\boldsymbol{\epsilon}\|^2$ as $t \to \infty$, exponentially fast with a rate no worse than $\eta/\bar{\lambda}_F$. □

Regarding the role of the design parameters and condition (30), we omit for space reasons observations similar to the ones made in Remark 2. The lower bound on the inter-event times allows the designer to compute bounds on the maximum energy spent by each agent on communication during any given time interval. Since the total number of agents is finite and each agent's inter-event times are lower bounded, it follows that the total number of events in any finite time interval is finite. In general, an explicit expression lower bounding the network inter-event times is not available. It is also worth noticing that the farther away the agents start from the final convergence point (larger θ in (31)), the smaller the guaranteed lower bound between inter-event times becomes. As before, τ^i in (31) depends on the graph topology, the parameters of the local cost function, the design parameters α and β, and the variable ϵ^i. One can use this dependency to maximize the value of τ^i in a similar fashion as discussed in Remark 11. For quadratic local cost functions of the form $f^i(\mathbf{x}) = \frac{1}{2}(\mathbf{x}^\top\mathbf{x} + \mathbf{x}^\top\mathbf{a}^i + \mathbf{b}^i)$, $i \in \mathcal{V}$ the claim of Theorem 13 holds for any $\alpha, \beta > 0$. Finally, we point out that the discrete-time communication strategies introduced above enjoy similar privacy preservation properties as the ones stated in Proposition 5, but we omit the details here for brevity.

6 Simulations

Here, we illustrate the performance of the algorithm (3) and its implementation with discrete-time communication (18). We consider a network of 10 agents, with strongly convex local cost functions on $\mathbb{R}$ given by

$$\begin{aligned}&f^1(x) = 0.5\,\mathrm{e}^{-0.5x} + 0.4\,\mathrm{e}^{0.3x}, && f^2(x) = (x-4)^2, \\ &f^3(x) = 0.5x^2\ln(1+x^2) + x^2, && f^4(x) = x^2 + \mathrm{e}^{0.1x}, \\ &f^5(x) = \ln(\mathrm{e}^{-0.1x} + \mathrm{e}^{0.3x}) + 0.1x^2, && f^6(x) = x^2/\ln(2+x^2), \\ &f^7(x) = 0.2\,\mathrm{e}^{-0.2x} + 0.4\,\mathrm{e}^{0.4x}, && f^8(x) = x^4 + 2x^2 + 2, \\ &f^9(x) = x^2/\sqrt{x^2+1} + 0.1x^2, && f^{10}(x) = (x+2)^2.\end{aligned}$$



(a) $\alpha = 1$, $\beta = 0.5$ (b) $\alpha = 1$, $\beta = 1$ (c) $\alpha = 1$, $\beta = 5$

Fig. 1. Executions of (3) over a time-varying digraph that remains weight-balanced and strongly connected.

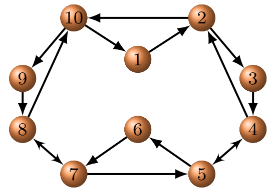
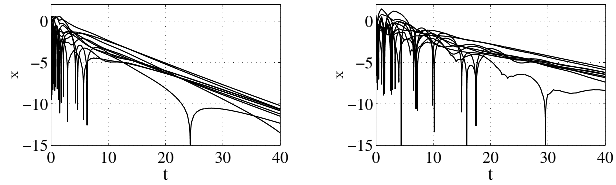
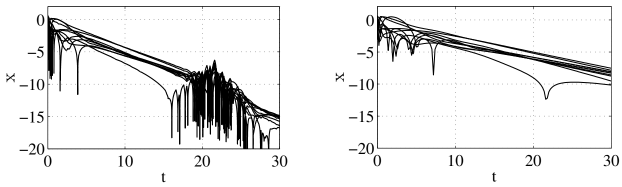


Fig. 2. One of the digraphs used in the simulations.



(a) Algorithm (18): $\alpha = 1$, $\beta = 1$, $\Delta = 0.5$ s. (b) Algorithm (18): $\alpha = 1$, $\beta = 0.5$, $\Delta = 1$ s.

Fig. 3. Performance evaluation of the algorithm (18) when the communication is periodic.



(a) Algorithm (18): $\alpha = 1$, $\beta = 2, \Delta = 0.2$ s. (b) Euler discretization of (3) ($\alpha = 1$, $\beta = 1, \Delta = 0.2$ s.)

Fig. 4. Performance evaluation of the algorithm (18) when the communication is periodic vs. the Euler-discretized implementation of the algorithm (3).

The gradient of the cost function of agents 1, 4, 7, 8 are locally Lipschitz, while the rest are globally Lipschitz. Figure 1 shows the executions of the algorithm (3) for different values of β when the network topology alternates every 2 seconds among three strongly connected, weight-balanced digraphs (with unitary edge weights, Figure 2 shows one of these graphs). Convergence is achieved as guaranteed by Proposition 4 (see also Remark 3). The plot also shows that larger values of β result in faster convergence, cf. Remark 2. In all the simulations we ran, convergence is achieved for any $\alpha, \beta > 0$.

Figures 3(a)-(b) show executions of (18) with periodic communication over the network depicted in Figure 2 for different β's and Δ's. Even though Theorem 10 is established for undirected graphs, our simulations show convergent behavior for strongly connected and weight-balanced digraphs. Figure 3 suggests a trade-off where larger Δ (corresponding to smaller β, see Remark 11) result on savings on the energy consumed by agents for communication at the cost of slower convergence.

Figures 4(a)-(b) compare the evolution of the agents for (18) with periodic communication and for an Euler discretization of (3) over the network in Figure 2. In these simulations, we fixed Δ and varied β until the algorithm becomes close to the divergence. The results show that (18) can use a larger β, which reveals that, for the same amount of communication effort, (18) achieves faster convergence.

Figure 5 shows the time history of the natural log error $x^i - \mathsf{x}^\star$ and the communication execution times of each agent $i \in \mathcal{V}$ of (18) with the distributed event-triggered communication law (29). The results illustrate the behavior guaranteed by Theorem 13: the communication times are asynchronous, the operation is Zeno-free, and the states converge to an $\|\boldsymbol{\epsilon}\|^2$-neighborhood of $\mathsf{x}^\star$.

7 Conclusions

We have presented a novel class of distributed continuous-time coordination algorithms that solve network optimization problems where the objective function is strictly convex and equal to a sum of local agent cost functions. For strongly connected and weight-balanced agent interactions, we have shown that our algorithms converge exponentially to the solution of the optimization problem when the local cost functions are strongly convex and their gradients are globally Lipschitz. This property is preserved in dynamic networks as long as the topology stays strongly connected and weight-balanced. For connected and undirected agent interactions, we have shown that exponential convergence still holds under the relaxed conditions of strongly convex local cost functions with locally Lipschitz gradients. In this case, asymptotic convergence also holds when the local cost functions are just convex. We have also explored the implementation of our algorithms with discrete-time communication. Specifically, we have established asymptotic

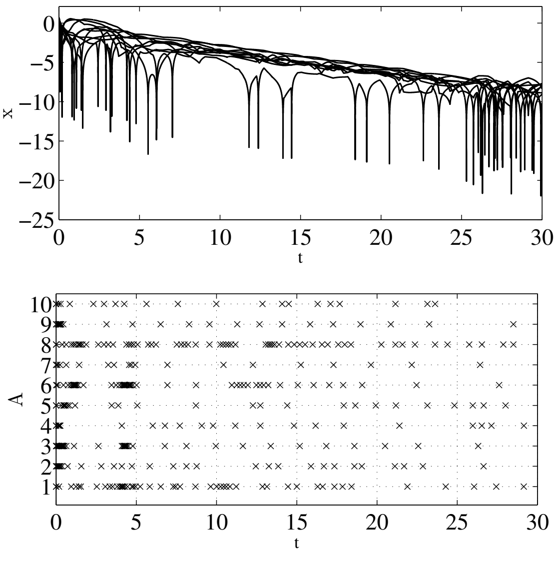


Fig. 5. Execution of algorithm (18) using $\alpha = \beta = 1$ when the distributed event-triggered communication law (29) with $\epsilon^i = 0.002$, $i \in \mathcal{V}$ is employed: in the bottom plot $\times$ shows the time an event is triggered by an agent.

convergence under periodic, centralized synchronous, and distributed asynchronous event-triggered communication schemes, paying special attention to establishing the Zeno-free nature of the algorithm executions. Future work will focus on strengthening the results to eliminate the offline computation of the design parameters, the study of the robustness against disturbances, time delays, and asynchronous agents' clocks, the exploration of agent abstractions for self-triggered implementations, and the use of triggered control methods in other coordination problems, including constrained, time-varying, and online scenarios, and networked games.

References

S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein. Distributed optimization and statistical learning via the alternating direction method of multipliers. *Foundations and Trends in Machine Learning*, 3(1):1–122, 2010.

F. Bullo, J. Cortés, and S. Martínez. *Distributed Control of Robotic Networks*. Applied Mathematics Series. Princeton University Press, 2009. ISBN 978-0-691-14195-4. Available at http://www.coordinationbook.info.

D. V. Dimarogonas, E. Frazzoli, and K. H. Johansson. Distributed event-triggered control for multi-agent systems. *IEEE Transactions on Automatic Control*, 57(5):1291–1297, 2012.

M. C. F. Donkers and W. P. M. H. Heemels. Output-based event-triggered control with guaranteed $\mathcal{L}_\infty$-gain and improved and decentralized event-triggering. *IEEE Transactions on Automatic Control*, 57(6):1362–1376, 2012.

J. Duchi, A. Agarwal, and M. Wainwright. Dual averaging for distributed optimization: Convergence analysis and network scaling. *IEEE Transactions on Automatic Control*, 57(3):592–606, 2012.

E. Garcia, Y. Cao, H. Yuc, P. Antsaklis, and D. Casbeer. Decentralised event-triggered cooperative control with limited communication. *International Journal of Control*, 86 (9):1479–1488, 2013.

B. Gharesifard and J. Cortés. Distributed continuous-time convex optimization on weight-balanced digraphs. *IEEE Transactions on Automatic Control*, 59(3):781–786, 2014.

W. P. M. H. Heemels, K.H Johansson, and P. Tabuada. An introduction to event-triggered and self-triggered control. In *IEEE Int. Conf. on Decision and Control*, pages 3270–3285, Maui, HI, 2012.

B. Johansson, M. Rabi, and M. Johansson. A randomized incremental subgradient method for distributed optimization in networked systems. *SIAM Journal on Optimization*, 20(3):1157–1170, 2009.

H. K. Khalil. *Nonlinear Systems*. Prentice Hall, 3 edition, 2002. ISBN 0130673897.

J. Lu and C. Y. Tang. Zero-gradient-sum algorithms for distributed convex optimization: The continuous-time case. *IEEE Transactions on Automatic Control*, 57(9):2348–2354, 2012.

R. Madan and S. Lall. Distributed algorithms for maximum lifetime routing in wireless sensor networks. *IEEE Transactions on Wireless Communications*, 5(8):2185–2193, 2006.

M. Mazo and P. Tabuada. Decentralized event-triggered control over wireless sensor/actuator networks. *IEEE Transactions on Automatic Control*, 56(10):2456–2461, 2011.

B. Mohar. Eigenvalues, diameter, and mean distance in graphs. *Graphs and Combinatorics*, 7:53–64, 1991.

A. Nedić and A. Ozdaglar. Distributed subgradient methods for multi-agent optimization. *IEEE Transactions on Automatic Control*, 54(1):48–61, 2009.

C. Nowzari and J. Cortés. Distributed event-triggered coordination for average consensus on weight-balanced digraphs. *Automatica*, 2014. Submitted.

S. S. Ram, A. Nedić, and V. V. Veeravalli. Distributed and recursive parameter estimation in parametrized linear state-space models. *IEEE Transactions on Automatic Control*, 55(2):488–492, 2010.

P. Wan and M.D. Lemmon. Event-triggered distributed optimization in sensor networks. In *Symposium on Information Processing of Sensor Networks*, pages 49–60, San Francisco, CA, 2009.

J. Wang and N. Elia. A control perspective for centralized and distributed convex optimization. In *IEEE Int. Conf. on Decision and Control*, Florida, USA, December 2011.

X. Wang and M. D. Lemmon. Event-triggering in distributed networked control systems. *IEEE Transactions on Automatic Control*, 56(3):586–601, 2011.

P. C. Weeraddana, G. Athanasiou, C. Fischione, and J. S. Baras. Per-se privacy preserving solution methods based on optimization. In *IEEE Int. Conf. on Decision and Control*, pages 206–211, Florence, Italy, December 2013.

F. Yan, S. Sundaram, S. Vishwanathan, and Y. Qi. Distributed autonomous online learning: regrets and intrinsic privacy-preserving properties. *IEEE Transactions on Knowledge and Data Engineering*, 35(11):2483–2493, 2013.

P. Yang, R. A. Freeman, G. J. Gordon, K. M. Lynch, S. S. Srinivasa, and R. Sukthankar. Decentralized estimation and control of graph connectivity for mobile sensor networks. *Automatica*, 46(2):390–396, 2010.

F. Zanella, D. Varagnolo, A. Cenedese, G. Pillonetto, and L. Schenato. Newton-Raphson consensus for distributed convex optimization. In *IEEE Int. Conf. on Decision and Control*, pages 5917–5922, Florida, USA, December 2011.

M. Zhu and S. Martínez. On distributed convex optimization under inequality and equality constraints. *IEEE Transactions on Automatic Control*, 57(1):151–164, 2012.