2011 American Control Conference
on O'Farrell Street, San Francisco, CA, USA
June 29 - July 01, 2011

Zero-Gradient-Sum Algorithms for Distributed Convex Optimization: The Continuous-Time Case

Jie Lu and Choon Yik Tang

***Abstract*— This paper presents a family of continuous-time distributed algorithms called *Zero-Gradient-Sum* (ZGS) algorithms, which solve unconstrained, separable, convex optimization problems over undirected networks with fixed topologies. The ZGS algorithms are derived using a Lyapunov function candidate that exploits convexity, and get their name from the fact that they yield nonlinear networked dynamical systems whose states slide along an invariant, zero-gradient-sum manifold and converge asymptotically to the unknown minimizer. We also present a systematic way to construct ZGS algorithms, show that a subset of them converge exponentially, and obtain lower bounds on their convergence rates in terms of the convexity characteristics of the problem and the network topology, including its algebraic connectivity. Finally, we show that some of the well-studied continuous-time distributed consensus algorithms are special cases of ZGS algorithms and discuss the ramifications.**

I. INTRODUCTION

This paper addresses the problem of solving an unconstrained, separable, convex optimization problem over an N-node multi-hop network, where each node i observes a convex function f_i, and all the N nodes wish to determine an optimizer x^*, which minimizes the sum of the f_i's, i.e.,

$$x^* \in \arg\min_x \sum_{i=1}^{N} f_i(x). \tag{1}$$

The optimization problem (1) has many applications in emerging and future multi-agent systems and wired/wireless/social networks, where agents or nodes often need to collaborate in order to jointly accomplish sophisticated tasks in distributed and optimal fashions [1].

To date, a family of discrete-time *subgradient algorithms*, aimed at solving problem (1) under general convexity assumptions, have been reported in the literature. These subgradient algorithms may be roughly classified into two groups. The first group of algorithms [1]–[4] are *incremental* in nature, relying on the passing of an estimate of x^* around the network to operate. Such passing may be carried out in several ways, including passing along a Hamiltonian cycle (i.e., a closed path that visit every node exactly once), random and equiprobable multi-hop passing, and probabilistic one-hop passing based on a Markov chain associated with the network. The second group of algorithms [5]–[7], in contrast to the first, are *non-incremental*, relying instead on a combination of subgradient updates and linear consensus iterations to operate, although gossip-based updates have also been considered [8]. For each of these algorithms, a number of convergence properties have been established, including the resulting error bounds, asymptotic convergence, and convergence rates.

In [9], we developed a gossip-style, distributed asynchronous algorithm, referred to as *Pairwise Equalizing* (PE), which solves the scalar version of problem (1), in a manner that is fundamentally different from the aforementioned subgradient algorithms. More recently in [10], we show that the two basic ideas behind PE—namely, the conservation of a certain gradient sum at zero and the use of a first-order-convexity-condition-based Lyapunov function—can be extended, leading to *Controlled Hopwise Equalizing* (CHE), a distributed asynchronous algorithm that allows individual nodes to use potential drops in the value of the Lyapunov function to control, on their own, when to initiate an iteration, so that problem (1) can be solved efficiently over wireless networks. In both the papers [9], [10], problem (1) was studied in a discrete-time, asynchronous setting, and only the scalar version of it was considered.

In this paper, we address problem (1) from a continuous-time and multi-dimensional standpoint, building upon the two basic ideas behind PE. Specifically, using the same Lyapunov function candidate as the one for PE and CHE, we first derive a family of continuous-time distributed algorithms called the *Zero-Gradient-Sum* (ZGS) algorithms, with which the states of the resulting nonlinear networked dynamical systems slide along an invariant, zero-gradient-sum manifold and converge asymptotically to the unknown minimizer x^* in (1). We then describe systematic and concrete ways to construct ZGS algorithms, including a class of algorithms, which turns out to be exponentially convergent. For this class of algorithms, we also provide lower bounds on their exponential convergence rates, which are expressible in terms of the convexity characteristics of the problem and the network topology, including its algebraic connectivity. As another contribution of this paper, we show that there is an intimate connection between the continuous-time distributed consensus algorithms in the literature (e.g., [11]–[16]) and the ZGS algorithms for distributed convex optimization. In particular, we found that the consensus algorithms studied in [11]–[14], [16] are only a Hessian inverse and an initial condition away from solving any convex optimization problem of the form (1).

The outline of this paper is as follows: Section II introduces the preliminaries. Section III formulates the problem. Section IV presents the ZGS algorithms. Section V analyzes the exponential convergence of a subset of the ZGS algorithms. Finally, Section VI concludes the paper.

II. PRELIMINARIES

A twice continuously differentiable function $f : \mathbb{R}^n \to \mathbb{R}$ is *locally strongly convex* if for any convex and compact

J. Lu and C. Y. Tang are with the School of Electrical and Computer Engineering, University of Oklahoma, Norman, OK 73019, USA (e-mail: {jie.lu-1,cytang}@ou.edu).

This work was supported by the National Science Foundation under grant CMMI-0900806.

set $D \subset \mathbb{R}^n$, there exists a constant $\theta > 0$ such that the following equivalent conditions hold [17], [18]:

$$f(y)-f(x)-\nabla f(x)^T(y-x) \geq \frac{\theta}{2}\|y-x\|^2, \quad \forall x, y \in D, \tag{2}$$

$$(\nabla f(y)-\nabla f(x))^T(y-x) \geq \theta\|y-x\|^2, \quad \forall x, y \in D, \tag{3}$$

$$\nabla^2 f(x) \geq \theta I_n, \quad \forall x \in D, \tag{4}$$

where $\|\cdot\|$ denotes the Euclidean norm, $\nabla f : \mathbb{R}^n \to \mathbb{R}^n$ is the gradient of f, $\nabla^2 f : \mathbb{R}^n \to \mathbb{R}^{n\times n}$ is the Hessian of f, and $I_n \in \mathbb{R}^{n\times n}$ is the identity matrix. The function f is *strongly convex* if there exists a constant $\theta > 0$ such that the equivalent conditions (2)–(4) hold for $D = \mathbb{R}^n$, in which case θ is called the *convexity parameter* of f [18]. Finally, for any twice continuously differentiable function $f : \mathbb{R}^n \to \mathbb{R}$, any convex set $D \subset \mathbb{R}^n$, and any constant $\Theta > 0$, the following conditions are equivalent [18], [19]:

$$f(y)-f(x)-\nabla f(x)^T(y-x) \leq \frac{\Theta}{2}\|y-x\|^2, \quad \forall x, y \in D, \tag{5}$$

$$(\nabla f(y)-\nabla f(x))^T(y-x) \leq \Theta\|y-x\|^2, \quad \forall x, y \in D, \tag{6}$$

$$\nabla^2 f(x) \leq \Theta I_n, \quad \forall x \in D. \tag{7}$$

III. PROBLEM FORMULATION

Consider a multi-hop network consisting of $N \geq 2$ nodes, connected by L bidirectional links in a fixed topology. The network is modeled as a connected, undirected graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, where $\mathcal{V} = \{1, 2, \ldots, N\}$ represents the set of N nodes and $\mathcal{E} \subset \{\{i, j\} : i, j \in \mathcal{V}, i \neq j\}$ represents the set of L links. Any two nodes $i, j \in \mathcal{V}$ are one-hop neighbors and can communicate if and only if $\{i, j\} \in \mathcal{E}$. The set of one-hop neighbors of each node $i \in \mathcal{V}$ is denoted as $\mathcal{N}_i = \{j \in \mathcal{V} : \{i, j\} \in \mathcal{E}\}$, and the communications are assumed to be delay- and error-free, with no quantization.

Suppose each node $i \in \mathcal{V}$ observes a function $f_i : \mathbb{R}^n \to \mathbb{R}$ satisfying the following assumption:

Assumption 1. For each $i \in \mathcal{V}$, the function f_i is twice continuously differentiable, strongly convex with convexity parameter $\theta_i > 0$, and has a locally Lipschitz Hessian $\nabla^2 f_i$.

Suppose, upon observing the f_i's, all the N nodes wish to solve the following unconstrained, separable, convex optimization problem:

$$\min_{x\in\mathbb{R}^n} F(x), \tag{8}$$

where the objective function $F : \mathbb{R}^n \to \mathbb{R}$ is defined as $F(x) = \sum_{i\in\mathcal{V}} f_i(x)$. The proposition below shows that F has a unique minimizer $x^* \in \mathbb{R}^n$, so that problem (8) is well-posed:

Proposition 1. *With Assumption 1, there exists a unique $x^* \in \mathbb{R}^n$ such that $F(x^*) \leq F(x)$ $\forall x \in \mathbb{R}^n$ and $\nabla F(x^*) = 0$.*

Proof. By Assumption 1, F is twice continuously differentiable and strongly convex with convexity parameter $\sum_{j\in\mathcal{V}} \theta_j > 0$. Pick any $x_o \in \mathbb{R}^n$ and define the set $D = \{x \in \mathbb{R}^n : F(x) \leq F(x_o)\}$. Since $x_o \in D$ and F is continuous, D is nonempty and closed. Pick any $y \in \mathbb{R}^n$ with $\|y\| = 1$ and consider the ray $\{x_o + \eta y \in \mathbb{R}^n : \eta \geq 0\}$. From (2), $F(x_o + \eta y) \geq F(x_o) + \eta\nabla F(x_o)^T y + \eta^2 \frac{\sum_{j\in\mathcal{V}}\theta_j}{2}\|y\|^2$. Since $\|y\| = 1$ and $\eta \geq 0$, $F(x_o + \eta y) \geq F(x_o) - \eta\|\nabla F(x_o)\| + \eta^2\frac{\sum_{j\in\mathcal{V}}\theta_j}{2}$. Therefore, $\forall \eta > \frac{2\|\nabla F(x_o)\|}{\sum_{j\in\mathcal{V}}\theta_j}$, $F(x_o + \eta y) > F(x_o)$, so that $x_o + \eta y \notin D$. Hence, D is bounded and, thus, compact. Since F is continuous, there exists an $x^* \in D$ such that $F(x^*) \leq F(x)$ $\forall x \in D$. By definition of D, $F(x^*) \leq F(x)$ $\forall x \in \mathbb{R}^n$. Because F is strongly convex, x^* is unique and satisfies $\nabla F(x^*) = 0$. □

Given the above network and problem, the aim of this paper is to devise a continuous-time distributed algorithm of the form

$$\dot{x}_i(t) = \varphi_i(x_i(t), \mathbf{x}_{\mathcal{N}_i}(t); f_i, \mathbf{f}_{\mathcal{N}_i}), \quad \forall t \geq 0, \forall i \in \mathcal{V}, \tag{9}$$

$$x_i(0) = \chi_i(f_i, \mathbf{f}_{\mathcal{N}_i}), \quad \forall i \in \mathcal{V}, \tag{10}$$

where $t \geq 0$ denotes time; $x_i(t) \in \mathbb{R}^n$ is a state representing node i's estimate of the unknown minimizer x^* at time t; $\mathbf{x}_{\mathcal{N}_i}(t) = (x_j(t))_{j\in\mathcal{N}_i} \in \mathbb{R}^{n|\mathcal{N}_i|}$ is a vector obtained by stacking $x_j(t)$ $\forall j \in \mathcal{N}_i$; $\mathbf{f}_{\mathcal{N}_i} = (f_j)_{j\in\mathcal{N}_i} : \mathbb{R}^n \to \mathbb{R}^{|\mathcal{N}_i|}$ is a function obtained by stacking f_j $\forall j \in \mathcal{N}_i$; $\varphi_i : \mathbb{R}^n \times \mathbb{R}^{n|\mathcal{N}_i|} \to \mathbb{R}^n$ is a locally Lipschitz function of $x_i(t)$ and $\mathbf{x}_{\mathcal{N}_i}(t)$ governing the dynamics of $x_i(t)$, whose definition may depend on f_i and $\mathbf{f}_{\mathcal{N}_i}$; $\chi_i \in \mathbb{R}^n$ is a constant determining the initial state $x_i(0)$, whose value may depend on f_i and $\mathbf{f}_{\mathcal{N}_i}$; $|\cdot|$ denotes the cardinality of a set; and $x_i(t)$, f_i, φ_i, and χ_i are maintained in node i's local memory. The goal of the algorithm (9) and (10) is to steer all the estimates $x_i(t)$'s asymptotically (or, better yet, exponentially) to the unknown x^*, i.e.,

$$\lim_{t\to\infty} x_i(t) = x^*, \quad \forall i \in \mathcal{V}, \tag{11}$$

enabling all the nodes to cooperatively solve problem (8). Note that to realize (9) and (10), for each $i \in \mathcal{V}$, every node $j \in \mathcal{N}_i$ must send node i its $x_j(t)$ at each time t if φ_i does depend on $x_j(t)$, and its f_j at time $t = 0$ if φ_i or χ_i does depend on f_j.

IV. ZERO-GRADIENT-SUM ALGORITHMS

In this section, we develop a family of algorithms that achieve the stated goal. To facilitate the development, we let $\mathbf{x}^* = (x^*, x^*, \ldots, x^*) \in \mathbb{R}^{nN}$ denote the vector of minimizers and $\mathbf{x}(t) = (x_1(t), x_2(t), \ldots, x_N(t)) \in \mathbb{R}^{nN}$, or simply $\mathbf{x} = (x_1, x_2, \ldots, x_N)$, denote the entire state vector.

Consider a Lyapunov function candidate $V : \mathbb{R}^{nN} \to \mathbb{R}$, defined in terms of the observed f_i's as

$$V(\mathbf{x}) = \sum_{i\in\mathcal{V}} f_i(x^*) - f_i(x_i) - \nabla f_i(x_i)^T(x^* - x_i). \tag{12}$$

Notice that V in (12) is continuously differentiable because of Assumption 1, and that it satisfies $V(\mathbf{x}^*) = 0$. Moreover, V is positive definite with respect to $\mathbf{x}^*$ and is radially unbounded, which can be seen by noting that Assumption 1 and the first-order strong convexity condition (2) imply

$$V(\mathbf{x}) \geq \sum_{i\in\mathcal{V}} \frac{\theta_i}{2}\|x^* - x_i\|^2, \quad \forall \mathbf{x} \in \mathbb{R}^{nN}, \tag{13}$$

and (13) in turn implies $V(\mathbf{x}) > 0$ $\forall \mathbf{x} \neq \mathbf{x}^*$ and $V(\mathbf{x}) \to \infty$ as $\|\mathbf{x}\| \to \infty$. Therefore, V in (12) is a legitimate Lyapunov function candidate, which may be used to derive algorithms that ensure (11).

Taking the time derivative of V along the state trajectory $\mathbf{x}(t)$ of the system (9) and calling it $\dot{V}:\mathbb{R}^{nN}\to\mathbb{R}$, we obtain

$$\dot{V}(\mathbf{x}(t))=\sum_{i\in\mathcal{V}}(x_i(t)-x^*)^T\nabla^2 f_i(x_i(t))\cdot\varphi_i(x_i(t),\mathbf{x}_{\mathcal{N}_i}(t);f_i,\mathbf{f}_{\mathcal{N}_i}),\quad\forall t\geq 0.\tag{14}$$

Due to Assumption 1 and to each φ_i being locally Lipschitz, $\dot{V}$ in (14) is continuous. In addition, it yields $\dot{V}(\mathbf{x}^*)=0$. Hence, if the functions φ_i $\forall i\in\mathcal{V}$ are such that $\dot{V}$ is negative definite with respect to $\mathbf{x}^*$, i.e.,

$$\sum_{i\in\mathcal{V}}(x_i-x^*)^T\nabla^2 f_i(x_i)\varphi_i(x_i,\mathbf{x}_{\mathcal{N}_i};f_i,\mathbf{f}_{\mathcal{N}_i})<0,\quad\forall\mathbf{x}\neq\mathbf{x}^*,\tag{15}$$

the system (9) would have a unique equilibrium point at $\mathbf{x}^*$, which by the Barbashin-Krasovskii theorem would be globally asymptotically stable. Consequently, regardless of how the constants χ_i $\forall i\in\mathcal{V}$ in (10) are chosen, the goal (11) would be accomplished.

As it follows from the above, the challenge lies in finding φ_i $\forall i\in\mathcal{V}$, which collectively satisfy (15). Such φ_i's, however, may be difficult to construct because x^* in (15) is unknown to any of the nodes, i.e., x^* depends on *every* f_i via (8), but φ_i maintained by each node $i\in\mathcal{V}$ can *only* depend on f_i and $\mathbf{f}_{\mathcal{N}_i}$. As a result, one cannot let the φ_i's depend on x^*, such as letting $\varphi_i(x_i,\mathbf{x}_{\mathcal{N}_i};f_i,\mathbf{f}_{\mathcal{N}_i})=x^*-x_i$ $\forall i\in\mathcal{V}$, even though this particular choice guarantees (15) (since each $\nabla^2 f_i(x_i)$ is positive definite, by (4)). Given that the required φ_i's are not readily apparent, instead of searching for them, below we present an alternative approach toward the goal (11), which uses the same V and $\dot{V}$ as in (12) and (14), but demands neither local nor global asymptotic stability.

To state the approach, we first introduce two definitions: let $\mathcal{A}\subset\mathbb{R}^{nN}$ represent the *agreement set* and $\mathcal{M}\subset\mathbb{R}^{nN}$ represent the *zero-gradient-sum manifold*, defined respectively as

$$\mathcal{A}=\{(y_1,y_2,\ldots,y_N)\in\mathbb{R}^{nN}:y_1=y_2=\cdots=y_N\},\tag{16}$$
$$\mathcal{M}=\{(y_1,y_2,\ldots,y_N)\in\mathbb{R}^{nN}:\sum_{i\in\mathcal{V}}\nabla f_i(y_i)=0\},\tag{17}$$

so that $\mathbf{x}\in\mathcal{A}$ if and only if all the x_i's agree, and $\mathbf{x}\in\mathcal{M}$ if and only if the sum of all the gradients ∇f_i's, evaluated respectively at the x_i's, is zero. Notice from (16) that $\mathbf{x}^*\in\mathcal{A}$, from (17) and Proposition 1 that $\mathbf{x}^*\in\mathcal{M}$, and from all of them that $\mathbf{x}\in\mathcal{A}\cap\mathcal{M}\Rightarrow\mathbf{x}=\mathbf{x}^*$. Thus, $\mathcal{A}\cap\mathcal{M}=\{\mathbf{x}^*\}$. Also note from the continuity of each ∇f_i that $\mathcal{M}$ is closed and from the Implicit Function Theorem and the nonsingularity of each $\nabla^2 f_i(x)$ $\forall x\in\mathbb{R}^n$ that $\mathcal{M}$ is indeed a manifold of dimension $n(N-1)$.

Having introduced $\mathcal{A}$ and $\mathcal{M}$, we now describe the approach, which is based on the following recognition: to attain the goal (11), condition (15)—which ensures that *every* trajectory $\mathbf{x}(t)$ goes to $\mathbf{x}^*$—is sufficient but not necessary. Rather, all that is needed is a *single* trajectory $\mathbf{x}(t)$, along which $\dot{V}(\mathbf{x}(t))\leq 0$ $\forall t\geq 0$ and $\lim_{t\to\infty}V(\mathbf{x}(t))=0$, since the latter implies (11). Recognizing this, we next derive three conditions on the φ_i's and χ_i's in (9) and (10) that produce such a trajectory. Assume, for a moment, that the χ_i's dictating the initial state $\mathbf{x}(0)$ have been decided, so that we may focus on the φ_i's that shape the trajectory $\mathbf{x}(t)$ leaving $\mathbf{x}(0)$. Observe that $\dot{V}$ in (14) takes the form $\dot{V}(\mathbf{x}(t))=\Phi_1(\mathbf{x}(t))-x^{*T}\Phi_2(\mathbf{x}(t))$ $\forall t\geq 0$, where $\Phi_1:\mathbb{R}^{nN}\to\mathbb{R}$ and $\Phi_2:\mathbb{R}^{nN}\to\mathbb{R}^n$. Thus, the unknown x^*—which may undesirably affect the sign of $\dot{V}(\mathbf{x}(t))$—can be eliminated by setting $\Phi_2(\mathbf{x})=0$ $\forall\mathbf{x}\in\mathbb{R}^{nN}$, i.e., by forcing the φ_i's to satisfy

$$\sum_{i\in\mathcal{V}}\nabla^2 f_i(x_i)\varphi_i(x_i,\mathbf{x}_{\mathcal{N}_i};f_i,\mathbf{f}_{\mathcal{N}_i})=0,\quad\forall\mathbf{x}\in\mathbb{R}^{nN}.\tag{18}$$

With this first condition (18), $\dot{V}$ becomes free of x^*, reducing to

$$\dot{V}(\mathbf{x}(t))=\sum_{i\in\mathcal{V}}x_i(t)^T\nabla^2 f_i(x_i(t))\varphi_i(x_i(t),\mathbf{x}_{\mathcal{N}_i}(t);f_i,\mathbf{f}_{\mathcal{N}_i}),\quad\forall t\geq 0.\tag{19}$$

Next, notice that whenever $\mathbf{x}(t)$ is in the agreement set $\mathcal{A}$, due to (16) and (18), $\dot{V}(\mathbf{x}(t))$ in (19) must vanish. However, whenever $\mathbf{x}(t)\notin\mathcal{A}$, there is no such restriction. Hence, any time $\mathbf{x}(t)\notin\mathcal{A}$, $\dot{V}(\mathbf{x}(t))$ can be made negative by forcing the φ_i's to also satisfy

$$\sum_{i\in\mathcal{V}}x_i^T\nabla^2 f_i(x_i)\varphi_i(x_i,\mathbf{x}_{\mathcal{N}_i};f_i,\mathbf{f}_{\mathcal{N}_i})<0,\quad\forall\mathbf{x}\in\mathbb{R}^{nN}-\mathcal{A}.\tag{20}$$

With this additional, second condition (20), no matter what x^* is, $\dot{V}(\mathbf{x}(t))\leq 0$ along $\mathbf{x}(t)$, with equality if and only if $\mathbf{x}(t)\in\mathcal{A}$. Finally, note that (18) and (9) imply $\frac{d}{dt}\sum_{i\in\mathcal{V}}\nabla f_i(x_i(t))=\sum_{i\in\mathcal{V}}\nabla^2 f_i(x_i(t))\dot{x}_i(t)=0$ $\forall t\geq 0$, while (11), the continuity of each ∇f_i, and Proposition 1 imply $\lim_{t\to\infty}\sum_{i\in\mathcal{V}}\nabla f_i(x_i(t))=\sum_{i\in\mathcal{V}}\nabla f_i(\lim_{t\to\infty}x_i(t))=\sum_{i\in\mathcal{V}}\nabla f_i(x^*)=\nabla F(x^*)=0$. The former says that by making the φ_i's satisfy (18), the gradient sum $\sum_{i\in\mathcal{V}}\nabla f_i(x_i(t))$ along $\mathbf{x}(t)$ would remain constant over time, while the latter says that to achieve $\lim_{t\to\infty}V(\mathbf{x}(t))=0$ or equivalently (11), this constant sum must be zero, i.e., $\sum_{i\in\mathcal{V}}\nabla f_i(x_i(t))=0$ $\forall t\geq 0$. Therefore, in view of (10), the χ_i's must be such that

$$\sum_{i\in\mathcal{V}}\nabla f_i(\chi_i(f_i,\mathbf{f}_{\mathcal{N}_i}))=0,\tag{21}$$

yielding the third and final condition.

By imposing algebraic constraints on the φ_i's and χ_i's, conditions (18), (20), and (21) characterize a family of algorithms. This family of algorithms share a number of properties, including one that has a nice geometric interpretation: observe from (21), (10), and (17) that $\mathbf{x}(0)\in\mathcal{M}$ and further from (18) and (9) that $\mathbf{x}(t)\in\mathcal{M}$ $\forall t>0$. Thus, every algorithm in the family produces a nonlinear networked dynamical system, whose trajectory $\mathbf{x}(t)$ begins on, and slides along, the zero-gradient-sum manifold $\mathcal{M}$, making $\mathcal{M}$ a positively invariant set. Due to this geometric interpretation, these algorithms are referred to as follows:

Definition 1. A continuous-time distributed algorithm of the form (9) and (10) is said to be a *Zero-Gradient-Sum* (ZGS) algorithm if φ_i $\forall i\in\mathcal{V}$ are locally Lipschitz and satisfy (18) and (20), and χ_i $\forall i\in\mathcal{V}$ satisfy (21).

The following theorem lists the properties shared by ZGS algorithms, showing that every one of them is capable of asymptotically driving $\mathbf{x}(t)$ to $\mathbf{x}^*$, solving problem (8):

Theorem 1. *Consider the network modeled in Section III and the use of a ZGS algorithm described in Definition 1. Suppose Assumption 1 holds. Then: (i) there exists a unique solution* $\mathbf{x}(t)$ $\forall t \geq 0$ *to* (9) *and* (10)*; (ii)* $\mathbf{x}(t) \in \mathcal{M}$ $\forall t \geq 0$*; (iii)* $\dot{V}(\mathbf{x}(t)) \leq 0$ $\forall t \geq 0$*, with equality if and only if* $\mathbf{x}(t) = \mathbf{x}^*$*; (iv)* $\lim_{t\to\infty} V(\mathbf{x}(t)) = 0$*; and (v)* $\lim_{t\to\infty} \mathbf{x}(t) = \mathbf{x}^*$*, i.e.,* (11) *holds.*

Proof. Since φ_i $\forall i \in \mathcal{V}$ are locally Lipschitz, to prove (i) it suffices to show that every solution $\mathbf{x}(t)$ of (9) and (10) lies entirely in a compact subset of $\mathbb{R}^{nN}$. To this end, let $\mathcal{B}(\mathbf{x}^*, r) \subset \mathbb{R}^{nN}$ denote the closed-ball of radius $r \in [0, \infty)$ centered at $\mathbf{x}^*$, i.e., $\mathcal{B}(\mathbf{x}^*, r) = \{\mathbf{y} \in \mathbb{R}^{nN} : \|\mathbf{y} - \mathbf{x}^*\| \leq r\}$. Note from (14), (18), and (20) that $\dot{V}(\mathbf{x}(t)) \leq 0$ along $\mathbf{x}(t)$. This, together with (13), implies that $V(\mathbf{x}(0)) \geq V(\mathbf{x}(t)) \geq \frac{\min_{i\in\mathcal{V}} \theta_i}{2}\|\mathbf{x}(t) - \mathbf{x}^*\|^2$ along $\mathbf{x}(t)$. Hence, $\mathbf{x}(t) \in \mathcal{B}(\mathbf{x}^*, \sqrt{\frac{2V(\mathbf{x}(0))}{\min_{i\in\mathcal{V}} \theta_i}})$ $\forall t \geq 0$, ensuring (i). Statement (ii) has been proven in the paragraph before Definition 1. To verify (iii), notice again from (14), (18), and (20) that $\dot{V}(\mathbf{x}(t)) = 0$ if and only if $\mathbf{x}(t) \in \mathcal{A}$. Due to (ii) and to $\mathcal{A} \cap \mathcal{M} = \{\mathbf{x}^*\}$ shown earlier, (iii) holds. To prove (iv), observe from (13) and (iii) that $V(\mathbf{x}(t))$ $\forall t \geq 0$ is nonnegative and non-increasing. Thus, there exists a $c \geq 0$ such that $\lim_{t\to\infty} V(\mathbf{x}(t)) = c$ and $V(\mathbf{x}(t)) \geq c$ $\forall t \geq 0$. To show that $c = 0$, assume to the contrary that $c > 0$. Then, because V in (12) is continuous and positive definite with respect to $\mathbf{x}^*$, there exists an $\epsilon > 0$ such that $\mathcal{B}(\mathbf{x}^*, \epsilon) \subset \{\mathbf{y} \in \mathbb{R}^{nN} : V(\mathbf{y}) < c\}$. With this ϵ, define a set $\mathcal{K} \subset \mathbb{R}^{nN}$ as $\mathcal{K} = \mathcal{M} \cap \{\mathbf{y} \in \mathbb{R}^{nN} : \epsilon \leq \|\mathbf{y} - \mathbf{x}^*\| \leq \sqrt{\frac{2V(\mathbf{x}(0))}{\min_{i\in\mathcal{V}} \theta_i}}\}$. Notice that $\mathbf{x}(t) \in \mathcal{K}$ $\forall t \geq 0$ because $\mathbf{x}(t) \in \mathcal{M}$, $V(\mathbf{x}(t)) \geq c$, and $\mathbf{x}(t) \in \mathcal{B}(\mathbf{x}^*, \sqrt{\frac{2V(\mathbf{x}(0))}{\min_{i\in\mathcal{V}} \theta_i}})$ $\forall t \geq 0$. Also note that $\mathcal{K} \subset \mathcal{M}$ but $\mathcal{K} \not\ni \mathbf{x}^*$. This, along with the properties $\mathcal{A} \cap \mathcal{M} = \{\mathbf{x}^*\}$ and $\dot{V}(\mathbf{y}) < 0$ $\forall \mathbf{y} \notin \mathcal{A}$, implies that $\dot{V}(\mathbf{y}) < 0$ $\forall \mathbf{y} \in \mathcal{K}$. Since $\dot{V}$ in (14) is continuous and $\mathcal{K}$ is nonempty and compact (due to $\mathcal{M}$ being a closed set), there exists an $\eta > 0$ such that $\max_{\mathbf{y}\in\mathcal{K}} \dot{V}(\mathbf{y}) = -\eta$. Since $\mathbf{x}(t) \in \mathcal{K}$ $\forall t \geq 0$, $V(\mathbf{x}(t)) = V(\mathbf{x}(0)) + \int_0^t \dot{V}(\mathbf{x}(\tau))\, d\tau \leq V(\mathbf{x}(0)) - \eta t$. This implies $V(\mathbf{x}(t)) < c$ $\forall t > \frac{V(\mathbf{x}(0)) - c}{\eta}$, which is a contradiction. Therefore, $c = 0$, establishing (iv). Finally, (v) is an immediate consequence of (13) and (iv). ☐

Having established Theorem 1, we now present a systematic way to construct ZGS algorithms. First, to find χ_i's that meet condition (21), consider the following proposition, which shows that each f_i has a unique minimizer $x_i^* \in \mathbb{R}^n$:

Proposition 2. *With Assumption 1, for each* $i \in \mathcal{V}$*, there exists a unique* $x_i^* \in \mathbb{R}^n$ *such that* $f_i(x_i^*) \leq f_i(x)$ $\forall x \in \mathbb{R}^n$ *and* $\nabla f_i(x_i^*) = 0$*.*

Proof. For each $i \in \mathcal{V}$, the proof is identical to that of Proposition 1 with x^*, F, and $\sum_{j\in\mathcal{V}} \theta_j$ replaced by x_i^*, f_i, and θ_i, respectively. ☐

Proposition 2 implies that $\sum_{i\in\mathcal{V}} \nabla f_i(x_i^*) = 0$. Hence, (21) can be met by simply letting

$$\chi_i(f_i, \mathbf{f}_{\mathcal{N}_i}) = x_i^*, \quad \forall i \in \mathcal{V}, \tag{22}$$

which is permissible since every x_i^* in (22) depends just on f_i. It follows that each node $i \in \mathcal{V}$ must solve a "local" convex optimization problem $\min_{x\in\mathbb{R}^n} f_i(x)$ for x_i^* before time $t = 0$, in order to execute (10) and (22).

Next, to generate locally Lipschitz φ_i's that ensure conditions (18) and (20), notice that each φ_i is premultiplied by $\nabla^2 f_i(x_i)$, which is nonsingular $\forall x_i \in \mathbb{R}^n$. Therefore, the impact of each $\nabla^2 f_i(x_i)$ can be absorbed by setting

$$\varphi_i(x_i, \mathbf{x}_{\mathcal{N}_i}; f_i, \mathbf{f}_{\mathcal{N}_i}) = (\nabla^2 f_i(x_i))^{-1} \phi_i(x_i, \mathbf{x}_{\mathcal{N}_i}; f_i, \mathbf{f}_{\mathcal{N}_i}), \quad \forall i \in \mathcal{V}, \tag{23}$$

where $\phi_i : \mathbb{R}^n \times \mathbb{R}^{n|\mathcal{N}_i|} \to \mathbb{R}^n$ is a locally Lipschitz function of x_i and $\mathbf{x}_{\mathcal{N}_i}$ maintained by node i. For each $i \in \mathcal{V}$, because $\nabla^2 f_i$ is locally Lipschitz (due to Assumption 1) and the determinant of $\nabla^2 f_i(x_i)$ for every $x_i \in \mathbb{R}^n$ is no less than a positive constant θ_i^n (due further to (4)), the mapping $(\nabla^2 f_i(\cdot))^{-1} : \mathbb{R}^n \to \mathbb{R}^{n\times n}$ in (23) is locally Lipschitz. Thus, as long as the ϕ_i's are locally Lipschitz, so would the resulting φ_i's, fulfilling the requirement. With (23), the dynamics (9) become

$$\dot{x}_i(t) = (\nabla^2 f_i(x_i(t)))^{-1} \phi_i(x_i(t), \mathbf{x}_{\mathcal{N}_i}(t); f_i, \mathbf{f}_{\mathcal{N}_i}), \quad \forall t \geq 0,\ \forall i \in \mathcal{V}, \tag{24}$$

and conditions (18) and (20) simplify to

$$\sum_{i\in\mathcal{V}} \phi_i(x_i, \mathbf{x}_{\mathcal{N}_i}; f_i, \mathbf{f}_{\mathcal{N}_i}) = 0, \quad \forall \mathbf{x} \in \mathbb{R}^{nN}, \tag{25}$$

$$\sum_{i\in\mathcal{V}} x_i^T \phi_i(x_i, \mathbf{x}_{\mathcal{N}_i}; f_i, \mathbf{f}_{\mathcal{N}_i}) < 0, \quad \forall \mathbf{x} \in \mathbb{R}^{nN} - \mathcal{A}. \tag{26}$$

Finally, to come up with locally Lipschitz ϕ_i's that assure conditions (25) and (26), suppose each ϕ_i is decomposed as

$$\phi_i(x_i, \mathbf{x}_{\mathcal{N}_i}; f_i, \mathbf{f}_{\mathcal{N}_i}) = \sum_{j\in\mathcal{N}_i} \phi_{ij}(x_i, x_j; f_i, f_j), \quad \forall i \in \mathcal{V}, \tag{27}$$

so that the dynamics (24) become

$$\dot{x}_i(t) = (\nabla^2 f_i(x_i(t)))^{-1} \sum_{j\in\mathcal{N}_i} \phi_{ij}(x_i(t), x_j(t); f_i, f_j), \quad \forall t \geq 0,\ \forall i \in \mathcal{V}, \tag{28}$$

where $\phi_{ij} : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}^n$ is a locally Lipschitz function of x_i and x_j maintained by node i. Then, (25) can be ensured by requiring that every ϕ_{ij} and ϕ_{ji} pair be negative of each other, i.e.,

$$\phi_{ij}(y, z; f_i, f_j) = -\phi_{ji}(z, y; f_j, f_i), \quad \forall i \in \mathcal{V},\ \forall j \in \mathcal{N}_i,\ \forall y, z \in \mathbb{R}^n, \tag{29}$$

since $\sum_{i\in\mathcal{V}} \phi_i = \sum_{i\in\mathcal{V}} \sum_{j\in\mathcal{N}_i} \phi_{ij} = \sum_{\{i,j\}\in\mathcal{E}} \phi_{ij} + \phi_{ji} = 0$. With (27) and (29), the left-hand side of (26) turns into

$$\sum_{i\in\mathcal{V}} x_i^T \phi_i(x_i, \mathbf{x}_{\mathcal{N}_i}; f_i, \mathbf{f}_{\mathcal{N}_i}) = \frac{1}{2} \sum_{i\in\mathcal{V}} \sum_{j\in\mathcal{N}_i} (x_i - x_j)^T \phi_{ij}(x_i, x_j; f_i, f_j), \quad \forall \mathbf{x} \in \mathbb{R}^{nN}. \tag{30}$$

Because the graph $\mathcal{G}$ is connected, for any $\mathbf{x} \in \mathbb{R}^{nN} - \mathcal{A}$, there exist $i \in \mathcal{V}$ and $j \in \mathcal{N}_i$ such that $x_i - x_j$ in (30) is nonzero. Hence, (26) can be guaranteed by requiring the ϕ_{ij}'s to also satisfy

$$(y - z)^T \phi_{ij}(y, z; f_i, f_j) < 0, \quad \forall i \in \mathcal{V},\ \forall j \in \mathcal{N}_i,\ \forall y, z \in \mathbb{R}^n,\ y \neq z. \tag{31}$$

Note that if (29) holds, then ϕ_{ij} satisfies the inequality in (31) if and only if ϕ_{ji} does. Therefore, every pair of neighboring nodes $i, j \in \mathcal{V}$ need only minimal coordination before time $t = 0$ to realize the dynamics (28): only one of them, say, node i, needs to construct a ϕ_{ij} that satisfies the inequality in (31), and the other, i.e., node j, only needs to make sure that $\phi_{ji} = -\phi_{ij}$.

Examples 1 and 2 below illustrate two concrete ways to construct ϕ_{ij}'s that obey (29) and (31):

Example 1. Let $\phi_{ij}(y, z; f_i, f_j) = (\psi_{ij1}(y_1, z_1), \psi_{ij2}(y_2, z_2), \ldots, \psi_{ijn}(y_n, z_n))$ $\forall i \in \mathcal{V}$ $\forall j \in \mathcal{N}_i$ $\forall y = (y_1, y_2, \ldots, y_n) \in \mathbb{R}^n$ $\forall z = (z_1, z_2, \ldots, z_n) \in \mathbb{R}^n$, where each $\psi_{ij\ell} : \mathbb{R}^2 \to \mathbb{R}$ can be any locally Lipschitz function satisfying $\psi_{ij\ell}(y_\ell, z_\ell) = -\psi_{ji\ell}(z_\ell, y_\ell)$ and $(y_\ell - z_\ell)\psi_{ij\ell}(y_\ell, z_\ell) < 0$ whenever $y_\ell \neq z_\ell$ (e.g., $\psi_{ij\ell}(y_\ell, z_\ell) = \tanh(z_\ell - y_\ell)$ or $\psi_{ij\ell}(y_\ell, z_\ell) = -\psi_{ji\ell}(z_\ell, y_\ell) = \frac{z_\ell - y_\ell}{1+y_\ell^2}$). Then, (29) and (31) hold. ∎

Example 2. Let $\phi_{ij}(y, z; f_i, f_j) = \nabla g_{\{i,j\}}(z) - \nabla g_{\{i,j\}}(y)$ $\forall i \in \mathcal{V}$ $\forall j \in \mathcal{N}_i$ $\forall y, z \in \mathbb{R}^n$, where each $g_{\{i,j\}} : \mathbb{R}^n \to \mathbb{R}$ can be any twice continuously differentiable and locally strongly convex function associated with link $\{i, j\} \in \mathcal{E}$ (e.g., $g_{\{i,j\}}(y) = \frac{1}{2} y^T A_{\{i,j\}} y$, where $A_{\{i,j\}} \in \mathbb{R}^{n\times n}$ is any symmetric positive definite matrix, or $g_{\{i,j\}}(y) = f_i(y) + f_j(y)$). Then, (29) and (31) hold. ∎

Examples 3 and 4 below show that some of the continuous-time distributed consensus algorithms in the literature are special cases of ZGS algorithms. In addition, they are just a slight modification away from solving general unconstrained, separable, convex optimization problems:

Example 3. Consider the scalar (i.e., $n = 1$) linear consensus algorithm $\dot{x}_i(t) = \sum_{j\in\mathcal{N}_i} a_{ij}(x_j(t) - x_i(t))$ $\forall t \geq 0$ $\forall i \in \mathcal{V}$ with symmetric parameters $a_{ij} = a_{ji} > 0$ $\forall \{i, j\} \in \mathcal{E}$ and arbitrary initial states $x_i(0) = y_i$ $\forall i \in \mathcal{V}$, studied in [12]–[14], [16]. By Definition 1 and Theorem 1, this algorithm is a ZGS algorithm that solves problem (8) for $f_i(x) = \frac{1}{2}(x - y_i)^2$ $\forall i \in \mathcal{V}$. Moreover, the algorithm is only a Hessian inverse and an initial condition away (i.e., $\dot{x}_i(t) = (\nabla^2 f_i(x_i(t)))^{-1} \sum_{j\in\mathcal{N}_i} a_{ij}(x_j(t) - x_i(t))$ with $x_i(0) = x_i^*$) from solving *any* convex optimization problem of the form (8) for any $n \geq 1$. Note that the same can be said about the scalar nonlinear consensus protocol in [11]. ∎

Example 4. Consider the multivariable (i.e., $n \geq 1$) weighted-average consensus algorithm $\dot{x}_i(t) = W_i^{-1} \sum_{j\in\mathcal{N}_i} (x_j(t) - x_i(t))$ $\forall t \geq 0$ $\forall i \in \mathcal{V}$ with $W_i = W_i^T > 0$ and $x_i(0) = y_i$, proposed in [15] as a step toward a distributed Kalman filter. This algorithm is a ZGS algorithm that solves problem (8) for $f_i(x) = \frac{1}{2}(x - y_i)^T W_i (x - y_i)$ $\forall i \in \mathcal{V}$. Indeed, it came close to solving for general f_i's. ∎

V. Convergence Rate Analysis

In this section, we derive lower bounds on the exponential convergence rates of a class of ZGS algorithms.

Reconsider the class of ZGS algorithms specified in Example 2, which takes the form

$$\dot{x}_i(t) = (\nabla^2 f_i(x_i(t)))^{-1} \sum_{j\in\mathcal{N}_i} \nabla g_{\{i,j\}}(x_j(t)) - \nabla g_{\{i,j\}}(x_i(t)), \quad \forall t \geq 0,\ \forall i \in \mathcal{V}. \tag{32}$$

Suppose the initial state $\mathbf{x}(0)$ is given, which may simply be $\mathbf{x}(0) = (x_1^*, x_2^*, \ldots, x_N^*)$ according to (22). To analyze the convergence behavior of this class of ZGS algorithms, let $\mathcal{C}_i = \{x \in \mathbb{R}^n : f_i(x^*) - f_i(x) - \nabla f_i^T(x)(x^* - x) \leq V(\mathbf{x}(0))\}$ $\forall i \in \mathcal{V}$ and let $\mathcal{C} = \operatorname{conv} \cup_{i\in\mathcal{V}} \mathcal{C}_i$, where conv denotes the convex hull of a set. Due to the strong convexity of each f_i and (5), each $\mathcal{C}_i$ is compact. Hence, $\mathcal{C}$ is convex and compact. Because of (iii) in Theorem 1, $V(\mathbf{x}(t))$ is nonincreasing. This, along with (12), implies that

$$x_i(t), x^* \in \mathcal{C}_i \subset \mathcal{C}, \quad \forall t \geq 0,\ \forall i \in \mathcal{V}. \tag{33}$$

Due to Assumption 1 and (4), for each $i \in \mathcal{V}$, there exists a $\Theta_i \geq \theta_i > 0$ such that

$$\nabla^2 f_i(x) \leq \Theta_i I_n, \quad \forall x \in \mathcal{C}. \tag{34}$$

In addition, for each $\{i, j\} \in \mathcal{E}$, because $g_{\{i,j\}}$ is locally strongly convex and because of (3), there exists a $\gamma_{\{i,j\}} > 0$ such that

$$(\nabla g_{\{i,j\}}(y) - \nabla g_{\{i,j\}}(x))^T (y - x) \geq \gamma_{\{i,j\}} \|y - x\|^2, \quad \forall x, y \in \mathcal{C}. \tag{35}$$

Note that Θ_i $\forall i \in \mathcal{V}$ and $\gamma_{\{i,j\}}$ $\forall \{i, j\} \in \mathcal{E}$ depend on the set $\mathcal{C}$, which in turn depends on the initial state $\mathbf{x}(0)$. This suggests that the convergence rate results presented below are dependent on $\mathbf{x}(0)$.

The following theorem establishes the exponential convergence of the aforementioned class of ZGS algorithms and provides a lower bound on their convergence rates:

Theorem 2. *Consider the network modeled in Section III and the use of a ZGS algorithm in the form of* (32)*. Suppose Assumption 1 holds. Then,*

$$V(\mathbf{x}(t)) \leq V(\mathbf{x}(0)) e^{-\rho t}, \quad \forall t \geq 0, \tag{36}$$

$$\sum_{i\in\mathcal{V}} \theta_i \|x_i(t) - x^*\|^2 \leq \sum_{i\in\mathcal{V}} \Theta_i \|x_i(0) - x^*\|^2 e^{-\rho t}, \quad \forall t \geq 0, \tag{37}$$

where $\rho = \sup\{\varepsilon \in \mathbb{R} : \varepsilon P \leq Q\} > 0$, $P = [P_{ij}] \in \mathbb{R}^{N\times N}$ is a positive semidefinite matrix given by

$$P_{ij} = \begin{cases} (\frac{1}{2} - \frac{1}{N})\Theta_i + \frac{1}{2N^2} \sum_{\ell\in\mathcal{V}} \Theta_\ell, & \textit{if } i = j, \\ -\frac{\Theta_i + \Theta_j}{2N} + \frac{1}{2N^2} \sum_{\ell\in\mathcal{V}} \Theta_\ell, & \textit{otherwise}, \end{cases}$$

and $Q = [Q_{ij}] \in \mathbb{R}^{N\times N}$ is a positive semidefinite matrix given by

$$Q_{ij} = \begin{cases} \sum_{\ell\in\mathcal{N}_i} \gamma_{\{i,\ell\}}, & \textit{if } i = j, \\ -\gamma_{\{i,j\}}, & \textit{if } \{i, j\} \in \mathcal{E}, \\ 0, & \textit{otherwise}. \end{cases}$$

Proof. Due to space limitation, we provide only a sketch of the proof here. Since $F(x^*) \leq F(\eta)$ $\forall \eta \in \mathbb{R}^n$ and $\mathbf{x}(t) \in \mathcal{M}$ $\forall t \geq 0$, $V(\mathbf{x}(t)) \leq \sum_{i\in\mathcal{V}} f_i(\frac{1}{N}\sum_{j\in\mathcal{V}} x_j(t)) - f_i(x_i(t)) - \nabla f_i^T(x_i(t))(\frac{1}{N}\sum_{j\in\mathcal{V}} x_j(t) - x_i(t))$ $\forall t \geq 0$. Due to (33) and the convexity of $\mathcal{C}$, $\frac{1}{N}\sum_{j\in\mathcal{V}} x_j(t) \in \mathcal{C}$. It follows from the strong convexity of each f_i, (34), (5), and (7) that

$$\begin{aligned} V(\mathbf{x}(t)) &\leq \sum_{i\in\mathcal{V}} \frac{\Theta_i}{2} \|x_i(t) - \frac{1}{N}\sum_{j\in\mathcal{V}} x_j(t)\|^2 \\ &= \mathbf{x}(t)^T (P \otimes I_n) \mathbf{x}(t), \quad \forall t \geq 0, \end{aligned} \tag{38}$$

where $\otimes$ denotes the Kronecker product. Moreover, due to (32), (9), and (19), $\dot{V}(\mathbf{x}(t)) = \frac{1}{2}\sum_{i\in\mathcal{V}} \sum_{j\in\mathcal{N}_i} (x_i(t) -$

$x_j(t))^T(\nabla g_{\{i,j\}}(x_j(t)) - \nabla g_{\{i,j\}}(x_i(t)))\ \forall t \geq 0$. This, along with (35) and (33), implies that

$$-\dot{V}(\mathbf{x}(t)) \geq \frac{1}{2}\sum_{i\in\mathcal{V}}\sum_{j\in\mathcal{N}_i}\gamma_{\{i,j\}}\|x_i(t) - x_j(t)\|^2$$
$$= \mathbf{x}(t)^T(Q\otimes I_n)\mathbf{x}(t), \quad \forall t \geq 0. \tag{39}$$

Based on (38) and (39), it can be shown that $\rho V(\mathbf{x}(t)) \leq -\dot{V}(\mathbf{x}(t))\ \forall t \geq 0$, where $\rho = \sup\{\varepsilon \in \mathbb{R} : \varepsilon P \leq Q\} > 0$, so that (36) holds. Finally, from the strong convexity of each f_i, (2), (36), (34), (33), (5), and (7), we obtain (37). □

Notice that the lower bound ρ on the convergence rate in Theorem 2 can be calculated by forming the matrices P and Q, finding the largest $\varepsilon > 0$ such that $\varepsilon P \leq Q$, and setting ρ to this largest ε. The following corollary to Theorem 2 presents a more conservative lower bound on the convergence rate, which, however, makes the algebraic connectivity of the network explicit in the result:

Corollary 1. *Consider the network modeled in Section III and the use of a ZGS algorithm in the form of* (32). *Suppose Assumption 1 holds. Then,*

$$V(\mathbf{x}(t)) \leq V(\mathbf{x}(0))e^{-\frac{2\gamma}{\Theta}\lambda_2 t}, \quad \forall t \geq 0, \tag{40}$$

$$\|\mathbf{x}(t) - \mathbf{x}^*\| \leq \sqrt{\frac{\Theta}{\theta}}\|\mathbf{x}(0) - \mathbf{x}^*\|e^{-\frac{\gamma}{\Theta}\lambda_2 t}, \quad \forall t \geq 0, \tag{41}$$

where $\gamma = \min_{\{i,j\}\in\mathcal{E}}\gamma_{\{i,j\}}$, $\Theta = \max_{i\in\mathcal{V}}\Theta_i$, $\theta = \min_{i\in\mathcal{V}}\theta_i$, *and* $\lambda_2 > 0$ *is the algebraic connectivity of the graph* $\mathcal{G}$.

Proof. Again, only a sketch of the proof is provided. From (38) and (39), we have $V(\mathbf{x}(t)) \leq \frac{\Theta}{2}\sum_{i\in\mathcal{V}}\|x_i(t) - \frac{1}{N}\sum_{j\in\mathcal{V}}x_j(t)\|^2 = \frac{\Theta}{2N}\mathbf{x}(t)^T(\mathcal{L}_{\bar{\mathcal{G}}}\otimes I_n)\mathbf{x}(t)\ \forall t \geq 0$ and $-\dot{V}(\mathbf{x}(t)) \geq \frac{\gamma}{2}\sum_{i\in\mathcal{V}}\sum_{j\in\mathcal{N}_i}\|x_i(t) - x_j(t)\|^2 = \gamma\mathbf{x}(t)^T(\mathcal{L}_{\mathcal{G}}\otimes I_n)\mathbf{x}(t)\ \forall t \geq 0$, where $\mathcal{L}_{\bar{\mathcal{G}}} \in \mathbb{R}^{N\times N}$ is the Laplacian matrix of the complete graph $\bar{G}$ whose vertex set is $\mathcal{V}$ and $\mathcal{L}_{\mathcal{G}} \in \mathbb{R}^{N\times N}$ is the Laplacian matrix of $\mathcal{G}$. Let $\lambda_1, \lambda_2, \ldots, \lambda_N$ with $\lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_N$ be the eigenvalues of $\mathcal{L}_{\mathcal{G}}$. Note that $\lambda_1 = 0$ and $\lambda_2 > 0$. Let $w_1, w_2, \ldots, w_N$ be the corresponding normalized eigenvectors and $W = (w_1, w_2, \ldots, w_N) \in \mathbb{R}^{N\times N}$. Then, $W^T\mathcal{L}_{\mathcal{G}}W = \operatorname{diag}(0, \lambda_2, \ldots, \lambda_N)$ and $W^T\mathcal{L}_{\bar{\mathcal{G}}}W = \operatorname{diag}(0, N, \ldots, N)$. Hence, $\lambda_2 W^T\mathcal{L}_{\bar{\mathcal{G}}}W \leq NW^T\mathcal{L}_{\mathcal{G}}W$, i.e., $\frac{2\gamma}{\Theta}\lambda_2\cdot\frac{\Theta}{2N}\mathcal{L}_{\bar{\mathcal{G}}} \leq \gamma\mathcal{L}_{\mathcal{G}}$. It follows that (40) and (41) hold. □

Notice that in the special case where (32) reduces to the scalar linear consensus algorithm described in Example 3, i.e., $f_i(x) = \frac{1}{2}(x - y_i)^2\ \forall i \in \mathcal{V}$, and where $g_{\{i,j\}}(x) = \frac{1}{2}a_{\{i,j\}}x^2$ with $a_{\{i,j\}} > 0\ \forall\{i,j\} \in \mathcal{E}$, we may let θ_i, $\Theta_i\ \forall i \in \mathcal{V}$ and $\gamma_{\{i,j\}}\ \forall\{i,j\} \in \mathcal{E}$ all be 1. In this case, Theorem 2 and Corollary 1 lead to the same lower bound on the convergence rate, i.e., $\|\mathbf{x}(t) - \mathbf{x}^*\| \leq \|\mathbf{x}(0) - \mathbf{x}^*\|e^{-\lambda_2 t}$ $\forall t \geq 0$, which coincides with the well-known convergence rate result, obtained in [12], for the linear consensus algorithm. Therefore, Theorem 2 and Corollary 1 may be regarded as a generalization of the convergence rate result for distributed consensus to distributed convex optimization.

VI. Conclusion

In this paper, using a convexity-based Lyapunov function candidate, we have developed and analyzed a family of continuous-time distributed algorithms, which solve a class of convex optimization problems over undirected networks with fixed topologies. Referred to as ZGS algorithms, we have shown that they produce nonlinear networked dynamical systems, whose states remain on a zero-gradient-sum manifold and move asymptotically toward the unknown minimizer. We have also provided concrete ways to construct ZGS algorithms and obtained, for a subset of them, lower bounds on their exponential convergence rates in terms of the convexity characteristics of the problem and the network topology. Finally, we have shown that ZGS algorithms may be viewed as a generalization of certain existing continuous-time distributed consensus algorithms, thereby providing a connection between distributed consensus and distributed convex optimization.

References

[1] M. G. Rabbat and R. D. Nowak, "Distributed optimization in sensor networks," in *Proc. International Symposium on Information Processing in Sensor Networks*, Berkeley, CA, 2004, pp. 20–27.

[2] D. Blatt, A. O. Hero, and H. Gauchman, "A convergent incremental gradient method with a constant step size," *SIAM Journal on Optimization*, vol. 18, no. 1, pp. 29–51, 2007.

[3] B. Johansson, M. Rabi, and M. Johansson, "A simple peer-to-peer algorithm for distributed optimization in sensor networks," in *Proc. IEEE Conference on Decision and Control*, New Orleans, LA, 2007, pp. 4705–4710.

[4] S. S. Ram, A. Nedić, and V. V. Veeravalli, "Incremental stochastic subgradient algorithms for convex optimization," *SIAM Journal on Optimization*, vol. 20, no. 2, pp. 691–717, 2009.

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

[6] 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.

[7] S. S. Ram, A. Nedić, and V. V. Veeravalli, "Distributed stochastic subgradient projection algorithms for convex optimization," *Journal of Optimization Theory and Applications*, vol. 147, no. 3, pp. 516–545, 2010.

[8] ——, "Asynchronous gossip algorithms for stochastic optimization," in *Proc. IEEE Conference on Decision and Control*, Shanghai, China, 2009, pp. 3581–3586.

[9] J. Lu, C. Y. Tang, P. R. Regier, and T. D. Bow, "A gossip algorithm for convex consensus optimization over networks," in *Proc. American Control Conference*, Baltimore, MD, 2010, pp. 301–308.

[10] J. Lu, P. R. Regier, and C. Y. Tang, "Control of distributed convex optimization," in *Proc. IEEE Conference on Decision and Control*, Atlanta, GA, 2010, pp. 489–495.

[11] R. Olfati-Saber and R. M. Murray, "Consensus protocols for networks of dynamic agents," in *Proc. American Control Conference*, Denver, CO, 2003, pp. 951–956.

[12] ——, "Consensus problems in networks of agents with switching topology and time-delays," *IEEE Transactions on Automatic Control*, vol. 49, no. 9, pp. 1520–1533, 2004.

[13] Y. Hatano and M. Mesbahi, "Agreement over random networks," *IEEE Transactions on Automatic Control*, vol. 50, no. 11, pp. 1867–1872, 2005.

[14] W. Ren and R. W. Beard, "Consensus seeking in multiagent systems under dynamically changing interaction topologies," *IEEE Transactions on Automatic Control*, vol. 50, no. 5, pp. 655–661, 2005.

[15] D. P. Spanos, R. Olfati-Saber, and R. M. Murray, "Distributed sensor fusion using dynamic consensus," in *Proc. IFAC World Congress*, Prague, Czech Republic, 2005.

[16] A. Tahbaz-Salehi and A. Jadbabaie, "Small world phenomenon, rapidly mixing Markov chains, and average consensus algorithms," in *Proc. IEEE Conference on Decision and Control*, New Orleans, LA, 2007, pp. 276–281.

[17] R. Goebel and R. T. Rockafellar, "Local strong convexity and local lipschitz continuity of the gradient of convex functions," *Journal of Convex Analysis*, vol. 15, no. 2, pp. 263–270, 2008.

[18] Y. Nesterov, *Introductory Lectures on Convex Optimization: A Basic Course*. Norwell, MA: Kluwer Academic Publishers, 2004.

[19] S. Boyd and L. Vandenberghe, *Convex Optimization*. New York, NY: Cambridge University Press, 2004.