

Continuous-time distributed convex optimization on weight-balanced digraphs

Bahman Ghahesifard Jorge Cortés

Abstract—This paper studies the continuous-time distributed optimization of a sum of convex functions over directed graphs. Contrary to what is known in the consensus literature, where the same dynamics works for both undirected and directed scenarios, we show that the consensus-based dynamics that solves the continuous-time distributed optimization problem for undirected graphs fails to converge when transcribed to the directed setting. This study sets the basis for the design of an alternative distributed dynamics which we show is guaranteed to converge, on any strongly connected weight-balanced digraph, to the set of minimizers of a sum of convex differentiable functions with globally Lipschitz gradients. Our technical approach combines notions of invariance and cocoercivity with the positive definiteness properties of graph matrices to establish the results.

I. INTRODUCTION

Distributed optimization of a sum of convex functions has applications in a variety of scenarios, including sensor networks, source localization, and robust estimation, and has been intensively studied in recent years, see e.g. [1], [2], [3], [4], [5], [6]. Most of these works build on consensus-based dynamics [7], [8], [9], [10] to design discrete-time algorithms that find the solution of the optimization problem. A recent exception are the works [11], [12] that deal with continuous-time distributed optimization on undirected networks. This paper further contributes to this body of work by studying continuous-time algorithms for distributed optimization in directed scenarios.

The unidirectional information flow among agents characteristic of directed networks often leads to significant technical challenges when establishing convergence and robustness properties of coordination algorithms. The results of this paper provide one more example in support of this assertion for the case of continuous-time consensus-based distributed optimization. This is somewhat surprising given that, for consensus, the same dynamics works for both undirected connected graphs and strongly connected, weight-balanced directed graphs, see e.g., [7], [8].

Statement of contributions: The contributions of this paper are the following. We first show that the solutions of the optimization problem of a sum of locally Lipschitz convex functions over a directed graph correspond to the saddle points of an aggregate objective function that depends on the graph topology through its Laplacian. This function is convex in its first argument and linear in the second. Moreover, its gradient is distributed when the graph is undirected. Next,

we consider the optimization problem over directed graphs, where we start by providing an example of a strongly connected, weight-balanced directed graph where the distributed version of the saddle-point dynamics does not converge. This motivates us to introduce a generalization of the dynamics that incorporates a design parameter. Our technical analysis establishes that, when the original functions are differentiable and convex with globally Lipschitz gradients, the design parameter can be appropriately chosen so that the resulting dynamics asymptotically converges to the set of minimizers of the objective function on any strongly connected and weight-balanced digraph. Our technical approach combines notions and tools from stability analysis, algebraic graph theory, and convex analysis. In particular, we also prove as an auxiliary result for our main derivations that any locally Lipschitz function with globally Lipschitz generalized gradient is differentiable.

Organization: Section II includes basic preliminaries on analysis, dynamical systems, and graph theory. In Section III, we review the continuous-time distributed optimization problem. Section IV reviews the convergence properties of the continuous-time distributed optimization dynamics for undirected networks. Section V considers the optimization problem over directed graphs. Section V-A presents an example of a strongly connected, weight-balanced directed graph where the distributed version of the saddle-point dynamics does not converge. Section V-B then introduces a distributed optimization dynamics on weight-balanced digraphs and characterizes its convergence properties. Finally, Section VI contains our conclusions and ideas for future work.

Some of the proofs are omitted for reasons of space and will appear elsewhere.

II. PRELIMINARIES

We start with some notational conventions. Let \mathbb{R} , $\mathbb{R}_{\geq 0}$, \mathbb{Z} , $\mathbb{Z}_{\geq 1}$ denote the set of real, nonnegative real, integer, and positive integer numbers, respectively. We denote by $\|\cdot\|$ the Euclidean norm on \mathbb{R}^d , $d \in \mathbb{Z}_{\geq 1}$ and also use the shorthand notation $\mathbf{1}_d = (1, \dots, 1)^T$ and $\mathbf{0}_d = (0, \dots, 0)^T \in \mathbb{R}^d$. We let I_d denote the identity matrix in $\mathbb{R}^{d \times d}$. For matrices $A \in \mathbb{R}^{d_1 \times d_2}$ and $B \in \mathbb{R}^{e_1 \times e_2}$, $d_1, d_2, e_1, e_2 \in \mathbb{Z}_{\geq 1}$, we let $A \otimes B$ denote their Kronecker product.

A. Basic notions from analysis

A function $f : X_1 \times X_2 \rightarrow \mathbb{R}$, with $X_1 \subset \mathbb{R}^{d_1}$, $X_2 \subset \mathbb{R}^{d_2}$ closed and convex, is *concave-convex* if it is concave in its first argument and convex in the second one [13]. A point

Bahman Ghahesifard and Jorge Cortés are with the Department of Mechanical and Aerospace Engineering, University of California, San Diego, {bghahesifard,cortes}@ucsd.edu.

$(x_1^*, x_2^*) \in X_1 \times X_2$ is a *saddle point* of f if $f(x_1, x_2^*) \leq f(x_1^*, x_2^*) \leq f(x_1^*, x_2)$ for all $x_1 \in X_1$ and $x_2 \in X_2$.

A function $f : \mathbb{R}^d \rightarrow \mathbb{R}$ is *locally Lipschitz* at $x \in \mathbb{R}^d$ if there exists a neighborhood \mathcal{U} of x and $C_x \in \mathbb{R}_{\geq 0}$ such that $|f(y) - f(z)| \leq C_x \|y - z\|$, for $y, z \in \mathcal{U}$. f is locally Lipschitz on \mathbb{R}^d if it is locally Lipschitz at x for all $x \in \mathbb{R}^d$ and *globally Lipschitz* on \mathbb{R}^d if there exists $C \in \mathbb{R}_{\geq 0}$ such that $|f(y) - f(z)| \leq C \|y - z\|$ for all $y, z \in \mathbb{R}^d$.

For a differentiable function f , a point $x \in \mathbb{R}^d$ with $\nabla f(x) = 0$ is a *critical point* of f . A differentiable convex function f satisfies, for all $x, x' \in \mathbb{R}^d$, the *first-order condition* of convexity,

$$f(x') - f(x) \geq \nabla f(x) \cdot (x' - x). \quad (1)$$

The notion of cocoercivity [14] plays a key role in our technical approach later. For $\delta \in \mathbb{R}_{>0}$, a differentiable function f is δ -*cocoercive* if, for all $x, x' \in \mathbb{R}^d$,

$$(x - x')^T (\nabla f(x) - \nabla f(x')) \geq \delta \|\nabla f(x) - \nabla f(x')\|.$$

The next result [14, Lemma 6.7] characterizes cocoercive differentiable convex functions.

Proposition 2.1: (Characterization of cocoercivity): Let f be a differentiable convex function. Then, ∇f is globally Lipschitz with constant $K \in \mathbb{R}_{>0}$ iff f is $\frac{1}{K}$ -cocoercive.

B. Stability analysis

Here, we recall some background on continuous-time dynamical systems following [15]. Consider a system on $X \subset \mathbb{R}^d$ given by

$$\dot{x}(t) = \Psi(x(t)), \quad (2)$$

where $t \in \mathbb{R}_{\geq 0}$ and $\Psi : X \subset \mathbb{R}^d \rightarrow \mathbb{R}^d$ is continuous. A solution to this dynamical system is a continuously differentiable curve $x : [0, T] \rightarrow X$ which satisfies (2). The set of equilibria of (2) is denoted by $\text{Eq}(\Psi) = \{x \in X \mid \Psi(x) = 0\}$.

The LaSalle Invariance Principle for continuous-time systems is helpful to establish the asymptotic stability properties of systems of the form (2). A set $W \subset X$ is *positively invariant* with respect to Ψ if each solution with initial condition in W remains in W for all subsequent times. The *Lie derivative* of a continuously differentiable function $V : \mathbb{R}^d \rightarrow \mathbb{R}$ along Ψ at $x \in \mathbb{R}^d$ is defined by $\mathcal{L}_\Psi V(x) = \nabla V(x) \cdot \Psi(x)$.

Theorem 2.2: (LaSalle Invariance Principle): Let $W \subset X$ be a positively invariant under (2) and $V : X \rightarrow \mathbb{R}$ a continuously differentiable function. Suppose the evolutions of (2) with initial conditions in W are bounded. Then any solution $x(t)$, $t \in \mathbb{R}_{\geq 0}$, starting in W converges to the largest positively invariant set M contained in $S_{\Psi, V} \cap W$, where $S_{\Psi, V} = \{x \in X \mid \mathcal{L}_\Psi V(x) = 0\}$. When M is a finite collection of points, then the limit of each solution equals one of them.

C. Graph theory

We present some basic notions from algebraic graph theory following the exposition in [9]. A *directed graph*, or simply *digraph*, is a pair $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, where \mathcal{V} is a finite set called the vertex set and $\mathcal{E} \subseteq \mathcal{V} \times \mathcal{V}$ is the edge set. A digraph is *undirected* if $(v, u) \in \mathcal{E}$ anytime $(u, v) \in \mathcal{E}$. We refer to an undirected digraph as a *graph*. A path is an ordered sequence of vertices such that any ordered pair of vertices appearing consecutively is an edge of the digraph. A digraph is *strongly connected* if there is a path between any pair of distinct vertices. For a graph, we refer to this notion simply as *connected*. A *weighted digraph* is a triplet $\mathcal{G} = (\mathcal{V}, \mathcal{E}, A)$, where $(\mathcal{V}, \mathcal{E})$ is a digraph and $A \in \mathbb{R}_{\geq 0}^{n \times n}$ is the *adjacency matrix* of \mathcal{G} , with the property that $a_{ij} > 0$ if $(v_i, v_j) \in \mathcal{E}$ and $a_{ij} = 0$, otherwise. The weighted out-degree and in-degree of v_i , $i \in \{1, \dots, n\}$, are respectively, $d_{\text{out}}^w(v_i) = \sum_{j=1}^n a_{ij}$ and $d_{\text{in}}^w(v_i) = \sum_{j=1}^n a_{ji}$. The *weighted out-degree matrix* D_{out} is the diagonal matrix defined by $(D_{\text{out}})_{ii} = d_{\text{out}}^w(v_i)$, for all $i \in \{1, \dots, n\}$. The *Laplacian* matrix is $L = D_{\text{out}} - A$. Note that $L \mathbf{1}_n = 0$. If \mathcal{G} is strongly connected, then zero is a simple eigenvalue of L . \mathcal{G} is undirected if $L = L^T$ and *weight-balanced* if $d_{\text{out}}^w(v) = d_{\text{in}}^w(v)$, for all $v \in \mathcal{V}$. Equivalently, \mathcal{G} is weight-balanced if and only if $\mathbf{1}_n^T L = 0$ if and only if $L + L^T$ is positive semidefinite. Furthermore, if \mathcal{G} is weight-balanced and strongly connected, then zero is a simple eigenvalue of $L + L^T$. Note that any undirected graph is weight-balanced.

III. PROBLEM STATEMENT AND EQUIVALENT FORMULATIONS

Consider a network composed by $n \in \mathbb{Z}_{\geq 1}$ agents v_1, \dots, v_n whose communication topology is described by strongly connected digraph \mathcal{G} . For each, $i \in \{1, \dots, n\}$, let $f^i : \mathbb{R}^d \rightarrow \mathbb{R}$ be continuously differentiable and convex, and only available to agent v_i . The network objective is to solve the following optimization problem in a distributed way,

$$\text{minimize } f(x) = \sum_{i=1}^n f^i(x). \quad (3)$$

Let $x^i \in \mathbb{R}^d$ denote the estimate of agent v_i about the value of the solution to (3) and define $\mathbf{x}^T = ((x^1)^T, \dots, (x^n)^T) \in \mathbb{R}^{nd}$. The following result provides an alternative formulation of (3).

Lemma 3.1: Let $L \in \mathbb{R}^{n \times n}$ be the Laplacian of \mathcal{G} and define $\mathbf{L} = L \otimes I_d \in \mathbb{R}^{nd \times nd}$. The problem (3) on \mathbb{R}^d is equivalent to the following problem on \mathbb{R}^{nd} ,

$$\text{minimize } \tilde{f}(\mathbf{x}) = \sum_{i=1}^n f^i(x^i), \quad (4a)$$

$$\text{subject to } \mathbf{L}\mathbf{x} = \mathbf{0}_{nd}. \quad (4b)$$

Proof: The proof follows by noting that (i) $\tilde{f}(\mathbf{1}_n \otimes x) = f(x)$ for all $x \in \mathbb{R}^d$ and (ii) since \mathcal{G} is strongly connected, $\mathbf{L}\mathbf{x} = \mathbf{0}_{nd}$ if and only if $\mathbf{x} = \mathbf{1}_n \otimes x$, for some $x \in \mathbb{R}^d$. ■

The formulation (4) is appealing because it brings together the estimates of each agent about the value of the solution to the original optimization problem. It is worth mentioning that

\tilde{f} is continuously differentiable and convex. Moreover, its gradient is of the form $\nabla \tilde{f}(\mathbf{x}) = (\nabla f^1(x^1), \dots, \nabla f^n(x^n))$. Since \tilde{f} is convex and the constraints in (4) are linear, the constrained optimization problem is feasible. The following result provides an equivalent formulation based on augmented Lagrangian techniques [16].

Proposition 3.2: (Solutions of the distributed optimization problem as saddle points): Let \mathcal{G} be strongly connected and weight-balanced, and define $F : \mathbb{R}^{nd} \times \mathbb{R}^{nd} \rightarrow \mathbb{R}$ by

$$F(\mathbf{x}, \mathbf{z}) = \tilde{f}(\mathbf{x}) + \mathbf{z}^T \mathbf{L}\mathbf{x} + \frac{1}{2} \mathbf{x}^T \mathbf{L}\mathbf{x}. \quad (5)$$

Then F is continuously differentiable and convex in its first argument and linear in its second, and

- (i) if $(\mathbf{x}^*, \mathbf{z}^*)$ is a saddle point of F , then so is $(\mathbf{x}^*, \mathbf{z}^* + \mathbf{1}_n \otimes a)$, for any $a \in \mathbb{R}^d$.
- (ii) if $(\mathbf{x}^*, \mathbf{z}^*)$ is a saddle point of F , then \mathbf{x}^* is a solution of (4).
- (iii) if \mathbf{x}^* is a solution of (4), then there exists \mathbf{z}^* with $\mathbf{L}\mathbf{z}^* = -\nabla \tilde{f}(\mathbf{x}^*)$ such that $(\mathbf{x}^*, \mathbf{z}^*)$ is a saddle point of F .

IV. CONTINUOUS-TIME DISTRIBUTED OPTIMIZATION ON UNDIRECTED NETWORKS

Here, we review the continuous-time solution to the optimization problem proposed in [11], [12] for undirected graphs. If \mathcal{G} is undirected, the gradient of F in (5) is distributed over \mathcal{G} . Given Proposition 3.2, it is natural to consider the saddle-point dynamics of F to solve (3),

$$\dot{\mathbf{x}} + \mathbf{L}\mathbf{x} + \mathbf{L}\mathbf{z} = -\nabla \tilde{f}(\mathbf{x}), \quad (6a)$$

$$\dot{\mathbf{z}} = \mathbf{L}\mathbf{x}. \quad (6b)$$

This dynamics is distributed over \mathcal{G} , in the sense that agent v_i can update x^i and z^i with knowledge of its own state and the state of its network neighbors. From Proposition 3.2, if $(\mathbf{x}^*, \mathbf{z}^*)$ is an equilibrium of (6), then \mathbf{x}^* is a solution to (4). According to [12], the dynamics (6) leads the network to agree on a global minimum of f for the case when \mathcal{G} is undirected and f is both strictly convex and the sum of continuously differentiable convex functions. This result, in fact, also holds true when \mathcal{G} is undirected and f is the sum of locally Lipschitz convex functions, see [17].

Theorem 4.1: (Asymptotic convergence of (6) on undirected networks): Let \mathcal{G} be a connected graph and consider the distributed optimization problem (3), where each f^i , $i \in \{1, \dots, n\}$ is continuously differentiable and convex. Then, the projection onto the first component of any trajectory of (6) asymptotically converges to the set of solutions to (4). Moreover, if f has a finite number of critical points, the limit of the projection onto the first component of each trajectory is a solution of (4).

V. CONTINUOUS-TIME DISTRIBUTED OPTIMIZATION ON DIRECTED NETWORKS

In this section, we consider the distributed optimization problem (3) on directed graphs. Note that, when \mathcal{G} is

directed, the gradient of F defined in (5) is no longer distributed over \mathcal{G} , and indeed the dynamics (6) does no longer correspond to the saddle-point dynamics. Nevertheless, it is natural to study whether the dynamics (6) enjoys the same convergence properties as in the undirected situation (as, for instance, is the case in the agreement problem [7], [8], [9], [10]). Surprisingly, this turns out not to be the case, as shown in Section V-A. This result motivates the introduction in Section V-B of an alternative provably correct dynamics on directed graphs.

A. Counterexample

Here, we provide an example of a strongly connected, weight-balanced digraph on which (6) fails to converge. For convenience, we let $\mathcal{S}_{\text{agree}} = \{(\mathbf{1}_n \otimes x, \mathbf{1}_n \otimes z) \in \mathbb{R}^{nd} \times \mathbb{R}^{nd} \mid x, z \in \mathbb{R}^d\}$ denote the set of agreement configurations. Our construction relies on the following result.

Lemma 5.1: (Necessary condition for the convergence of (6) on digraphs): Let \mathcal{G} be a strongly connected digraph and $f^i = 0$, $i \in \{1, \dots, n\}$. Then $\mathcal{S}_{\text{agree}}$ is stable under (6) iff, for any nonzero eigenvalue λ of the Laplacian L , one has $\sqrt{3}|\text{Im}(\lambda)| \leq \text{Re}(\lambda)$.

The next example shows that the criterium of Lemma 5.1 can fail even for strongly connected weight-balanced digraphs.

Example 5.2: Consider the strongly connected, weight-balanced digraph with

$$A = \begin{pmatrix} 0 & 0.5326 & 0.1654 & 0.0004 & 0.0002 \\ 0.0595 & 0 & 0.6676 & 0.0681 & 0.1230 \\ 0.0213 & 0.0004 & 0 & 0.5809 & 0.3181 \\ 0.0248 & 0.2458 & 0 & 0 & 0.5587 \\ 0.5930 & 0.1394 & 0.0877 & 0.1799 & 0 \end{pmatrix}$$

as adjacency matrix. Note that $\lambda = 0.8833 \pm 0.5197i$ is an eigenvalue of the Laplacian. Since $\sqrt{3}|\text{Im}(\lambda)| - \text{Re}(\lambda) = 0.0171 > 0$, Lemma 5.1 implies that (6) fails to converge. •

Lemma 5.1, together with Example 5.2, motivates the search for alternative dynamics to solve the optimization problem (3) on directed graphs in a distributed way.

B. Provably correct distributed dynamics on directed graphs

Here, given the result in Section V-A, we introduce an alternative continuous-time distributed dynamics for strongly connected weight-balanced digraphs. Let $\alpha \in \mathbb{R}_{>0}$ and consider the dynamics

$$\dot{\mathbf{x}} + \alpha \mathbf{L}\mathbf{x} + \mathbf{L}\mathbf{z} = -\nabla \tilde{f}(\mathbf{x}), \quad (7a)$$

$$\dot{\mathbf{z}} = \mathbf{L}\mathbf{x}. \quad (7b)$$

We first show that appropriate choices of α allow to circumvent the problem raised in Lemma 5.1.

Lemma 5.3: (Sufficient conditions for the convergence of (7) on digraphs with trivial objective function): Let \mathcal{G} be a strongly connected and weight-balanced digraph and $f^i = 0$, $i \in \{1, \dots, n\}$. If $\alpha \geq 2\sqrt{2}$, then $\mathcal{S}_{\text{agree}}$ is asymptotically stable under (7).

Proof: When all f_i , $i \in \{1, \dots, n\}$, are identically zero, the dynamics (7) is linear and has $\mathcal{S}_{\text{agree}}$ as equilibria. Consider the coordinate transformation from (\mathbf{x}, \mathbf{z}) to $(\mathbf{x}, \mathbf{y}) = (\mathbf{x}, \beta\mathbf{x} + \mathbf{z})$, with $\beta \in \mathbb{R}_{>0}$ to be chosen later. The dynamics can be rewritten as

$$\begin{pmatrix} \dot{\mathbf{x}} \\ \dot{\mathbf{y}} \end{pmatrix} = A \begin{pmatrix} \mathbf{x} \\ \mathbf{y} \end{pmatrix}$$

where

$$A = \begin{pmatrix} -(\alpha - \beta)\mathbf{L} & -\mathbf{L} \\ (-\beta(\alpha - \beta) + 1)\mathbf{L} & -\beta\mathbf{L} \end{pmatrix}. \quad (8)$$

Consider the candidate Lyapunov function $V(\mathbf{x}, \mathbf{y}) = \mathbf{x}^T \mathbf{x} + \mathbf{y}^T \mathbf{y}$. Its Lie derivative is the quadratic form defined by the matrix

$$Q = \mathbf{l}_{2nd} A + A^T \mathbf{l}_{2nd} = \begin{pmatrix} -(\alpha - \beta)(\mathbf{L} + \mathbf{L}^T) & -\mathbf{L} + (-\beta(\alpha - \beta) + 1)\mathbf{L}^T \\ (-\beta(\alpha - \beta) + 1)\mathbf{L} - \mathbf{L}^T & -\beta(\mathbf{L} + \mathbf{L}^T) \end{pmatrix}.$$

Select β now satisfying $\beta^2 - \alpha\beta + 2 = 0$ (this equation has a real solution if $\alpha \geq 2\sqrt{2}$). Then,

$$Q = \begin{pmatrix} -(\frac{\beta^2+2}{\beta} - \beta) & -1 \\ -1 & -\beta \end{pmatrix} \otimes (\mathbf{L} + \mathbf{L}^T). \quad (9)$$

Each eigenvalue η of Q is of the form $\eta = \lambda \frac{-(\beta^2+2) \pm \sqrt{(\beta^2+2)^2 - 4\beta^2}}{2\beta}$, where λ is an eigenvalue of $\mathbf{L} + \mathbf{L}^T$. Since \mathcal{G} is strongly connected and weight-balanced, $\mathbf{L} + \mathbf{L}^T$ is positive semidefinite with a simple eigenvalue at zero, and hence $\eta \leq 0$. By the LaSalle invariance principle, the solutions of (7) from any initial condition $(\mathbf{x}_0, \mathbf{y}_0) \in \mathbb{R}^{nd} \times \mathbb{R}^{nd}$, asymptotically converge to the set $S = \{(\mathbf{x}, \mathbf{y}) \mid Q(\mathbf{x}, \mathbf{y})^T = \mathbf{0}_{2nd}\} \cap W_{\mathbf{z}_0}$. To conclude the result, we need to show that $S \subseteq \mathcal{S}_{\text{agree}}$. This follows from noting that, for $\beta > 0$, $Q(\mathbf{x}, \mathbf{y})^T = \mathbf{0}_{2nd}$ implies that $(\mathbf{L} + \mathbf{L}^T)\mathbf{x} = \mathbf{0}_{nd}$ and $(\mathbf{L} + \mathbf{L}^T)\mathbf{y} = \mathbf{0}_{nd}$, i.e., $(\mathbf{x}, \mathbf{y}) \in \mathcal{S}_{\text{agree}}$. ■

The reason behind the introduction of the parameter α in (7) comes from the following observation: if one tries to reproduce the proof of Theorem 4.1 for a digraph, one encounters indefinite terms of the form $(\mathbf{x} - \mathbf{x}^*)^T (\mathbf{L} - \mathbf{L}^T)(\mathbf{z} - \mathbf{z}^*)$ in the Lie derivative of V , invalidating it as a Lyapunov function. However, the proof of Lemma 5.3 shows that an appropriate choice of α , together with a suitable change of coordinates, makes the quadratic form defined by the identity matrix a valid Lyapunov function. We next build on these observations to establish our main result: the dynamics (7) solves in a distributed way the optimization problem (3) on strongly connected weight-balanced digraphs.

Theorem 5.4: (Asymptotic convergence of (7) on directed networks): Let \mathcal{G} be a strongly connected, weight-balanced digraph and consider the distributed optimization problem (3), where each f^i , $i \in \{1, \dots, n\}$, is convex and differentiable with globally Lipschitz continuous gradients. Let $K \in \mathbb{R}_{>0}$ be the Lipschitz constant of $\nabla \tilde{f}$ and define

$h : \mathbb{R}_{>0} \rightarrow \mathbb{R}$ by

$$h(r) = \frac{1}{2} \Lambda_*(\mathbf{L} + \mathbf{L}^T) \left(-\frac{r^4 + 3r^2 + 2}{r} + \sqrt{\left(\frac{r^4 + 3r^2 + 2}{r} \right)^2 - 4} \right) + \frac{Kr^2}{(1+r^2)}, \quad (10)$$

where $\Lambda_*(\cdot)$ denotes the non-zero eigenvalue with smallest absolute value. Then, there exists $\beta^* \in \mathbb{R}_{>0}$ with $h(\beta^*) = 0$ such that, for all $0 < \beta < \beta^*$, the projection onto the first component of any trajectory of (7) with $\alpha = \frac{\beta^2+2}{\beta}$ asymptotically converges to the set of solutions of (4). Moreover, if f has a finite number of critical points, the limit of the projection onto the first component of each trajectory is a solution of (4).

Proof: For convenience, we denote the dynamics (7) by $\Psi_{\alpha\text{-dis-opt}} : \mathbb{R}^{nd} \times \mathbb{R}^{nd} \rightarrow \mathbb{R}^{nd} \times \mathbb{R}^{nd}$. Note that the equilibria of $\Psi_{\alpha\text{-dis-opt}}$ are precisely the set of saddle points of F in (5). Let $\mathbf{x}^* = \mathbf{1}_n \otimes \mathbf{x}^*$ be a solution of (4). First, note that given any initial condition $(\mathbf{x}_0, \mathbf{z}_0) \in \mathbb{R}^{nd} \times \mathbb{R}^{nd}$, the set $W_{\mathbf{z}_0}$ defined by

$$W_{\mathbf{z}_0} = \{(\mathbf{x}, \mathbf{z}) \mid (\mathbf{1}_n^T \otimes \mathbf{l}_d)\mathbf{z} = (\mathbf{1}_n^T \otimes \mathbf{l}_d)\mathbf{z}_0\},$$

is invariant under the evolutions of (7). By Proposition 3.2(i) and (iii), there exists $(\mathbf{x}^*, \mathbf{z}^*) \in \text{Eq}(\Psi_{\alpha\text{-dis-opt}}) \cap W_{\mathbf{z}_0}$. Consider the function $V : \mathbb{R}^{nd} \times \mathbb{R}^{nd} \rightarrow \mathbb{R}_{\geq 0}$,

$$V(\mathbf{x}, \mathbf{z}) = \frac{1}{2} (\mathbf{x} - \mathbf{x}^*)^T (\mathbf{x} - \mathbf{x}^*) + \frac{1}{2} (\mathbf{y}_{(\mathbf{x}, \mathbf{z})} - \mathbf{y}_{(\mathbf{x}^*, \mathbf{z}^*)})^T (\mathbf{y}_{(\mathbf{x}, \mathbf{z})} - \mathbf{y}_{(\mathbf{x}^*, \mathbf{z}^*)}),$$

where $\mathbf{y}_{(\mathbf{x}, \mathbf{z})} = \beta\mathbf{x} + \mathbf{z}$ and $\beta \in \mathbb{R}_{>0}$ satisfies $\beta^2 - \alpha\beta + 2 = 0$. This function is quadratic, hence smooth. Next, we consider its Lie derivative along $\Psi_{\alpha\text{-dis-opt}}$ on $W_{\mathbf{z}_0}$. For $(\mathbf{x}, \mathbf{z}) \in W_{\mathbf{z}_0}$, let

$$\begin{aligned} \xi &= \mathcal{L}_{\Psi_{\alpha\text{-dis-opt}}} V(\mathbf{x}, \mathbf{z}) \\ &= (-\alpha\mathbf{L}\mathbf{x} - \mathbf{L}\mathbf{z} - \nabla \tilde{f}(\mathbf{x}), \mathbf{L}\mathbf{x}) \cdot \nabla V(\mathbf{x}, \mathbf{z}) \\ &= \frac{1}{2} ((\mathbf{x} - \mathbf{x}^*)^T, (\mathbf{y}_{(\mathbf{x}, \mathbf{z})} - \mathbf{y}_{(\mathbf{x}^*, \mathbf{z}^*)})^T) A(\mathbf{x}, \mathbf{y}_{(\mathbf{x}, \mathbf{z})})^T \\ &\quad + \frac{1}{2} (\mathbf{x}^T, \mathbf{y}_{(\mathbf{x}, \mathbf{z})}^T) A^T (\mathbf{x} - \mathbf{x}^*, \mathbf{y}_{(\mathbf{x}, \mathbf{z})} - \mathbf{y}_{(\mathbf{x}^*, \mathbf{z}^*)})^T \\ &\quad - (\mathbf{x} - \mathbf{x}^*)^T \nabla \tilde{f}(\mathbf{x}) - \beta (\mathbf{y}_{(\mathbf{x}, \mathbf{z})} - \mathbf{y}_{(\mathbf{x}^*, \mathbf{z}^*)})^T \nabla \tilde{f}(\mathbf{x}), \end{aligned}$$

where A is given by (8). This equation can be written as

$$\begin{aligned} \xi &= \frac{1}{2} ((\mathbf{x} - \mathbf{x}^*)^T, (\mathbf{y}_{(\mathbf{x}, \mathbf{z})} - \mathbf{y}_{(\mathbf{x}^*, \mathbf{z}^*)})^T) \\ &\quad Q(\mathbf{x} - \mathbf{x}^*, \mathbf{y}_{(\mathbf{x}, \mathbf{z})} - \mathbf{y}_{(\mathbf{x}^*, \mathbf{z}^*)})^T - (\mathbf{x} - \mathbf{x}^*)^T \nabla \tilde{f}(\mathbf{x}) \\ &\quad + ((\mathbf{x} - \mathbf{x}^*)^T, (\mathbf{y}_{(\mathbf{x}, \mathbf{z})} - \mathbf{y}_{(\mathbf{x}^*, \mathbf{z}^*)})^T) A(\mathbf{x}^*, \mathbf{y}_{(\mathbf{x}^*, \mathbf{z}^*)})^T \\ &\quad - \beta (\mathbf{y}_{(\mathbf{x}, \mathbf{z})} - \mathbf{y}_{(\mathbf{x}^*, \mathbf{z}^*)})^T \nabla \tilde{f}(\mathbf{x}), \end{aligned}$$

where Q is given by (9). Note that $A(\mathbf{x}^*, \mathbf{y}_{(\mathbf{x}^*, \mathbf{z}^*)})^T = -(\mathbf{L}\mathbf{y}_{(\mathbf{x}^*, \mathbf{z}^*)}, \beta\mathbf{L}\mathbf{y}_{(\mathbf{x}^*, \mathbf{z}^*)})^T = (\nabla \tilde{f}(\mathbf{x}^*), \beta \nabla \tilde{f}(\mathbf{x}^*))^T$. Thus,

after substituting for $\mathbf{y}_{(\mathbf{x}, \mathbf{z})}$, we have

$$\begin{aligned} \xi &= \frac{1}{2} ((\mathbf{x} - \mathbf{x}^*)^T, (\mathbf{z} - \mathbf{z}^*)^T)^T \tilde{Q} (\mathbf{x} - \mathbf{x}^*, \mathbf{z} - \mathbf{z}^*)^T \\ &\quad - (1 + \beta^2)(\mathbf{x} - \mathbf{x}^*)^T (\nabla \tilde{f}(\mathbf{x}) - \nabla \tilde{f}(\mathbf{x}^*)) \\ &\quad - \beta(\mathbf{z} - \mathbf{z}^*)^T (\nabla \tilde{f}(\mathbf{x}) - \nabla \tilde{f}(\mathbf{x}^*)), \end{aligned} \quad (11)$$

where

$$\tilde{Q} = \begin{pmatrix} -\beta^3 - (\frac{\beta^2+2}{\beta}) - \beta & -(1 + \beta^2) \\ -(1 + \beta^2) & -\beta \end{pmatrix} \otimes (\mathbf{L} + \mathbf{L}^T).$$

Each eigenvalue of \tilde{Q} is of the form

$$\tilde{\eta} = \lambda \times \frac{-(\beta^4 + 3\beta^2 + 2) \pm \sqrt{(\beta^4 + 3\beta^2 + 2)^2 - 4\beta^2}}{2\beta}, \quad (12)$$

where λ is an eigenvalue of $\mathbf{L} + \mathbf{L}^T$. Using the cocoercivity of \tilde{f} , we can upper bound ξ as,

$$\xi \leq \frac{1}{2} X^T \underbrace{\begin{pmatrix} \tilde{Q}_{11} & \tilde{Q}_{12} & 0 \\ \tilde{Q}_{21} & \tilde{Q}_{22} & -\beta \mathbf{l}_{nd} \\ 0 & -\beta \mathbf{l}_{nd} & -\frac{1}{K}(1 + \beta^2) \mathbf{l}_{nd} \end{pmatrix}}_{\mathbf{Q}} X, \quad (13)$$

where $K \in \mathbb{R}_{>0}$ is the Lipschitz constant for the gradient of \tilde{f} and

$$X^T = ((\mathbf{x} - \mathbf{x}^*), (\mathbf{z} - \mathbf{z}^*), (\nabla \tilde{f}(\mathbf{x}) - \nabla \tilde{f}(\mathbf{x}^*))).$$

Since $(\mathbf{x}, \mathbf{z}) \in W_{\mathbf{z}_0}$, we have $(\mathbf{1}_n^T \otimes \mathbf{l}_d)(\mathbf{z} - \mathbf{z}^*) = \mathbf{0}_d$ and hence it is enough to establish that \mathbf{Q} is negative semidefinite on the subspace $\mathcal{W} = \{(v_1, v_2, v_3) \in (\mathbb{R}^{nd})^3 \mid (\mathbf{1}_n^T \otimes \mathbf{l}_d)v_2 = \mathbf{0}_d\}$. Using the fact that $-\frac{1}{K}(1 + \beta^2)\mathbf{l}_{nd}$ is invertible, we can express \mathbf{Q} as

$$\mathbf{Q} = N \begin{pmatrix} \tilde{Q} & 0 \\ 0 & -\frac{1}{K}(1 + \beta^2)\mathbf{l}_{nd} \end{pmatrix} N^T,$$

where

$$\begin{aligned} \bar{Q} &= \tilde{Q} + \frac{K\beta^2}{(1 + \beta^2)} \begin{pmatrix} 0 & 0 \\ 0 & \mathbf{l}_{nd} \end{pmatrix}, \\ N &= \begin{pmatrix} \mathbf{l}_{nd} & 0 & 0 \\ 0 & \mathbf{l}_{nd} & \frac{\beta K}{1 + \beta^2} \mathbf{l}_{nd} \\ 0 & 0 & \mathbf{l}_{nd} \end{pmatrix}. \end{aligned}$$

Noting that \mathcal{W} is invariant under N^T (i.e., $N^T \mathcal{W} = \mathcal{W}$), all we need to check is that the matrix $\begin{pmatrix} \tilde{Q} & 0 \\ 0 & -\frac{1}{K}(1 + \beta^2)\mathbf{l}_{nd} \end{pmatrix}$ is negative semidefinite on \mathcal{W} . Clearly, $-\frac{1}{K}(1 + \beta^2)\mathbf{l}_{nd}$ is negative definite. On the other hand, on $(\mathbb{R}^{nd})^2$, 0 is an eigenvalue of \tilde{Q} with multiplicity $2d$ and eigenspace generated by vectors of the form $(\mathbf{1}_n \otimes a, 0)$ and $(0, \mathbf{1}_n \otimes b)$, with $a, b \in \mathbb{R}^d$. However, on $\{(v_1, v_2) \in (\mathbb{R}^{nd})^2 \mid (\mathbf{1}_n^T \otimes \mathbf{l}_d)v_2 = \mathbf{0}_d\}$, 0 is an eigenvalue of \tilde{Q} with multiplicity d and eigenspace generated by vectors of the form $(\mathbf{1}_n \otimes a, 0)$. Moreover, on $\{(v_1, v_2) \in (\mathbb{R}^{nd})^2 \mid (\mathbf{1}_n^T \otimes \mathbf{l}_d)v_2 = \mathbf{0}_d\}$, the eigenvalues of $\frac{K\beta^2}{(1 + \beta^2)} \begin{pmatrix} 0 & 0 \\ 0 & \mathbf{l}_{nd} \end{pmatrix}$ are $\frac{K\beta^2}{(1 + \beta^2)}$ with multiplicity $nd - d$ and 0 with multiplicity nd . Therefore, using Weyl's theorem [18, Theorem 4.3.7], we deduce that the nonzero eigenvalues of

the sum \bar{Q} are upper bounded by $\Lambda_*(\bar{Q}) + \frac{K\beta^2}{(1 + \beta^2)}$. From (12) and the definition of h in (10), we conclude that the nonzero eigenvalues of \bar{Q} are upper bounded by $h(\beta)$. It remains to show that there exists $\beta^* \in \mathbb{R}_{>0}$ with $h(\beta^*) = 0$ such that for all $0 < \beta < \beta^*$ we have $h(\beta) < 0$. For $r > 0$ small enough, $h(r) < 0$, since $h(r) = -\frac{1}{2}\Lambda_*(\mathbf{L} + \mathbf{L}^T)r + O(r^2)$. Furthermore, $\lim_{r \rightarrow \infty} h(r) = K > 0$. Hence, the existence of β^* follows from the Mean Value Theorem. Therefore we conclude $\mathcal{L}_{\Psi_{\alpha\text{-dis-opt}}} V(\mathbf{x}, \mathbf{z}) \leq 0$. As a by-product, the trajectories of (7) are bounded. Consequently, all assumptions of the LaSalle Invariance Principle are satisfied and its application yields that any trajectory of (7) starting from an initial condition $(\mathbf{x}_0, \mathbf{z}_0)$ converges to the largest positively invariant set M in $S_{\Psi_{\alpha\text{-dis-opt}}, V} \cap W_{\mathbf{z}_0}$. Note that if $(\mathbf{x}, \mathbf{z}) \in S_{\Psi_{\alpha\text{-dis-opt}}, V} \cap W_{\mathbf{z}_0}$, then $N^T \begin{pmatrix} \mathbf{x} - \mathbf{x}^* \\ \mathbf{z} - \mathbf{z}^* \\ \nabla \tilde{f}(\mathbf{x}) - \nabla \tilde{f}(\mathbf{x}^*) \end{pmatrix} \in \ker(\bar{Q}) \times \{0\}$. From the discussion above, we know $\ker(\bar{Q})$ is generated by vectors of the form $(\mathbf{1}_n \otimes a, 0)$, and hence this implies that $\mathbf{x} = \mathbf{x}^* + \mathbf{1}_n \otimes a$, $\mathbf{z} = \mathbf{z}^*$, and $\nabla \tilde{f}(\mathbf{x}) = \nabla \tilde{f}(\mathbf{x}^*)$, from where we deduce that \mathbf{x} is also a solution to (4). Finally, for $(\mathbf{x}, \mathbf{z}) \in M$, an argument similar to the one in the proof of Theorem 4.1 establishes $(\mathbf{x}, \mathbf{z}) \in \text{Eq}(\Psi_{\alpha\text{-dis-opt}})$. If the set of equilibria is finite, convergence to a point is also guaranteed. ■

Figure 1 illustrates the result of Theorem 5.4 for the network of Example 5.2.

Remark 5.5: (Locally Lipschitz objective functions): Our simulations suggests that the convergence result in Theorem 5.4 holds true for any locally Lipschitz objective function. However, our proof cannot be reproduced for this case because it would rely on the generalized gradient of the function being globally Lipschitz which, by Proposition A.1, would imply that the function is continuously differentiable.

•

Remark 5.6 (Selection of α in (7)): According to Theorem 5.4, the parameter α is determined by β as $\alpha = \frac{\beta^2 + 2}{\beta}$. In turn, one can observe from (10) that the range of suitable values for β increases with higher network connectivity and smaller variability of the gradient of the objective function. From a control design viewpoint, it is reasonable to choose the value of β that yields the smallest α while satisfying the conditions of the theorem statement. •

VI. CONCLUSIONS AND FUTURE WORK

We have studied the distributed optimization of a sum of convex functions over directed networks using consensus-based dynamics. Somewhat surprisingly, we have established that the convergence results established in the literature for undirected networks do not carry over to the directed scenario. Nevertheless, our analysis has allowed us to introduce a slight generalization of the saddle-point dynamics of the undirected case which incorporates a design parameter. We have proved that, for appropriate parameter choices, this dynamics solves the distributed optimization problem for differentiable convex functions with globally Lipschitz gradients on strongly connected and weight-balanced digraphs.

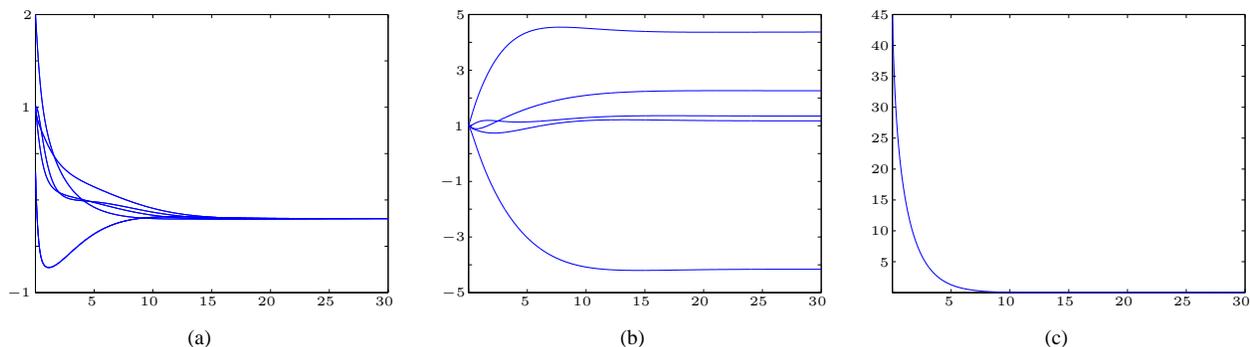


Fig. 1. Execution of (7) for the network of Example 5.2 with $f^1(x) = e^x$, $f^2(x) = (x-3)^2$, $f^3(x) = (x+3)^2$, $f^4(x) = x^4$, $f^5(x) = 4$. (a) and (b) show the evolution of the agent's values in x and z , respectively, and (c) shows the value of the Lyapunov function. Here, $\alpha = 3$, $\mathbf{x}_0 = (1, 2, 0.3, 1, 1)^T$, and $\mathbf{z}_0 = \mathbf{1}_5$. The equilibrium $(\mathbf{x}^*, \mathbf{z}^*)$ is $\mathbf{x}^* = -0.2005 \cdot \mathbf{1}_5$ and $\mathbf{z}^* = (1.1784, 4.3717, -4.1598, 2.2598, 1.3499)^T$.

Our technical approach relies on a careful combination of notions from stability analysis, algebraic graph theory, and convex analysis. Future work will focus on the extension of the convergence results to locally Lipschitz functions, the incorporation of local and global constraints, and the design of distributed algorithms that allow the network to agree on an optimal value of the design parameter.

ACKNOWLEDGMENTS

This work was supported in part by Award FA9550-10-1-0499 and NSF award CMMI-0908508.

REFERENCES

- [1] M. Rabbat and R. Nowak, "Distributed optimization in sensor networks," in *Symposium on Information Processing of Sensor Networks*, (Berkeley, CA), pp. 20–27, Apr. 2004.
- [2] A. Nedic and A. Ozdaglar, "Distributed subgradient methods for multi-agent optimization," *IEEE Transactions on Automatic Control*, vol. 54, no. 1, pp. 48–61, 2009.
- [3] P. Wan and M. D. Lemmon, "Event-triggered distributed optimization in sensor networks," in *Symposium on Information Processing of Sensor Networks*, (San Francisco, CA), pp. 49–60, 2009.
- [4] A. Nedic, A. Ozdaglar, and P. A. Parrilo, "Constrained consensus and optimization in multi-agent networks," *IEEE Transactions on Automatic Control*, vol. 55, no. 4, pp. 922–938, 2010.
- [5] B. Johansson, M. Rabi, and M. Johansson, "A randomized incremental subgradient method for distributed optimization in networked systems," *SIAM Journal on Control and Optimization*, vol. 20, no. 3, pp. 1157–1170, 2009.
- [6] M. Zhu and S. Martínez, "On distributed convex optimization under inequality and equality constraints," *IEEE Transactions on Automatic Control*, vol. 57, no. 1, pp. 151–164, 2012.
- [7] R. Olfati-Saber, J. A. Fax, and R. M. Murray, "Consensus and cooperation in networked multi-agent systems," *Proceedings of the IEEE*, vol. 95, no. 1, pp. 215–233, 2007.
- [8] W. Ren and R. W. Beard, *Distributed Consensus in Multi-vehicle Cooperative Control*. Communications and Control Engineering, Springer, 2008.
- [9] F. Bullo, J. Cortés, and S. Martínez, *Distributed Control of Robotic Networks*. Applied Mathematics Series, Princeton University Press, 2009. Electronically available at <http://coordinationbook.info>.
- [10] M. Mesbahi and M. Egerstedt, *Graph Theoretic Methods in Multiagent Networks*. Applied Mathematics Series, Princeton University Press, 2010.
- [11] J. Wang and N. Elia, "Control approach to distributed optimization," in *Allerton Conf. on Communications, Control and Computing*, (Monticello, IL), pp. 557–561, Oct. 2010.
- [12] J. Wang and N. Elia, "A control perspective for centralized and distributed convex optimization," in *IEEE Conf. on Decision and Control*, (Orlando, Florida), pp. 3800–3805, 2011.

- [13] R. T. Rockafellar, *Convex Analysis*. Princeton Landmarks in Mathematics and Physics, Princeton, NJ: Princeton University Press, 1997. Reprint of 1970 edition.
- [14] E. G. Golshtein and N. V. Tretyakov, *Modified Lagrangians and Monotone Maps in Optimization*. New York: Wiley, 1996.
- [15] H. K. Khalil, *Nonlinear Systems*. Prentice Hall, 3 ed., 2002.
- [16] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge University Press, 2004.
- [17] B. Gharesifard and J. Cortés, "Distributed convergence to Nash equilibria by adversarial networks with undirected topologies," in *American Control Conference*, (Montréal, Canada), pp. 5881–5886, 2012.
- [18] R. A. Horn and C. R. Johnson, *Matrix Analysis*. Cambridge University Press, 1985.
- [19] J. Cortés, "Discontinuous dynamical systems - a tutorial on solutions, nonsmooth analysis, and stability," *IEEE Control Systems Magazine*, vol. 28, no. 3, pp. 36–73, 2008.

APPENDIX

The *generalized gradient* of a locally Lipschitz function f is

$$\partial f(x) = \text{co} \left\{ \lim_{k \rightarrow \infty} \nabla f(x_k) \mid x_k \rightarrow x, x_k \notin \Omega_f \cup S \right\},$$

where Ω_f is the set of points where f fails to be differentiable and S is any set of measure zero. The next result, invoked in our observation in Remark 5.5, shows that the differentiability hypothesis of Proposition 2.1 cannot be relaxed.

Proposition A.1: (Lipschitz generalized gradient and differentiability): Any locally Lipschitz function with globally Lipschitz generalized gradient is continuously differentiable.

Proof: Let $f : \mathbb{R}^d \rightarrow \mathbb{R}$ be a locally Lipschitz function and has a globally Lipschitz generalized gradient map [19]. Take $x \in \mathbb{R}^d$ and let us show that $\partial f(x)$ is a singleton. Since f is differentiable almost everywhere, there exists a sequence of points $\{x_n\}_{n=1}^{\infty}$, where f is differentiable such that $\lim_{n \rightarrow \infty} x_n = x$. Using the set-valued Lipschitz property of ∂f , we have $\partial f(x) \subset \nabla f(x_n) + K \|x_n - x\| B(0, 1)$, where $K \in \mathbb{R}_{>0}$ is the Lipschitz constant and $B(0, 1)$ is the ball centered at $0 \in \mathbb{R}^d$ of radius one. Hence, any element $v \in \partial f(x)$ can be written as $v = \nabla f(x_n) + K \|x_n - x\| u_n$, where u_n is a unit vector in \mathbb{R}^d . Now, taking the limit, $v = \lim_{n \rightarrow \infty} \nabla f(x_n)$. Hence the generalized gradient is singleton-valued. Differentiability follows now from the set-valued Lipschitz condition. ■