# Distributed continuous-time convex optimization on weight-balanced digraphs

Bahman Gharesifard 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], [7], [8]. Most of these works build on consensus-based dynamics [9], [10], [11], [12] to design discrete-time algorithms that find the solution of the optimization problem. Recent exceptions are the works [13], [14] that deal with continuous-time strategies 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

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

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., [9], [10].

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 (or digraph) 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. Secondly, we study the convergence properties of the saddle-point dynamics when the graph is undirected and provide a complete, original proof of its asymptotic correctness when the original functions are locally Lipschitz (i.e., not necessarily differentiable) and convex. Finally, we consider the optimization problem over directed graphs. We first provide an example of a strongly connected, weight-balanced digraph 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. We show 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. In case the gradients are only locally Lipschitz on compact sets, then the convergence is semiglobal. Our technical approach combines notions and tools from set-valued stability analysis, algebraic graph theory, and convex analysis.

## II. Preliminaries

We start with notational conventions. Let $\mathbb{R}$ and $\mathbb{R}_{\geq 0}$ denote the set of reals and nonnegative reals, respectively. We let $||\cdot||$ denote the Euclidean norm on $\mathbb{R}^d$, $\mathbf{1}_d = (1,\dots,1)^T$, $\mathbf{0}_d = (0,\dots,0)^T \in \mathbb{R}^d$, and $\mathsf{I}_d$ denote the identity matrix in $\mathbb{R}^{d\times d}$. For $A \in \mathbb{R}^{d_1\times d_2}$ and $B \in \mathbb{R}^{e_1\times e_2}$, $A\otimes B$ is the Kronecker product. A function $f : \mathsf{X}_1 \times \mathsf{X}_2 \to \mathbb{R}$, with $\mathsf{X}_1 \subset \mathbb{R}^{d_1}$, $\mathsf{X}_2 \subset \mathbb{R}^{d_2}$ closed and convex, is *concave-convex* if it is concave in its first argument and convex in the second one. A *saddle point* $(x_1^*, x_2^*) \in \mathsf{X}_1 \times \mathsf{X}_2$ of $f$ satisfies $f(x_1, x_2^*) \leq f(x_1^*, x_2^*) \leq f(x_1^*, x_2)$ for all $x_1 \in \mathsf{X}_1$ and $x_2 \in \mathsf{X}_2$. A set-valued map $f : \mathbb{R}^d \rightrightarrows \mathbb{R}^d$ takes elements of $\mathbb{R}^d$ to subsets of $\mathbb{R}^d$.

### A. Graph theory

We review basic notions from graph theory [11]. A *directed graph*, or *digraph*, is a pair $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, where $\mathcal{V}$ is 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 $(u, v) \in \mathcal{E}$. An undirected digraph is a *graph*. A path is an ordered sequence of vertices such that any pair of vertices appearing consecutively is an edge. A digraph is *strongly connected* if there is a path between any pair of distinct vertices. For a graph, this notion is referred to as *connected*. A *weighted digraph* $\mathcal{G} = (\mathcal{V}, \mathcal{E}, \mathsf{A})$ consists of a digraph $(\mathcal{V}, \mathcal{E})$ and an *adjacency matrix* $\mathsf{A} \in \mathbb{R}^{n\times n}_{\geq 0}$ with $a_{ij} > 0$ if $(v_i, v_j) \in \mathcal{E}$ and $a_{ij} = 0$, otherwise. The weighted out-degree and in-degree of $v_i$ are respectively, $d^{\mathrm{w}}_{\mathrm{out}}(v_i) = \sum_{j=1}^{n} a_{ij}$ and $d^{\mathrm{w}}_{\mathrm{in}}(v_i) = \sum_{j=1}^{n} a_{ji}$. The *weighted out-degree matrix* $\mathsf{D}_{\mathrm{out}}$ is diagonal with $(\mathsf{D}_{\mathrm{out}})_{ii} = d^{\mathrm{w}}_{\mathrm{out}}(i)$, for $i \in \{1, \dots, n\}$. The *Laplacian* matrix is $\mathsf{L} = \mathsf{D}_{\mathrm{out}} - \mathsf{A}$. Note that $\mathsf{L}\mathbf{1}_n = 0$. If $\mathcal{G}$ is strongly connected, then zero is a simple eigenvalue of $\mathsf{L}$. $\mathcal{G}$ is undirected if $\mathsf{L} = \mathsf{L}^T$ and *weight-balanced* if $d^{\mathrm{w}}_{\mathrm{out}}(v) = d^{\mathrm{w}}_{\mathrm{in}}(v)$, for all $v \in \mathcal{V}$. Any undirected graph is weight-balanced. The following are equivalent: (i) $\mathcal{G}$ weight-balanced, (ii) $\mathbf{1}_n^T\mathsf{L} = 0$, and (iii) $\mathsf{L} + \mathsf{L}^T$ positive semidefinite, see e.g., [11, Theorem 1.37]. If $\mathcal{G}$ is weight-balanced and strongly connected, then zero is a simple eigenvalue of $\mathsf{L} + \mathsf{L}^T$.

### *B. Nonsmooth analysis*

We recall some notions from nonsmooth analysis [15]. A function $f : \mathbb{R}^d \to \mathbb{R}$ 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 for all $y, z \in \mathbb{R}^d$ there exists $C \in \mathbb{R}_{\geq 0}$ such that $|f(y) - f(z)| \leq C||y - z||$. Locally Lipschitz functions are differentiable almost everywhere. If $\Omega_f$ denotes the set of points where $f$ fails to be differentiable, the *generalized gradient* of $f$ is

$$\partial f(x) = \mathrm{co}\{\lim_{k\to\infty} \nabla f(x_k) \mid x_k \to x, x_k \notin \Omega_f \cup S\},$$

where $S$ is any set of measure zero and co denotes convex hull.

*Lemma 2.1: (Continuity of the generalized gradient map):* Let $f : \mathbb{R}^d \to \mathbb{R}$ be a locally Lipschitz function at $x \in \mathbb{R}^d$. Then the set-valued map $\partial f : \mathbb{R}^d \rightrightarrows \mathbb{R}^d$ is upper semicontinuous and locally bounded at $x \in \mathbb{R}^d$ and moreover, $\partial f(x)$ is nonempty, compact, and convex.

For $f : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ and $z \in \mathbb{R}^d$, we let $\partial_x f(x, z)$ denote the generalized gradient of $x \mapsto f(x, z)$. Similarly, for $x \in \mathbb{R}^d$, we let $\partial_z f(x, z)$ denote the generalized gradient of $z \mapsto f(x, z)$. A *critical point* $x \in \mathbb{R}^d$ of $f$ satisfies $\mathbf{0} \in \partial f(x)$. A function $f : \mathbb{R}^d \to \mathbb{R}$ is *regular* at $x \in \mathbb{R}$ if for all $v \in \mathbb{R}^d$ the right directional derivative of $f$, in the direction of $v$, exists at $x$ and coincides with the generalized directional derivative of $f$ at $x$ in the direction of $v$, see [15] for definitions of these notions. A convex and locally Lipschitz function at $x$ is regular [15, Proposition 2.3.6].

*Lemma 2.2: (Finite sum of locally Lipschitz functions):* Let $\{f^i\}_{i=1}^n$ be locally Lipschitz at $x \in \mathbb{R}^d$. Then $\partial(\sum_{i=1}^n f^i)(x) \subseteq \sum_{i=1}^n \partial f^i(x)$, and equality holds if $f^i$ is regular for $i \in \{1, \dots, n\}$ (here, the summation of sets is the set of points of the form $\sum_{i=1}^n g_i$, with $g_i \in \partial f^i(x)$).

A locally Lipschitz and convex function $f$ satisfies, for all $x, x' \in \mathbb{R}^d$ and $\xi \in \partial f(x)$, the *first-order condition* of convexity,

$$f(x') - f(x) \geq \xi^T (x' - x). \tag{1}$$

The notion of cocoercivity [16] plays a key role in our technical approach later. For $\delta \in \mathbb{R}_{>0}$, a locally Lipschitz function $f$ is $\delta$-*cocoercive* if, for all $x, x' \in \mathbb{R}^d$ and $g_x \in \partial f(x)$, $g_{x'} \in \partial f(x')$,

$$(x - x')^T (g_x - g_{x'}) \geq \delta (g_x - g_{x'})^T (g_x - g_{x'}).$$

The next result [16, Lemma 6.7] characterizes cocoercive differentiable convex functions.

*Proposition 2.3: (Characterization of cocoercivity):* Let $f$ be a differentiable convex function. Then, $\nabla f$ is globally Lipschitz with constant $K \in \mathbb{R}_{>0}$ iff $f$ is $\frac{1}{K}$-cocoercive.

### *C. Set-valued dynamical systems*

Here, we recall some background on set-valued dynamical systems following [17]. A continuous-time set-valued dynamical system on $\mathsf{X} \subset \mathbb{R}^d$ is a differential inclusion

$$\dot{x}(t) \in \Psi(x(t)) \tag{2}$$

where $t \in \mathbb{R}_{\geq 0}$ and $\Psi : \mathsf{X} \subset \mathbb{R}^d \rightrightarrows \mathbb{R}^d$ is a set-valued map. A solution to this dynamical system is an absolutely continuous curve $x : [0, T] \to \mathsf{X}$ which satisfies (2) almost everywhere. The set of equilibria of (2) is denoted by $\mathrm{Eq}(\Psi) = \{x \in \mathsf{X} \mid 0 \in \Psi(x)\}$.

*Lemma 2.4: (Existence of solutions):* For $\Psi : \mathbb{R}^d \rightrightarrows \mathbb{R}^d$ upper semicontinuous with nonempty, compact, and convex values, there exists a solution to (2) from any initial condition.

The LaSalle Invariance Principle is helpful to establish the asymptotic convergence of systems of the form (2). A set $W \subset \mathsf{X}$ is *weakly positively invariant* under (2) if, for each $x \in W$, there exists at least one solution of (2) starting from $x$ entirely contained in $W$. Similarly, $W$ is *strongly positively invariant* under (2) if, for each $x \in W$, all solutions of (2) starting from $x$ are entirely contained in $W$. Finally, the *set-valued Lie derivative* of a differentiable function $V : \mathbb{R}^d \to \mathbb{R}$ with respect to $\Psi$ at $x \in \mathbb{R}^d$ is $\widetilde{\mathcal{L}}_\Psi V(x) = \{v^T \nabla V(x) \mid v \in \Psi(x)\}$.

*Theorem 2.5: (Set-valued LaSalle Invariance Principle):* Let $W \subset \mathsf{X}$ be strongly positively invariant under (2) and $V : \mathsf{X} \to \mathbb{R}$ a continuously differentiable function. Suppose the evolutions of (2) are bounded and $\max \widetilde{\mathcal{L}}_\Psi V(x) \leq 0$ or $\widetilde{\mathcal{L}}_\Psi V(x) = \emptyset$, for all $x \in W$. Let $S_{\Psi,V} = \{x \in \mathsf{X} \mid 0 \in \widetilde{\mathcal{L}}_\Psi V(x)\}$. Then any solution $x(t)$, $t \in \mathbb{R}_{\geq 0}$, starting in $W$ converges to the largest weakly positively invariant set $M$ contained in $\bar{S}_{\Psi,V} \cap W$. When $M$ is a finite collection of points, then the limit of each solution equals one of them.

## III. Problem statement and equivalent formulations

Consider a network composed by $n$ agents $v_1, \dots, v_n$ whose communication topology is described by a strongly connected digraph $\mathcal{G}$. An edge $(v_i, v_j)$ represents the fact that $v_i$ can receive information from $v_j$. For each $i \in \{1, \dots, n\}$, let $f^i : \mathbb{R}^d \to \mathbb{R}$ be locally Lipschitz and convex, and only available to agent $v_i$. The network objective is to solve

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

in a distributed way. Let $x^i \in \mathbb{R}^d$ denote the estimate of agent $v_i$ about the value of the solution to (3) and let $\boldsymbol{x}^T = ((x^1)^T, \dots, (x^n)^T) \in \mathbb{R}^{nd}$. Next, we provide an alternative formulation of (3).

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

$$\text{minimize} \quad \tilde{f}(\boldsymbol{x}) = \sum_{i=1}^{n} f^i(x^i), \qquad \text{subject to} \quad \mathbf{L}\boldsymbol{x} = \mathbf{0}_{nd}. \tag{4}$$

*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, $\mathbf{L}\boldsymbol{x} = \mathbf{0}_{nd}$ if and only if $\boldsymbol{x} = \mathbf{1}_n \otimes x$, 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. Note that $\tilde{f}$ is locally Lipschitz and convex. Moreover, from Lemma 2.2, the elements of its generalized gradient are of the form $\tilde{g}_{\boldsymbol{x}} = (g^1_{x^1}, \dots, g^n_{x^n}) \in \partial\tilde{f}(\boldsymbol{x})$, where $g^i_{x^i} \in \partial f^i(x^i)$, for $i \in \{1, \dots, n\}$. Since $\tilde{f}$ is convex and the constraints in (4) are linear, the constrained optimization problem is feasible [18].

The next result introduces a function which corresponds to the Lagrangian function associated to the constrained optimization problem (4) plus an additional quadratic term that vanishes if the agreement constraint is satisfied. Interestingly, the saddle points of this function correspond to the solutions of the constrained optimization problem, as we show next.

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

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

Then $F$ is locally Lipschitz and convex in its first argument and linear in its second, and

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

*Proof:* First, note that for $\mathcal{G}$ weight-balanced, $\mathbf{L}+\mathbf{L}^T$ is positive semi-definite. Since the sum of convex functions is convex, one deduces that $F$ is convex in its first argument. By inspection, $F$ is linear in its second argument. The statement (i) is immediate. To show (ii), using that $\mathcal{G}$ is strongly connected, one can see that the saddle points of $F$ are of the form $(\boldsymbol{x}^*, \boldsymbol{z}^*)$ with $\boldsymbol{x}^* = \mathbf{1}_n \otimes x^*$, $x^* \in \mathbb{R}^d$, and $\mathbf{L}\boldsymbol{z}^* \in -\partial \tilde{f}(\boldsymbol{x}^*)$. The last inclusion implies that there exist $g^i_{x^*} \in \partial f^i(x^*)$, $i \in \{1, \dots, n\}$, such that $\mathbf{L}\boldsymbol{z}^* = -(g^1_{x^*}, \dots, g^n_{x^*})^T$. Noting that

$$(\mathbf{1}_n^T \otimes \mathsf{I}_d)\mathbf{L} = (\mathbf{1}_n^T \otimes \mathsf{I}_d)(\mathsf{L} \otimes \mathsf{I}_d) = \mathbf{1}_n^T \mathsf{L} \otimes \mathsf{I}_d = \mathbf{0}_{d \times dn},$$

we deduce $\mathbf{0}_d = (\mathbf{1}_n^T \otimes \mathsf{I}_d)\mathbf{L}\boldsymbol{z}^* = -\sum_{i=1}^n g^i_{x^*}$. As a result, using Lemma 2.2, $\boldsymbol{x}^*$ is a solution of (4). Finally, (iii) follows by noting $\boldsymbol{x}^* = \mathbf{1}_n \otimes x^*$ and the fact that $0 \in \partial f(x^*)$ implies that there exists $\boldsymbol{z}^* \in \mathbb{R}^{nd}$ with $\mathbf{L}\boldsymbol{z}^* \in -\partial \tilde{f}(\boldsymbol{x}^*)$, yielding that $(\boldsymbol{x}^*, \boldsymbol{z}^*)$ is a saddle point of $F$. ■

## IV. Continuous-time distributed optimization on undirected networks

Here, we consider the case of 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),

$$\dot{\boldsymbol{x}} + \mathbf{L}\boldsymbol{x} + \mathbf{L}\boldsymbol{z} \in -\partial \tilde{f}(\boldsymbol{x}), \tag{6a}$$

$$\dot{\boldsymbol{z}} = \mathbf{L}\boldsymbol{x}. \tag{6b}$$

Note that (6) is a set-valued dynamical system. Lemmas 2.1 and 2.4 guarantee the existence of solutions. Moreover, from Proposition 3.2, if $(\boldsymbol{x}^*, \boldsymbol{z}^*)$ is an equilibrium of (6), then $\boldsymbol{x}^*$ is a solution to (4). This continuous-time dynamics was originally proposed in [13] (see also [14]), unfortunately without a formal analysis of its convergence properties. Here, we provide a complete, original convergence proof for the case when $f$ is the sum of locally Lipschitz convex functions. The proof also serves to illustrate later the challenges in solving the distributed optimization problem over directed graphs.

*Theorem 4.1: (Asymptotic convergence of* (6) *on graphs):* Let $\mathcal{G}$ be a connected graph and consider the optimization problem (3), where each $f^i$, $i \in \{1, \dots, n\}$ is locally Lipschitz 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).

*Proof:* For convenience, we denote the dynamics (6) by $\Psi_{\text{dis-opt}} : \mathbb{R}^{nd} \times \mathbb{R}^{nd} \rightrightarrows \mathbb{R}^{nd} \times \mathbb{R}^{nd}$. Let $\boldsymbol{x}^* = \mathbf{1}_n \otimes x^*$ be a solution of (4). By Proposition 3.2(iii), there exists $\boldsymbol{z}^*$ such that $(\boldsymbol{x}^*, \boldsymbol{z}^*) \in$

$\mathrm{Eq}(\Psi_{\text{dis-opt}})$. First, note that given any initial condition $(\boldsymbol{x}_0, \boldsymbol{z}_0) \in \mathbb{R}^{nd} \times \mathbb{R}^{nd}$, the set

$$W_{\boldsymbol{z}_0} = \{(\boldsymbol{x}, \boldsymbol{z}) \mid (\mathbf{1}_n^T \otimes \mathsf{I}_d)\boldsymbol{z} = (\mathbf{1}_n^T \otimes \mathsf{I}_d)\boldsymbol{z}_0\} \tag{7}$$

is strongly positively invariant under (6). Consider then the function $V : \mathbb{R}^{nd} \times \mathbb{R}^{nd} \to \mathbb{R}_{\geq 0}$,

$$V(\boldsymbol{x}, \boldsymbol{z}) = \frac{1}{2}(\boldsymbol{x} - \boldsymbol{x}^*)^T(\boldsymbol{x} - \boldsymbol{x}^*) + \frac{1}{2}(\boldsymbol{z} - \boldsymbol{z}^*)^T(\boldsymbol{z} - \boldsymbol{z}^*). \tag{8}$$

The function $V$ is smooth. Let us examine its set-valued Lie derivative. For each $\xi \in \widetilde{\mathcal{L}}_{\Psi_{\text{dis-opt}}} V(\boldsymbol{x}, \boldsymbol{z})$, there exists $v = (-\mathbf{L}\boldsymbol{x} - \mathbf{L}\boldsymbol{z} - \tilde{g}_{\boldsymbol{x}}, \mathbf{L}\boldsymbol{x}) \in \Psi_{\text{dis-opt}}(\boldsymbol{x}, \boldsymbol{z})$, with $\tilde{g}_{\boldsymbol{x}} \in \partial\tilde{f}(\boldsymbol{x})$, such that

$$\xi = v^T \nabla V(\boldsymbol{x}, \boldsymbol{z}) = -(\boldsymbol{x} - \boldsymbol{x}^*)^T(\mathbf{L}\boldsymbol{x} + \mathbf{L}\boldsymbol{z} + \tilde{g}_{\boldsymbol{x}}) + (\boldsymbol{z} - \boldsymbol{z}^*)^T\mathbf{L}\boldsymbol{x}. \tag{9}$$

Since $F$ is convex in its first argument and $\mathbf{L}\boldsymbol{x} + \mathbf{L}\boldsymbol{z} + \tilde{g}_{\boldsymbol{x}} \in \partial_{\boldsymbol{x}} F(\boldsymbol{x}, \boldsymbol{z})$, using the first-order condition of convexity (1), we deduce $(\boldsymbol{x}^* - \boldsymbol{x})^T(\mathbf{L}\boldsymbol{x} + \mathbf{L}\boldsymbol{z} + \tilde{g}_{\boldsymbol{x}}) \leq F(\boldsymbol{x}^*, \boldsymbol{z}) - F(\boldsymbol{x}, \boldsymbol{z})$. On the other hand, the linearity of $F$ in its second argument implies that $(\boldsymbol{z} - \boldsymbol{z}^*)^T\mathbf{L}\boldsymbol{x} = F(\boldsymbol{x}, \boldsymbol{z}) - F(\boldsymbol{x}, \boldsymbol{z}^*)$. Therefore, $\xi \leq F(\boldsymbol{x}^*, \boldsymbol{z}) - F(\boldsymbol{x}^*, \boldsymbol{z}^*) + F(\boldsymbol{x}^*, \boldsymbol{z}^*) - F(\boldsymbol{x}, \boldsymbol{z}^*)$. Since the equilibria of $\Psi_{\text{dis-opt}}$ are the saddle points of $F$, we deduce that $\xi \leq 0$. Since $\xi$ is arbitrary, we conclude $\max \widetilde{\mathcal{L}}_{\Psi_{\text{dis-opt}}} V(\boldsymbol{x}, \boldsymbol{z}) \leq 0$. As a by-product, the trajectories of (6) are bounded. Consequently, all assumptions of the set-valued version of the LaSalle Invariance Principle, cf. Theorem 2.5, are satisfied. This result then implies that any trajectory of (6) starting from an initial condition $(\boldsymbol{x}_0, \boldsymbol{z}_0)$ converges to the largest weakly positively invariant set $M$ in $S_{\Psi_{\text{dis-opt}}, V} \cap W_{\boldsymbol{z}_0}$. Our final step consists of characterizing $M$. Let $(\boldsymbol{x}, \boldsymbol{z}) \in M$. Then $F(\boldsymbol{x}^*, \boldsymbol{z}^*) - F(\boldsymbol{x}, \boldsymbol{z}^*) = 0$, i.e.,

$$\tilde{f}(\boldsymbol{x}^*) - \tilde{f}(\boldsymbol{x}) - (\boldsymbol{z}^*)^T\mathbf{L}\boldsymbol{x} - \frac{1}{2}\boldsymbol{x}^T\mathbf{L}\boldsymbol{x} = 0. \tag{10}$$

Define now $G : \mathbb{R}^{nd} \times \mathbb{R}^{nd} \to \mathbb{R}$ by $G(\boldsymbol{x}, \boldsymbol{z}) = \tilde{f}(\boldsymbol{x}) + \boldsymbol{z}^T\mathbf{L}\boldsymbol{x}$. Note that $G$ is convex in its first argument and linear in its second, and that it has the same saddle points as $F$. As a result, $G(\boldsymbol{x}^*, \boldsymbol{z}^*) - G(\boldsymbol{x}, \boldsymbol{z}^*) \leq 0$, or equivalently, $\tilde{f}(\boldsymbol{x}^*) - \tilde{f}(\boldsymbol{x}) - (\boldsymbol{z}^*)^T\mathbf{L}\boldsymbol{x} \leq 0$. Combining this with (10), we have $\mathbf{L}\boldsymbol{x} = 0$ and $-\tilde{f}(\boldsymbol{x}) + \tilde{f}(\boldsymbol{x}^*) = 0$, i.e., $\boldsymbol{x}$ is solution to (4). Since $M$ is weakly positively invariant, there exists at least a solution of (6) starting from $(\boldsymbol{x}, \boldsymbol{z})$ that remains in $M$. This implies that, along the solution, the components of $\boldsymbol{x}$ remain in agreement, i.e., $\boldsymbol{x}(t) = \mathbf{1}_n \otimes a(t)$ with $a(t) \in \mathbb{R}^d$ a solution of (3). Applying $\mathbf{1}_n^T \otimes \mathsf{I}_d$ on both sides of $\mathbf{1}_n \otimes \dot{a}(t) + \mathbf{L}\boldsymbol{z} \in -\partial\tilde{f}(\boldsymbol{x}(t))$, we deduce $n\dot{a}(t) \in -\sum_{i=1}^n \partial f^i(a(t))$. Lemma A.2 then implies that $\dot{a}(t) = 0$, i.e., $\mathbf{L}\boldsymbol{z} \in -\partial\tilde{f}(\boldsymbol{x})$ and thus $(\boldsymbol{x}, \boldsymbol{z}) \in \mathrm{Eq}(\Psi_{\text{dis-opt}})$. Finally, if the set of equilibria is finite, the last statement holds true. ∎

*Remark 4.2: (Asymptotic convergence of saddle-point dynamics):* The work [19] studies saddle-point dynamics and guarantees asymptotic convergence to a saddle point when the function's

Hessian in one argument is positive definite and the function is linear in the other. Such result, however, cannot be applied to establish Theorem 4.1 because the generality of the hypotheses on $f$ mean that $F$ might not satisfy these conditions. Instead, our proof shows that a careful study of the invariance properties of the flow yields the desired result. •

## V. Continuous-time distributed optimization on directed networks

Here, we consider the optimization problem (3) on digraphs. When $\mathcal{G}$ is directed, the gradient of $F$ defined in (5) is no longer distributed over $\mathcal{G}$ because it contains terms that involve $\mathbf{L}^T$ and hence requires agents to receive information from its in-neighbors. In fact, the dynamics (6), which is distributed over $\mathcal{G}$, does no longer correspond to the saddle-point dynamics of $F$. Nevertheless, it is natural to study whether (6) enjoys the same convergence properties as in the undirected setting (as, for instance, is the case in the agreement problem [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 weight-balanced 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 $\lambda$ of the Laplacian $\mathsf{L}$, one has $\sqrt{3}|\mathrm{Im}(\lambda)| \leq \mathrm{Re}(\lambda)$.

*Proof:* By assumption, the dynamics (6) is linear with matrix $\left(\begin{smallmatrix} -1 & -1 \\ 1 & 0 \end{smallmatrix}\right) \otimes \mathbf{L}$ and has $\mathcal{S}_{\text{agree}}$ as equilibria. The eigenvalues of the matrix are of the form $\lambda\left(\frac{-1}{2} \pm \frac{\sqrt{3}}{2} i\right)$, with $\lambda$ an eigenvalue of $\mathbf{L}$ (because the eigenvalues of a Kronecker product are just the product of the eigenvalues of the corresponding matrices). Since $\mathbf{L} = \mathsf{L} \otimes \mathsf{I}_d$, each eigenvalue of $\mathbf{L}$ is an eigenvalue of $\mathsf{L}$. Finally, $\mathrm{Re}\big(\lambda\big(\frac{-1}{2} \pm \frac{\sqrt{3}}{2} i\big)\big) = \frac{1}{2}(\mp\sqrt{3}\mathrm{Im}(\lambda) - \mathrm{Re}(\lambda))$, from which the result follows. ■

It is not difficult to construct examples of convex functions that have zero contribution to the linearization of (6) around the solution. Therefore, such systems cannot be convergent if they fail the criterium identified in Lemma 5.1. The next example shows that this criterium 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 $\sqrt{3}|\mathrm{Im}(\lambda)| - \mathrm{Re}(\lambda) = 0.0171 > 0$, Lemma 5.1 implies that (6) fails to converge. •

### *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. For reasons that will be made clear later in Remark 5.5, we restrict our attention to the case when the functions $f^i$, $i \in \{1, \dots, n\}$ are continuously differentiable. Let $\alpha \in \mathbb{R}_{>0}$ and consider the dynamics

$$\dot{\boldsymbol{x}} + \alpha \mathbf{L}\boldsymbol{x} + \mathbf{L}\boldsymbol{z} = -\nabla \tilde{f}(\boldsymbol{x}), \tag{11a}$$

$$\dot{\boldsymbol{z}} = \mathbf{L}\boldsymbol{x}. \tag{11b}$$

The existence of solutions is guaranteed by Lemmas 2.1 and 2.4. We first show that appropriate choices of $\alpha$ allow to circumvent the problem raised in Lemma 5.1.

*Lemma 5.3: (Sufficient conditions for the convergence of* (11) *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 (11).

*Proof:* When all $f_i$, $i \in \{1, \dots, n\}$, are identically zero, the dynamics (11) is linear and has $\mathcal{S}_{\text{agree}}$ as equilibria. Consider the coordinate transformation from $(\boldsymbol{x}, \boldsymbol{z})$ to $(\boldsymbol{x}, \boldsymbol{y}) = (\boldsymbol{x}, \beta\boldsymbol{x} + \boldsymbol{z})$, with $\beta \in \mathbb{R}_{>0}$ to be chosen later. The dynamics can be rewritten as

$$\begin{pmatrix} \dot{\boldsymbol{x}} \\ \dot{\boldsymbol{y}} \end{pmatrix} = A \begin{pmatrix} \boldsymbol{x} \\ \boldsymbol{y} \end{pmatrix}, \quad \text{where} \quad A = \begin{pmatrix} -(\alpha - \beta)\mathbf{L} & -\mathbf{L} \\ (-\beta(\alpha-\beta)+1)\mathbf{L} & -\beta\mathbf{L} \end{pmatrix}. \tag{12}$$

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

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

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

$$Q = \begin{pmatrix} -(\frac{\beta^2+2}{\beta} - \beta) & -1 \\ -1 & -\beta \end{pmatrix} \otimes (\mathbf{L} + \mathbf{L}^T). \tag{13}$$

Each eigenvalue $\eta$ of $Q$ is of the form $\eta = \lambda \frac{-(\beta^2+2) \pm \sqrt{(\beta^2+2)^2 - 4\beta^2}}{2\beta}$, where $\lambda$ is an eigenvalue of $\mathsf{L} + \mathsf{L}^T$. 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 (11) from any initial condition $(\boldsymbol{x}_0, \boldsymbol{y}_0) \in \mathbb{R}^{nd} \times \mathbb{R}^{nd}$, asymptotically converge to the set $S = \{(\boldsymbol{x}, \boldsymbol{y}) \mid Q(\boldsymbol{x}, \boldsymbol{y})^T = \mathbf{0}_{2nd}\} \cap W_{\boldsymbol{z}_0}$. To conclude the result, we need to show that $S \subseteq \mathcal{S}_{\text{agree}}$. This follows from noting that, for $\beta > 0$, $Q(\boldsymbol{x}, \boldsymbol{y})^T = \mathbf{0}_{2nd}$ implies that $(\mathbf{L} + \mathbf{L}^T)\boldsymbol{x} = \mathbf{0}_{nd}$ and $(\mathbf{L} + \mathbf{L}^T)\boldsymbol{y} = \mathbf{0}_{nd}$, i.e., $(\boldsymbol{x}, \boldsymbol{y}) \in \mathcal{S}_{\text{agree}}$. ■

The reason behind the introduction of the parameter $\alpha$ in (11) 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 $(\boldsymbol{x} - \boldsymbol{x}^*)^T(\mathbf{L} - \mathbf{L}^T)(\boldsymbol{z} - \boldsymbol{z}^*)$ 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 $\alpha$, 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 (11) solves in a distributed way the optimization problem (3) on strongly connected weight-balanced digraphs.

*Theorem 5.4: (Asymptotic convergence of* (11) *on weight-balanced digraphs):* Let $\mathcal{G}$ be a strongly connected, weight-balanced digraph and consider the optimization problem (3), where each $f^i$, $i \in \{1, \ldots, n\}$, is convex and differentiable with globally Lipschitz continuous gradient. Let $K \in \mathbb{R}_{>0}$ be the Lipschitz constant of $\nabla \tilde{f}$ and define $h : \mathbb{R}_{>0} \to \mathbb{R}$ by

$$h(r) = \frac{1}{2}\Lambda_*(\mathsf{L} + \mathsf{L}^T)\left(-\frac{r^4 + 3r^2 + 2}{r} + \sqrt{\left(\frac{r^4 + 3r^2 + 2}{r}\right)^2 - 4}\right) + \frac{Kr^2}{(1 + r^2)}, \tag{14}$$

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 (11) 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 (11) by $\Psi_{\alpha\text{-dis-opt}} : \mathbb{R}^{nd} \times \mathbb{R}^{nd} \to \mathbb{R}^{nd} \times \mathbb{R}^{nd}$. Note that the equilibria of $\Psi_{\alpha\text{-dis-opt}}$ are precisely the set of saddle points of $F$ in (5). Let

$\boldsymbol{x}^* = \mathbf{1}_n \otimes x^*$ be a solution of (4). First, note that given any initial condition $(\boldsymbol{x}_0, \boldsymbol{z}_0) \in \mathbb{R}^{nd} \times \mathbb{R}^{nd}$, the set $W_{\boldsymbol{z}_0}$ defined by (7) is invariant under the evolutions of (11). By Proposition 3.2(i) and (iii), there exists $(\boldsymbol{x}^*, \boldsymbol{z}^*) \in \mathrm{Eq}(\Psi_{\alpha\text{-dis-opt}}) \cap W_{\boldsymbol{z}_0}$. Consider the function $V : \mathbb{R}^{nd} \times \mathbb{R}^{nd} \to \mathbb{R}_{\geq 0}$,

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

where $\boldsymbol{y}_{(\boldsymbol{x},\boldsymbol{z})} = \beta\boldsymbol{x} + \boldsymbol{z}$ 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 $W_{\boldsymbol{z}_0}$. For $(\boldsymbol{x}, \boldsymbol{z}) \in W_{\boldsymbol{z}_0}$, let

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

where $A$ is given by (12). This equation can be written as

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

where $Q$ is given by (13). Note that $A(\boldsymbol{x}^*, \boldsymbol{y}_{(\boldsymbol{x}^*,\boldsymbol{z}^*)})^T = -(\mathbf{L}\boldsymbol{y}_{(\boldsymbol{x}^*,\boldsymbol{z}^*)}, \beta\mathbf{L}\boldsymbol{y}_{(\boldsymbol{x}^*,\boldsymbol{z}^*)})^T = (\nabla\tilde{f}(\boldsymbol{x}^*), \beta\nabla\tilde{f}(\boldsymbol{x}^*))^T$. Thus, after substituting for $\boldsymbol{y}_{(\boldsymbol{x},\boldsymbol{z})}$, we have

$$\begin{aligned}
\xi =& \frac{1}{2}\Big((\boldsymbol{x} - \boldsymbol{x}^*)^T, (\boldsymbol{z} - \boldsymbol{z}^*)^T\Big)^T \tilde{Q} \Big(\boldsymbol{x} - \boldsymbol{x}^*, \boldsymbol{z} - \boldsymbol{z}^*\Big)^T \\
&- (1 + \beta^2)(\boldsymbol{x} - \boldsymbol{x}^*)^T(\nabla\tilde{f}(\boldsymbol{x}) - \nabla\tilde{f}(\boldsymbol{x}^*)) - \beta(\boldsymbol{z} - \boldsymbol{z}^*)^T(\nabla\tilde{f}(\boldsymbol{x}) - \nabla\tilde{f}(\boldsymbol{x}^*)),
\end{aligned} \tag{15}$$

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

$$\tilde{\eta} = \lambda \times \frac{-(\beta^4 + 3\beta^2 + 2) \pm \sqrt{(\beta^4 + 3\beta^2 + 2)^2 - 4\beta^2}}{2\beta}, \tag{16}$$

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

$$\xi \leq \frac{1}{2}\begin{pmatrix} \boldsymbol{x} - \boldsymbol{x}^* \\ \boldsymbol{z} - \boldsymbol{z}^* \\ \nabla\tilde{f}(\boldsymbol{x}) - \nabla\tilde{f}(\boldsymbol{x}^*) \end{pmatrix}^T \underbrace{\begin{pmatrix} \tilde{Q}_{11} & \tilde{Q}_{12} & 0 \\ \tilde{Q}_{21} & \tilde{Q}_{22} & -\beta\mathsf{I}_{nd} \\ 0 & -\beta\mathsf{I}_{nd} & -\frac{1}{K}(1+\beta^2)\mathsf{I}_{nd} \end{pmatrix}}_{\mathbf{Q}} \begin{pmatrix} \boldsymbol{x} - \boldsymbol{x}^* \\ \boldsymbol{z} - \boldsymbol{z}^* \\ \nabla\tilde{f}(\boldsymbol{x}) - \nabla\tilde{f}(\boldsymbol{x}^*) \end{pmatrix}, \tag{17}$$

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

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

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

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 $\begin{pmatrix} \bar{Q} & 0 \\ 0 & -\frac{1}{K}(1+\beta^2)\mathsf{I}_{nd} \end{pmatrix}$ is negative semidefinite on $\mathcal{W}$. Clearly, $-\frac{1}{K}(1+\beta^2)\mathsf{I}_{nd}$ 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 $\{(v_1, v_2) \in (\mathbb{R}^{nd})^2 \mid (\mathbf{1}_n^T \otimes \mathsf{I}_d) v_2 = \mathbf{0}_d\}$, $0$ 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 $\{(v_1, v_2) \in (\mathbb{R}^{nd})^2 \mid (\mathbf{1}_n^T \otimes \mathsf{I}_d) v_2 = \mathbf{0}_d\}$, the eigenvalues of $\frac{K\beta^2}{(1+\beta^2)} \begin{pmatrix} 0 & 0 \\ 0 & \mathsf{I}_{nd} \end{pmatrix}$ are $\frac{K\beta^2}{(1+\beta^2)}$ with multiplicity $nd - d$ and $0$ with multiplicity $nd$. Therefore, using Weyl's theorem [20, Theorem 4.3.7], we deduce that the nonzero eigenvalues of the sum $\bar{Q}$ are upper bounded by $\Lambda_*(\tilde{Q}) + \frac{K\beta^2}{(1+\beta^2)}$. From (16) and the definition of $h$ in (14), 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 $h(r) = -\frac{1}{2}\Lambda_*(\mathsf{L} + \mathsf{L}^T) r + O(r^2)$. Furthermore, $\lim_{r \to \infty} h(r) = K > 0$. Hence, the existence of $\beta^*$ follows from the Mean Value Theorem. Therefore we conclude $\mathcal{L}_{\Psi_{\alpha\text{-dis-opt}}} V(\boldsymbol{x}, \boldsymbol{z}) \le 0$. As a by-product, the trajectories of (11) are bounded. Consequently, all assumptions of the LaSalle Invariance Principle are satisfied and its application yields that any trajectory of (11) starting from an initial condition $(\boldsymbol{x}_0, \boldsymbol{z}_0)$ converges to the largest positively invariant set $M$ in $S_{\Psi_{\alpha\text{-dis-opt}}, V} \cap W_{\boldsymbol{z}_0}$. Note that if $(\boldsymbol{x}, \boldsymbol{z}) \in S_{\Psi_{\alpha\text{-dis-opt}}, V} \cap W_{\boldsymbol{z}_0}$, then $N^T \begin{pmatrix} \boldsymbol{x} - \boldsymbol{x}^* \\ \boldsymbol{z} - \boldsymbol{z}^* \\ \nabla \tilde{f}(\boldsymbol{x}) - \nabla \tilde{f}(\boldsymbol{x}^*) \end{pmatrix} \in \ker(\bar{Q}) \times \{0\}$. 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 $\boldsymbol{x} = \boldsymbol{x}^* + \mathbf{1}_n \otimes a$, $\boldsymbol{z} = \boldsymbol{z}^*$, and $\nabla \tilde{f}(\boldsymbol{x}) = \nabla \tilde{f}(\boldsymbol{x}^*)$, from where we deduce that $\boldsymbol{x}$ is also a solution to (4). Finally, for $(\boldsymbol{x}, \boldsymbol{z}) \in M$, an argument similar to the one in the proof of Theorem 4.1 establishes $(\boldsymbol{x}, \boldsymbol{z}) \in \mathrm{Eq}(\Psi_{\alpha\text{-dis-opt}})$. If the set of equilibria is finite, convergence to a point is also guaranteed. ■

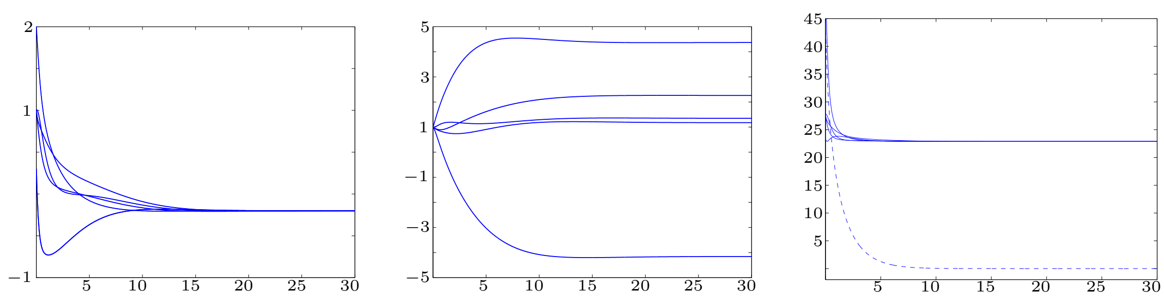
*Remark 5.5 (Locally Lipschitz objective functions):* Our simulations suggests that the convergence result in Theorem 5.4 holds true for locally Lipschitz objective functions. However, our proof cannot be reproduced for this case because it would rely on the generalized gradient being globally Lipschitz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

*Remark 5.6 (Locally Lipschitz gradients):* From the proof of Theorem 5.4, one can observe that if the requirement on the Lipschitzness of the gradient of each $f^i$, $i \in \{1, \dots, n\}$ is relaxed from globally to locally on compact sets, then the convergence result is semiglobal. In other words, given an arbitrary compact set $C \subset \mathbb{R}^{nd} \times \mathbb{R}^{nd}$, one can always find $\beta^*$ (that would depend on $C$) such that, for $0 < \beta < \beta^*$, $C$ belongs to the region of attraction of the set of saddle points for the dynamics (11). This relaxed requirement is satisfied, for instance, by any twice continuously differentiable function. •

*Remark 5.7 (Selection of $\alpha$ in (11)):* According to Theorem 5.4, the parameter $\alpha$ is determined by $\beta$ as $\alpha = \frac{\beta^2+2}{\beta}$. In turn, one can observe from (14) that the range of suitable values for $\beta$ 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 $\beta$ that yields the smallest $\alpha$ while satisfying the conditions of the theorem statement. •

*Remark 5.8 (Discrete-time counterpart of (6) and (11)):* It is worth noticing that the discretization of (6) for undirected graphs (performed in [13] for the case of continuously differentiable, strictly convex functions) and (11) for weight-balanced digraphs gives rise to different discrete-time optimization algorithms from the ones considered in [1], [2], [3], [4], [5], [6]. •

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



(a) (b) (c)

Fig. 1. Execution of (11) for the network of Example 5.2 to optimize $f(x) = \sum_{i=1}^{5} f^i(x)$, where $f^1(x) = e^x$, $f^2(x) = (x-3)^2$, $f^3(x) = (x+3)^2$, $f^4(x) = x^4$, $f^5(x) = 4$. Note that these functions are all smooth and convex, with locally Lipschitz gradients on compact sets, cf. Remark 5.6. (a) and (b) show the evolution of the agent's values in $x$ and $z$, respectively. (c) shows the value of $f$ along the agents' estimates of the optimizer plotted in (a) (solid lines), eventually converging towards the same optimal value, and the value of the Lyapunov function (dashed line). The initial condition is $\boldsymbol{x}_0 = (1, 2, 0.3, 1, 1)^T$, $\boldsymbol{z}_0 = \mathbf{1}_5$, and the parameter is $\alpha = 3$. The asymptotic agreement value is the equilibrium $(\boldsymbol{x}^*, \boldsymbol{z}^*)$, with $\boldsymbol{x}^* = -0.2005 \cdot \mathbf{1}_5$ and $\boldsymbol{z}^* = (1.1784, 4.3717, -4.1598, 2.2598, 1.3499)^T$. Note that $-0.2005$ is the global optimizer of $f$, with $f(-0.2005) = 22.9003$.

## 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 result for undirected networks does 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. Our technical approach relies on a careful combination of stability analysis, algebraic graph theory, and convex analysis. Future work will focus on the extension of the convergence results to locally Lipschitz functions in the weight-balanced directed case and to general digraphs, the incorporation of local and global constraints, the design of distributed algorithms that allow the network to agree on an optimal value of the design parameter, the discretization of the algorithms, and the study of the connection with dynamic consensus strategies.

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

[8] M. Burger, G. Notarstefano, F. Bullo, and F. Allgower, "A distributed simplex algorithm for degenerate linear programs and mulit-agent assignment," *Automatica*, vol. 48, no. 9, pp. 2298–2304, 2012.

[9] 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.

[10] W. Ren and R. W. Beard, *Distributed Consensus in Multi-vehicle Cooperative Control*. Communications and Control Engineering, Springer, 2008.

[11] 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.

[12] M. Mesbahi and M. Egerstedt, *Graph Theoretic Methods in Multiagent Networks*. Applied Mathematics Series, Princeton University Press, 2010.

[13] 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.

[14] 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.

[15] F. H. Clarke, *Optimization and Nonsmooth Analysis*. Canadian Mathematical Society Series of Monographs and Advanced Texts, Wiley, 1983.

[16] E. G. Golshtein and N. V. Tretyakov, *Modified Lagrangians and Monotone Maps in Optimization*. New York: Wiley, 1996.

[17] 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.

[18] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge University Press, 2004.

[19] K. Arrow, L. Hurwitz, and H. Uzawa, *Studies in Linear and Non-Linear Programming*. Stanford, California: Stanford University Press, 1958.

[20] R. A. Horn and C. R. Johnson, *Matrix Analysis*. Cambridge University Press, 1985.

[21] A. Bacciotti and F. Ceragioli, "Stability and stabilization of discontinuous systems and nonsmooth Lyapunov functions," *ESAIM: Control, Optimisation & Calculus of Variations*, vol. 4, pp. 361–376, 1999.

## APPENDIX

The next result shows that the differentiability hypothesis of Proposition 2.3 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 $f : \mathbb{R}^d \to \mathbb{R}$ be locally Lipschitz with a globally Lipschitz generalized gradient map [17]. 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 $\{x_n\}_{n=1}^{\infty}$ where $f$ is differentiable with $\lim_{n\to\infty} x_n = x$. Using the set-valued Lipschitz property of $\partial f$, 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 $1$. Hence, any element $v \in \partial f(x)$ can be written as $v = \nabla f(x_n) + K\|x_n - x\| u_n$, with $u_n \in \mathbb{R}^d$, $\|u_n\| = 1$. Now, taking the limit, $v = \lim_{n\to\infty} \nabla f(x_n)$. Hence the generalized gradient is singleton-valued. Continuous differentiability now follows from the set-valued Lipschitz condition. ■

*Lemma A.2 (Generalized gradient flow from a critical point):* Let $f : \mathbb{R}^d \to \mathbb{R}$ be locally Lipschitz and convex, and let $x^*$ be a minimizer of $f$. Then, the only solution of $\dot{x}(t) \in -\partial f(x(t))$ starting from $x^*$ is $x(t) = x^*$, for all $t \geq 0$.

*Proof:* We reason by contradiction. Assume $x(t)$ is not identically $x^*$. Since $f$ is monotonically nonincreasing along the gradient flow, the trajectory must stay in the set of minimizers of $f$, and hence $t \mapsto f(x(t))$ is constant. Let $t'$ be the smallest time such that $-\partial f(x^*) \ni v = \dot{x}(t') \neq 0$. Using [21, Lemma 1], we have $0 = \frac{d}{dt} f(x(t)) = v^T \xi$, for all $\xi \in \partial f(x^*)$. In particular, for $\xi = -v$, we get $0 = -\|v\|_2^2$, which is a contradiction. ■