This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TAC.2021.3071651, IEEE Transactions on Automatic Control

Distributed Adaptive Optimization with Weight-Balancing

Dongdong Yue, Simone Baldi, *Senior Member, IEEE,* Jinde Cao, *Fellow, IEEE,* and Bart De Schutter, *Fellow, IEEE*

***Abstract*—This paper addresses the continuous-time distributed optimization of a strictly convex summation-separable cost function with possibly non-convex local functions over strongly connected digraphs. Distributed optimization methods in the literature require convexity of local functions, or balanced weights, or vanishing step sizes, or algebraic information (eigenvalues or eigenvectors) of the Laplacian matrix. The solution proposed here covers both weight-balanced and unbalanced digraphs in a unified way, without any of the aforementioned requirements.**

***Index Terms*—Distributed optimization, weight balancing, directed graphs, multi-agent systems.**

I. Introduction

Distributed optimization stands for the strategy of solving an optimization problem cooperatively, using a network of agents. This problem has attracted a lot of interest in recent years due to its possible application in several domains, spanning from autonomous vehicles to smart grids and distributed computing [1]–[11]. In its standard (unconstrained) formulation, distributed optimization involves the problem of solving

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

where F is a global summation-separable cost function, and f_i are the local cost functions, one for each agent.

In order to solve the distributed optimization problem (1), different assumptions can be made regarding the topology with which the agents in the network communicate with each other: e.g., undirected and connected graphs [1]–[3], [10], [11] or strongly connected weight-balanced (i.e., weighted in-degree equals to weighted out-degree for each node) digraphs [4], [5], or unbalanced digraphs [6]–[9]. In the weight-balanced case of [4], [5], the algorithms require the selection of a parameter based on the eigenvalues of the global Laplacian matrix. The algorithms in [6], [7], [9] also rely on algebraic information (e.g., eigenvalues or eigenvectors) of the Laplacian matrix. From a *distributed* optimization perspective, removing the knowledge of the Laplacian matrix is a crucial aspect for advanced concurrency and enhanced flexibility. A modified saddle-point algorithm proposed in [8] does not require such knowledge, but it requires vanishing step sizes, possibly leading to slow convergence. All aforementioned algorithms except [9] require convexity of the local functions; [9] allows local non-convexity at the price of requiring global strong convexity and monotonously increasing coupling gains in the whole network, which can possibly lead to slow convergence with strong oscillations (cf. our simulations in Sect. V).

The idea of balancing an originally unbalanced digraph in a distributed way, as proposed in the discrete-time algorithms [12]–[15], provides an interesting perspective for rethinking distributed optimization over unbalanced digraphs, e.g. by achieving balancing in finite time before starting the distributed optimization algorithm. Nevertheless, finite-time convergence can only be guaranteed for the case of weight-balancing with integer weights in [13]–[15]. Balancing a digraph in finite time with real weights is still an open issue.

We focus on continuous-time algorithms, and develop a novel adaptive framework for solving problem (1) over strongly connected (balanced or unbalanced) digraphs without requiring convexity of local functions, nor vanishing step sizes, nor algebraic information of the Laplacian matrix. For weight-balanced digraphs, we introduce a saddle-point-like[1] algorithm with adaptive gains designed along a directed-spanning-tree (DST) of the graph. The adaptive gains are designed to promote the consensus over the estimates of the global minimizer of (1). In this sense, the proposed methodology is inspired by the DST adaptive control approach of [16]–[18], although the presence of the saddle-point-like dynamics used for optimization requires a completely different design and stability analysis. For weight-unbalanced digraphs, we propose a novel distributed continuous-time weight-balancing algorithm to obtain a balanced digraph in finite time with real weights, after which the previous saddle-point-like algorithm can start. This balancing idea departs from those in [12]–[15] in the sense that instead of adjusting all the weights to balance a digraph, we show that the digraph can be balanced along an aforementioned DST, which makes the proposed framework consistent for both balanced and unbalanced cases.

This work was supported in part by the China Scholarship Council under Grant 201906090134, in part by the Double Innovation Plan under Grant 4207012004, in part by the National Key Research and Development Project of China under Grant 2020YFA0714301, and in part by the Natural Science Foundation of China under Grant 62073074 and Grant 61833005.
(Corresponding authors: Simone Baldi and Jinde Cao).

D. Yue is with School of Automation, Southeast University, Nanjing, China (e-mail: yueseu@gmail.com).

S. Baldi is with School of Mathematics, Southeast University, Nanjing, China, and with Delft Center for Systems and Control, Delft University of Technology, Delft, The Netherlands (e-mail: S.Baldi@tudelft.nl).

J. Cao is with School of Mathematics, Southeast University, Nanjing, China, and with Yonsei Frontier Lab, Yonsei University, Seoul, South Korea. (e-mail: jdcao@seu.edu.cn)

B. De Schutter is with Delft Center for Systems and Control, Delft University of Technology, Delft, The Netherlands (e-mail: B.DeSchutter@tudelft.nl).

[1]We use “saddle-point-like” because the resulting system may not be a true saddle-point dynamical system due to the possibly non-convex local functions.

II. Preliminaries and Problem Setup

A. Notations

In this paper, $\mathbb{R}$ with appropriate dimensions represent the real spaces and $\mathbb{R}^+$ is the real positive scalar subspace. Let $\mathbf{I}_n$ and $\mathbf{1}_n$ be the $n\times n$ identity matrix, and the column vector with n elements being one, respectively. Zero vectors and zero matrices are all denoted by 0. Let $\|x\|$ denote the Euclidean norm of a vector x. For a real matrix A, denote $A^{\mathrm{s}} = (A + A^T)/2$: when A represents the adjacency matrix of a digraph, A^{s} is its undirected version. If A is symmetric, $\lambda_{\max}(A)$ (resp. $\lambda_{\min}(A)$) is its maximum (resp. minimum) eigenvalue, and $A > 0$ (resp. $A \geq 0$) means that A is positive definite (resp. semi-definite). Denote $\mathrm{col}(x_1, \cdots, x_N) = ({x_1}^T, \cdots, {x_N}^T)^T$ as the column vectorization. The abbreviation $\mathrm{diag}(\cdot)$ denotes the diagonalization operator and $\otimes$ stands for the Kronecker product. The difference of the sets $\mathcal{S}_1$ and $\mathcal{S}_2$ is denoted by $\mathcal{S}_1 \setminus \mathcal{S}_2$. Denote $\mathcal{I}_N = \{1, 2, \cdots, N\}$ as the set of natural numbers up to N. For $x, q \in \mathbb{R}$, $\mathrm{sgn}(x)$ is the sign function and $\mathrm{sig}^q(x) = \mathrm{sgn}(x)|x|^q$, where $|x|$ is the absolute value of x. For a differentiable function $g : \mathbb{R}^n \to \mathbb{R}$, ∇g is its gradient; g is *strictly convex* over a convex set $\Omega \subseteq \mathbb{R}^n$ if $(x - y)^T(\nabla g(x) - \nabla g(y)) > 0$, $\forall x, y \in \Omega$ with $x \neq y$.

B. Graph Theory and Technical Lemmas

A weighted *directed graph* [19] (or simply *digraph*) $\mathcal{G}(\mathcal{V}, \mathcal{E}, \mathcal{W})$ is specified by the node set $\mathcal{V} = \mathcal{I}_N$, the edge set $\mathcal{E} = \{e_{ij}, i \neq j | i \to j\}$, and the weighted adjacency matrix $\mathcal{W} = (w_{ij}) \in \mathbb{R}^{N\times N}$, representing the coupling strengths among nodes, such that $w_{ij} > 0$ if $e_{ji} \in \mathcal{E}$, and $w_{ij} = 0$ otherwise. If $e_{ij} \in \mathcal{E}$, i is called an in-neighbor of j, denoted by $i \in \mathcal{N}_{\mathrm{in}}(j)$ (resp. j is an out-neighbor of i: $j \in \mathcal{N}_{\mathrm{out}}(i)$). The Laplacian matrix $\mathcal{L} = (\mathcal{L}_{ij}) \in \mathbb{R}^{N\times N}$ of $\mathcal{G}$ is defined as follows: $\mathcal{L}_{ij} = -w_{ij}, i \neq j$, and $\mathcal{L}_{ii} = \sum_{k=1,k\neq i}^{N} w_{ik}, i = 1, \cdots, N$. A path is a sequence of edges connecting a pair of nodes. A digraph $\mathcal{G}$ is *strongly connected* if any pair of nodes is connected by a directed path, and is *weakly connected* if any pair of nodes is connected by a path disregarding the directions of the edges. Moreover, $\mathcal{G}$ is *weight-balanced* or *balanced* if $\sum_{j\in\mathcal{N}_{\mathrm{in}}(i)} w_{ij} = \sum_{j\in\mathcal{N}_{\mathrm{out}}(i)} w_{ji}$, $\forall i \in \mathcal{V}$.

A directed-spanning-tree (DST) $\bar{\mathcal{G}}(\mathcal{V}, \bar{\mathcal{E}}, \bar{\mathcal{W}})$ of $\mathcal{G}$ is a subgraph where there is a node called the root, which has no in-neighbors, such that one can find a unique directed path from the root to every other node. In $\bar{\mathcal{G}}$, each node has a unique in-neighbor, except for the root. Moreover, a node is called a stem if it has at least one out-neighbor, and a leaf otherwise; the root is called a hub if all its out-neighbors are leaves. Without loss of generality, we suppose that the root is node 1. Following the notations in [16]–[18], let i_k denote the unique in-neighbor of node $k+1$ in $\bar{\mathcal{G}}$, $k \in \mathcal{I}_{N-1}$. Then, $\bar{\mathcal{E}} = \{e_{i_k,k+1} | k \in \mathcal{I}_{N-1}\} \subseteq \mathcal{E}$. Correspondingly, $\bar{\mathcal{L}}$ (resp. $\bar{\mathcal{W}}$) is the Laplacian (resp. weighted adjacency) matrix of $\bar{\mathcal{G}}$ and $\bar{\mathcal{N}}_{\mathrm{out}}(i)$ is the set of out-neighbors of i in $\bar{\mathcal{G}}$.

Lemma 1: Suppose $\mathcal{G}$ contains a DST $\bar{\mathcal{G}}$. Let $\tilde{\mathcal{L}} = \mathcal{L} - \bar{\mathcal{L}}$. Define $\Xi \in \mathbb{R}^{(N-1)\times N}$ as

$$\Xi_{kj} = \begin{cases} -1, & \text{if } j = k+1, \\ 1, & \text{if } j = i_k, \\ 0, & \text{otherwise.} \end{cases} \tag{2}$$

Define $Q \in \mathbb{R}^{(N-1)\times(N-1)} := \tilde{Q} + \bar{Q}$ with

$$Q_{kj} = \underbrace{\sum_{c\in\bar{\mathcal{V}}_{j+1}} (\tilde{\mathcal{L}}_{k+1,c} - \tilde{\mathcal{L}}_{i_k,c})}_{\tilde{Q}_{kj}} + \underbrace{\sum_{c\in\bar{\mathcal{V}}_{j+1}} (\bar{\mathcal{L}}_{k+1,c} - \bar{\mathcal{L}}_{i_k,c})}_{\bar{Q}_{kj}}, \tag{3}$$

where $\bar{\mathcal{V}}_{j+1}$ represents the vertex set of the subtree of $\bar{\mathcal{G}}$ rooting at node $j+1$. Then, the following statements hold:

1) $\mathcal{L}$ has a simple zero eigenvalue corresponding to the right eigenvector $\mathbf{1}_N$, and the other eigenvalues have positive real parts.
2) $\Xi\mathcal{L} = Q\Xi$.
3) $\bar{Q}$ can be explicitly written as

$$\bar{Q}_{kj} = \begin{cases} \bar{w}_{j+1,i_j}, & \text{if } j = k, \\ -\bar{w}_{j+1,i_j}, & \text{if } j = i_k - 1, \\ 0, & \text{otherwise.} \end{cases}$$

4) The eigenvalues of Q are exactly the nonzero eigenvalues of $\mathcal{L}$.

Proof. The proof of statement 1) can be found in [19, Lemma 2.4]; statement 1) is sufficient and necessary for the existence of $\bar{\mathcal{G}}$. The proofs of statements 2) and 3) can be found in [18]. Statement 4) is a direct application of [20, Lemma 10]. ■

Remark 1: Matrix Ξ^T is the incidence matrix associated to $\bar{\mathcal{G}}$. Intuitively, statement 2) reveals that the information of the Laplacian $\mathcal{L}$ can be encoded into the reduced-order matrix Q through a commutative-like relation.

Lemma 2 ([21, Theorem 1.37]): A digraph $\mathcal{G}$ with N nodes is weight-balanced iff $\mathbf{1}_N^T\mathcal{L} = 0$.

C. Problem Setup

Consider N agents communicating over a digraph $\mathcal{G}(\mathcal{V}, \mathcal{E}, \mathcal{W})$, aiming to seek a global minimizer of (1), denoted by z^*, cooperatively. Each agent $i \in \mathcal{V}$ is associated to the local cost function $f_i(\cdot)$ of (1).

Assumption 1: The global cost function $F(\cdot)$ is differentiable and strictly convex over $\mathbb{R}^n$. Each local cost function $f_i(\cdot)$ is differentiable; and $\nabla f_i(x) = \Upsilon x + \psi_i(x)$, where $\Upsilon \in \mathbb{R}^{n\times n}$ with $\Upsilon \geq 0$, and $\|\psi_i(x)\| \leq K$ for some $K \in \mathbb{R}^+$ (could be unknown), for all $x \in \mathbb{R}^n$ and all $i \in \mathcal{V}$.

Assumption 2: The digraph $\mathcal{G}$ is strongly connected.

Remark 2: Strict convexity of $F(\cdot)$ is sufficient for the existence and uniqueness of the optimizer z^* of (1), and is milder than strong convexity[2] of $F(\cdot)$ required in [9], [11]. The structure of ∇f_i encompasses standard assumptions in literature, such as the boundedness assumption in [1], [8], [22] when $\Upsilon = 0$, and the decoupled multivariable assumption in [3] when $\Upsilon = \sigma\mathbf{I}_n$ with scalar $\sigma \geq 0$. Classes of functions with bounded gradients $\psi_i(\cdot)$ include trigonometric functions, logarithmic and fractional loss functions used for classification/regression [10], among others [11]. Furthermore, differently from [1]–[8], no convexity of $f_i(\cdot)$ is assumed here.

Remark 3: Under Assumption 2, one can find at least one DST rooting at any agent. The assumption on strong connectivity is standard in the distributed optimization literature

[2]Strong convexity means there exists $M \in \mathbb{R}^+$, such that $(x - y)^T(\nabla F(x) - \nabla F(y)) > M\|x - y\|^2$, $\forall x, y \in \mathbb{R}^n$ with $x \neq y$.

[4]–[6], [9], and it is more general than the case of undirected connected graphs [1]–[3], [10], [11].

The goal of this paper is to solve (1) in a systematic way for both weight-balanced and unbalanced digraphs. Let us start with the weight-balanced digraph case.

III. Adaptive Saddle-Point-like Dynamics for Weight-Balanced Digraphs

In [4, Eqn. (11)], static saddle-point dynamics with a fixed coupling gain α has been proposed for solving (1) over weight-balanced digraphs, with α determined based on the knowledge of the Laplacian eigenvalues. In this section, a novel *adaptive* saddle-point-like dynamics with dynamic coupling gains is proposed to resolve problem (1) without such knowledge.

Let each agent $i \in \mathcal{V}$ keep a local estimation $x_i \in \mathbb{R}^n$ of the optimal decision variable z^*, and an auxiliary variable $y_i \in \mathbb{R}^n$. Each agent i can only receive information from its in-neighbors for adjusting its own x_i and y_i. Let the agents communicate y_i over $\mathcal{G}$ and communicate x_i over $\mathcal{G}^{\mathcal{A}}(\mathcal{V},\mathcal{E},\mathcal{A}(t))$, where $\mathcal{A}(t) = (\alpha_{ij}(t))$ is the weight matrix of dynamic coupling gains with $\mathcal{A}(0) = \mathcal{W}$. Upon selecting[3] any DST $\bar{\mathcal{G}}$ of $\mathcal{G}$, consider the following algorithm:

$$\begin{aligned}\dot{x}_i =& -\gamma_1 \nabla f_i(x_i) - \sum_{j\in\mathcal{N}_{\text{in}}(i)} \alpha_{ij}(t)(x_i - x_j)\\ & - \sum_{j\in\mathcal{N}_{\text{in}}(i)} w_{ij}(y_i - y_j) \qquad (4a)\\ \dot{y}_i =& \sum_{j\in\mathcal{N}_{\text{in}}(i)} \alpha_{ij}(t)(x_i - x_j) \qquad (4b)\end{aligned}$$

with dynamic coupling gains

$$\alpha_{ij}(t) = \begin{cases} w_{ij}, & \text{if } e_{ji} \in \mathcal{E}\setminus\bar{\mathcal{E}},\\ \bar{a}_{k+1,i_k}(t), & \text{if } e_{ji}\in\bar{\mathcal{E}}\end{cases} \qquad (5a)$$

$$\dot{\bar{a}}_{k+1,i_k} = \gamma_2\Big((x_{i_k} - x_{k+1}) - \sum_{j\in\bar{\mathcal{N}}_{\text{out}}(k+1)} (x_{k+1} - x_j)\Big)^T (x_{i_k} - x_{k+1}) \qquad (5b)$$

for $k \in \mathcal{I}_{N-1}$, where $\gamma_1, \gamma_2 \in \mathbb{R}^+$. From (5), the coupling gain between agent i and its in-neighbor j is updating only when they communicate x_j, and when the edge e_{ji} appears in $\bar{\mathcal{G}}$ (i.e., $j = i_k$ and $i = k+1$ for some $k \in \mathcal{I}_{N-1}$). Upon defining the variables for the whole network $x = \text{col}(x_1, \cdots, x_N)$ and $y = \text{col}(y_1, \cdots, y_N)$, the algorithm (4) reads

$$\begin{aligned}\dot{x} =& -\gamma_1\nabla f(x) - (\mathcal{L}^{\mathcal{A}}(t)\otimes \mathbf{I}_n)x - (\mathcal{L}\otimes\mathbf{I}_n)y \qquad (6a)\\ \dot{y} =& (\mathcal{L}^{\mathcal{A}}(t)\otimes\mathbf{I}_n)x \qquad (6b)\end{aligned}$$

where $f(x) \triangleq \sum_{i=1}^N f_i(x_i) : \mathbb{R}^{Nn} \to \mathbb{R}$ is the cumulative cost function of the network state x, with $\nabla f(x) = \text{col}(\nabla f_1(x_1), \cdots, \nabla f_N(x_N))$ being its gradient. Here and in the following, let us explicitly use the time index t only for the matrices related to $\mathcal{G}^{\mathcal{A}}$ (e.g., the Laplacian $\mathcal{L}^{\mathcal{A}}(t)$) to highlight the time-varying property of the weights in $\mathcal{A}(t)$.

[3] Note that for a strongly connected digraph, a DST can be found in a distributed way without any knowledge of the Laplacian matrix [23].

Lemma 3: Suppose that $\mathcal{G}$ is weight-balanced and Assumptions 1-2 hold. If $(\tilde{x}, \tilde{y})$ is an equilibrium point of (6), then $\tilde{x} = \mathbf{1}_N \otimes z^*$, i.e., the global minimizer of (1).

Proof. We first obtain the equilibrium point of (6), $(\tilde{x}, \tilde{y})$, by

$$\begin{aligned}0 =& -\gamma_1\nabla f(\tilde{x}) - (\mathcal{L}^{\mathcal{A}}(t)\otimes\mathbf{I}_n)\tilde{x} - (\mathcal{L}\otimes\mathbf{I}_n)\tilde{y} \qquad (7a)\\ 0 =& (\mathcal{L}^{\mathcal{A}}(t)\otimes\mathbf{I}_n)\tilde{x}. \qquad (7b)\end{aligned}$$

By statement 1) of Lemma 1, $\mathbf{1}_N$ is the right eigenvector of $\mathcal{L}$ and $\mathcal{L}^{\mathcal{A}}(t)$, $\forall t$, associated to the corresponding simple zero eigenvalue, so it is guaranteed by (7b) that $\tilde{x} = \mathbf{1}_N \otimes z$, for some $z \in \mathbb{R}^n$. By Lemma 2, we have $\mathbf{1}_N^T\mathcal{L} = 0$. After left-multiplying (7a) by $\mathbf{1}_N^T \otimes \mathbf{I}_n$, we get $\sum_{i=1}^N \nabla f_i(z) = 0$, i.e., $\nabla F(z) = 0$, which implies $\tilde{x} = \mathbf{1}_N \otimes z^*$ due to the strict convexity of $F(\cdot)$. Note that if $(\tilde{x}, \tilde{y})$ is an equilibrium of (6), so is $(\tilde{x}, \tilde{y} + \mathbf{1}_N\otimes\kappa)$, for any $\kappa\in\mathbb{R}^n$. ■

Remark 4: Problem (1) is equivalent to the following constrained non-convex optimization problem:

$$\min_{x\in\mathbb{R}^{Nn}} f(x), \text{ subject to } x_1 = x_2 = \cdots = x_N, \qquad (8)$$

where non-convexity arises because $f(\cdot)$ (defined after (6)) may be non-convex on $\mathbb{R}^{Nn}$ under Assumption 1. However, instead of solving a non-convex problem as in standard literature [24], we approach problem (8) via a novel saddle-point-like dynamics. The algorithm (6) can also be interpreted as a gradient descent in x and a gradient ascent in y [4].

Based on Lemma 3, the remaining is to show that each trajectory of (6) converges to an equilibrium point. Let us transfer any equilibrium $(\tilde{x}, \tilde{y})$ to the origin and apply a change of coordinates:

$$\begin{aligned}\mu &= x - \tilde{x}, & \nu &= y - \tilde{y} \qquad (9a)\\ \bar{\mu} &= (\Xi\otimes\mathbf{I}_n)\mu, & \bar{\nu} &= (\Xi\otimes\mathbf{I}_n)\nu \qquad (9b)\end{aligned}$$

where Ξ is defined as in (2). In component-wise form, $\bar{\mu} = \text{col}(\bar{\mu}_1, \cdots, \bar{\mu}_{N-1})$ where $\bar{\mu}_k = \mu_{i_k} - \mu_{k+1}$, $k\in\mathcal{I}_{N-1}$. In these new coordinates, the algorithm (4) and the adaptive law (5b) read

$$\begin{aligned}\dot{\bar{\mu}} &= -\gamma_1(\Xi\otimes\mathbf{I}_n)h - (Q^{\mathcal{A}}(t)\otimes\mathbf{I}_n)\bar{\mu} - (Q\otimes\mathbf{I}_n)\bar{\nu} \qquad (10a)\\ \dot{\bar{\nu}} &= (Q^{\mathcal{A}}(t)\otimes\mathbf{I}_n)\bar{\mu} \qquad (10b)\\ \dot{\bar{a}}_{k+1,i_k} &= \gamma_2\Big(\bar{\mu}_k - \sum_{j\in\bar{\mathcal{N}}_{\text{out}}(k+1)}\bar{\mu}_{j-1}\Big)^T\bar{\mu}_k \qquad (10c)\end{aligned}$$

where $h = \nabla f(\mu+\tilde{x}) - \nabla f(\tilde{x})$, and Q as well as $Q^{\mathcal{A}}(t)$, $\forall t$, are defined as in (3) based on the DST $\bar{\mathcal{G}}$. More specifically, $Q^{\mathcal{A}}(t) = \tilde{Q} + \bar{Q}^{\mathcal{A}}(t)$ contains the fixed matrix $\tilde{Q}$, and the time-varying matrix

$$\bar{Q}^{\mathcal{A}}_{kj}(t) = \begin{cases}\bar{a}_{j+1,i_j}(t), & \text{if } j = k,\\ -\bar{a}_{j+1,i_j}(t), & \text{if } j = i_k - 1,\\ 0, & \text{otherwise.}\end{cases} \qquad (11)$$

To get (10), we have used statement 2) of Lemma 1 and the properties of the Kronecker product.

Lemma 4: For system (10) with arbitrary initial conditions, $(\bar{\mu}, \bar{\nu})$ asymptotically converges to the origin, and the weights $\bar{a}_{k+1,i_k}$, $k\in\mathcal{I}_{N-1}$, converge to some finite constant values.

Proof. See the appendix. ■

Now we formulate the main theorem of this section:

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TAC.2021.3071651, IEEE Transactions on Automatic Control

Theorem 1: Suppose that $\mathcal{G}$ is weight-balanced and Assumptions 1-2 hold. Then, algorithm (4) along with adaptive law (5) drives x_i to z^* asymptotically, for all $i \in \mathcal{V}$, and for any $x_i(0)$, $y_i(0) \in \mathbb{R}^n$. Moreover, the weights $\bar{a}_{k+1,i_k}$, $k \in \mathcal{I}_{N-1}$, in $\bar{\mathcal{G}}^{\mathcal{A}}$ converge to some finite constant values.

Proof. In the original coordinates, $(x,\ y)$ in (6) converges to $(\tilde{x} + \mathbf{1}_N \otimes \tau,\ \tilde{y} + \mathbf{1}_N \otimes \kappa)$, for some $\tau, \kappa \in \mathbb{R}^n$, due to Lemma 4 and the fact that the null-space of Ξ is spanned by $\mathbf{1}_N$ [17, Lemma 3.2].

Next, we show that $\tau = 0$ by seeking a contradiction based on the uniqueness of the optimizer z^*. Assume $\tau \neq 0$. Then, the steady-state dynamics of τ can be obtained by (6) as

$$\begin{aligned} 0 = \dot{\tau} =& \frac{1}{N}(\mathbf{1}_N^T \otimes \mathbf{I}_n)\dot{x} \\ =& -\frac{\gamma_1}{N}\nabla F(z^* + \tau) - \frac{1}{N}(\mathbf{1}_N^T \mathcal{L}^{\mathcal{A}}(t)\mathbf{1}_N \otimes (z^* + \tau)) \\ &- \frac{1}{N}(\mathbf{1}_N^T \mathcal{L} \otimes \mathbf{I}_n)(\tilde{y} + \mathbf{1}_N \otimes \kappa) \\ =& -\frac{\gamma_1}{N}\nabla F(z^* + \tau) \neq 0, \end{aligned}$$

which is a contradiction. Thus $\tau = 0$.

The above shows that any trajectory of (6) converges to an equilibrium point $(\tilde{x}, \tilde{y} + \mathbf{1}_N \otimes \kappa)$, for some $\kappa \in \mathbb{R}^n$. By lemma 3, the agents' estimates converge to the optimizer z^* of (1). ■

Remark 5: In line with classical adaptive control [25], the upper bound K in Assumption 1 is only used for proving convergence; thus, it can be unknown. Convergence of algorithm (4) can be guaranteed globally for any initial $x(0), y(0) \in \mathbb{R}^{Nn}$ and any parameters $\gamma_1, \gamma_2 \in \mathbb{R}^+$. These last two parameters can be tuned taking into account that increasing γ_1 allows larger step sizes towards decreasing the local costs, and increasing γ_2 enhances the importance of communicating the estimates of the global minimizer.

Remark 6: To elaborate on the features of the proposed algorithm with respect to the most recent state-of-the-art [9] (in the balanced digraph case), consider that the algorithm (4)-(5) allows only certain gains along a DST to be adaptive, while all the coupling gains in the network are made adaptive in [9]. For the stability analysis, Theorem 1 is built upon the identification of a proper form of local gradients, which differs from [9] as the invariant set analysis used therein is not needed for our approach. The benefits of the DST perspective become even more evident in the unbalanced digraph scenario considered in Sect. IV, as the proposed solution overcomes the knowledge of the Laplacian matrix required in [9] (cf. Remark 6 in [9]).

IV. Distributed Finite-time Weight-Balancing

The balanced property may not be satisfied in many cases. Our interest now is to extend the framework of Sect. III by including a finite-time weight-balancing algorithm consistently based on the same DST $\bar{\mathcal{G}}$ as the one used in Sect. III. Accordingly, consider the following balancing law:

$$\beta_{ij}(t) = \begin{cases} w_{ij}, & \text{if } e_{ji} \in \mathcal{E} \setminus \bar{\mathcal{E}}, \\ \bar{b}_{k+1,i_k}(t), & \text{if } e_{ji} \in \bar{\mathcal{E}} \end{cases} \tag{12a}$$

$$\dot{\bar{b}}_{k+1,i_k} = -\gamma_3 \mathrm{sig}^q\Big(\sum_{p \in \mathcal{N}_{\text{in}}(k+1)} \beta_{k+1,p} - \sum_{c \in \mathcal{N}_{\text{out}}(k+1)} \beta_{c,k+1} \Big) \tag{12b}$$

for $k \in \mathcal{I}_{N-1}$, where $q \in (0,1)$, $\gamma_3 \in \mathbb{R}^+$. Moreover, $\beta_{ij}(0) = w_{ij}$, $\forall i, j \in \mathcal{V}$. Denote $\bar{b} = (\bar{b}_{2,i_1}, \bar{b}_{3,i_2}, \cdots, \bar{b}_{N,i_{N-1}})^T$.

The design of (12b) is partly inspired by finite-time stabilization [26]. The intuition behind (12b) is to adjust the weight $\bar{b}_{k+1,i_k}$ in order to balance the node $k+1$, $k \in \mathcal{I}_{N-1}$. Two questions must be addressed:

1) whether these nodes $2, \cdots, N$ will be balanced;
2) if the answer to question 1) is possible, whether the root itself will be balanced subsequently.

Question 2) has a positive answer thanks to the following:

Lemma 5: The digraph $\mathcal{G}$ is weight-balanced iff $\mathbf{1}_N^T \mathcal{L}_{2:N} = 0$, where $\mathcal{L}_{2:N}$ is the block submatrix of $\mathcal{L}$ containing the second to the last columns.

Proof. Since $\mathcal{L}\mathbf{1}_N = 0$, we have $\mathbf{1}_N^T \mathcal{L} = 0$ iff $\mathbf{1}_N^T \mathcal{L}_{2:N} = 0$. Then, the lemma holds following Lemma 2. ■

In order to answer question 1), some preliminary analysis is needed.

Lemma 6: Suppose $\bar{\mathcal{G}}$ is a DST of $\mathcal{G}$. Let us take the weights in $\bar{\mathcal{G}}$ as independent variables $\bar{b}_{k+1,i_k}$, $k \in \mathcal{I}_{N-1}$, and denote the according Laplacian (resp. weighted adjacency) matrix with independent variables as $\mathcal{L}^{\mathcal{B}}$ (resp. $\mathcal{W}^{\mathcal{B}}$). Then, there exists a unique solution $\bar{b}^* := (\bar{b}^*_{2,i_1}, \bar{b}^*_{3,i_2}, \cdots, \bar{b}^*_{N,i_{N-1}})^T \in \mathbb{R}^{N-1}$ to the system of implicit linear equations $\mathbf{1}_N^T \mathcal{L}^{\mathcal{B}}_{2:N} = 0$.

Proof. Note that the k-th row of $\mathcal{W}^{\mathcal{B}}$, $k \in \{2, \cdots, N\}$, contains exactly one variable entry $\bar{b}_{k,i_{k-1}}$. Then, for the implicit linear equations $\mathbf{1}_N^T \mathcal{L}^{\mathcal{B}}_{2:N} = 0$ with variables $\bar{b}_{k+1,i_k}$, $k \in \mathcal{I}_{N-1}$, the coefficient matrix is square and full rank, indicating the existence and uniqueness of the solution. ■

Combining Lemma 5 with Lemma 6, we have that a balanced digraph is obtained by replacing the weights in $\bar{\mathcal{G}}$ with the corresponding components of $\bar{b}^*$, while keeping the other weights of $\mathcal{G}$ unchanged. More precisely, the resulting digraph, call it $\mathcal{G}^*_{\text{pseudo}}$, is actually a pseudo-balanced (i.e., balanced with possibly negative weights) digraph since $\bar{b}^*_{k+1,i_k}$ may be less than zero. However, $\mathcal{G}^*_{\text{pseudo}}$ can be used to obtain an actual balanced graph $\mathcal{G}^*$, upon changing the edges of $\mathcal{G}^*_{\text{pseudo}}$ whenever necessary. To illustrate how the balancing process works, let us give a simple example in Fig. 1.

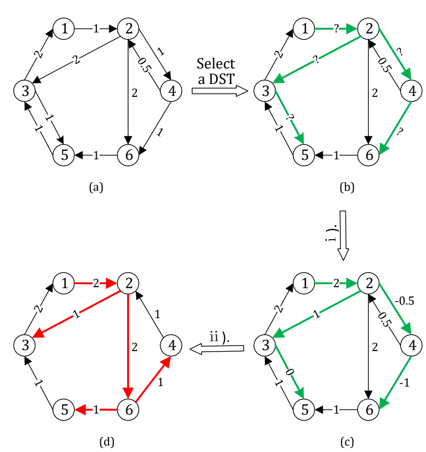

Fig. 1. (a): An unbalanced digraph $\mathcal{G}$; (b): $\mathcal{G}$ with a DST $\bar{\mathcal{G}}$, highlighted in green, where the marks "?" are weights to be adjusted; (c): The pseudo-balanced digraph $\mathcal{G}^*_{\text{pseudo}}$; (d): The balanced digraph $\mathcal{G}^*$ with a new DST $\bar{\mathcal{G}}^*$, highlighted in red.

The proposed balancing procedure consists of two steps:

i). Determine the weights in $\bar{\mathcal{G}}$ to obtain the pseudo-balanced digraph $\mathcal{G}^*_{\text{pseudo}}$ by (12);

ii). If $\bar{b}^*_{k+1,i_k} = 0$ and $w_{i_k,k+1} \neq 0$, remove the edge $e_{i_k,k+1}$; if $\bar{b}^*_{k+1,i_k} = 0$ and $w_{i_k,k+1} = 0$, introduce two bidirectional edges $e_{i_k,k+1}$ and e_{k+1,i_k} with equal positive weights, e.g., unitary weights. If $\bar{b}^*_{k+1,i_k} < 0$, introduce an opposite edge e_{k+1,i_k} with weight $|\bar{b}^*_{k+1,i_k}|$ (or increase its weight by $|\bar{b}^*_{k+1,i_k}|$ if e_{k+1,i_k} already exists). This leads to an actual balanced digraph $\mathcal{G}^*$.

Remark 7: It is known that a weight-balanced and weakly connected digraph is also strongly connected [27, Theorem 1]. After balancing steps i)-ii), the obtained $\mathcal{G}^*$ is weakly connected and weight-balanced, thus it remains strongly connected.

Now we are in the position of formulating the main convergence results for (12), which is the answer to question 1).

Theorem 2: Let Assumption 2 hold. Then, the balancing law (12) drives $\bar{b}$ to $\bar{b}^*$ in finite time.

Proof. By Lemma 6, the uniqueness of $\bar{b}^*$ guarantees that the components $\bar{b}^*_{k+1,i_k}$ have the following recursive form:

$$\bar{b}^*_{k+1,i_k} = \begin{cases} \sum_{c\in\bar{\mathcal{N}}_{\text{out}}(k+1)} \bar{b}^*_{c,k+1} + \sum_{c\in\mathcal{N}_{\text{out}}(k+1)\setminus\bar{\mathcal{N}}_{\text{out}}(k+1)} w_{c,k+1} \\ \quad - \sum_{p\in\mathcal{N}_{\text{in}}(k+1),p\neq i_k} w_{k+1,p}, \\ \qquad\qquad \text{if } k+1 \text{ is a stem,} \\ \sum_{c\in\mathcal{N}_{\text{out}}(k+1)} w_{c,k+1} - \sum_{p\in\mathcal{N}_{\text{in}}(k+1),p\neq i_k} w_{k+1,p}, \\ \qquad\qquad \text{if } k+1 \text{ is a leaf.} \end{cases} \tag{13}$$

In the following, we prove the convergence of the weights in the tree from the bottom to the top. Let d denote the depth of $\bar{\mathcal{G}}$, i.e., the number of edges in the longest path of $\bar{\mathcal{G}}$.

First, let us denote $E_1 = \{k \in \mathcal{I}_{N-1} | k+1 \text{ is a leaf}\}$. For each $k \in E_1$, consider the scalar candidate Lyapunov function

$$V_1 = (\bar{b}_{k+1,i_k} - \bar{b}^*_{k+1,i_k})^2. \tag{14}$$

By (12)-(14), the time derivative of V_1 can be obtained as

$$\begin{aligned}\dot{V}_1 =& -2\gamma_3(\bar{b}_{k+1,i_k} - \bar{b}^*_{k+1,i_k})\text{sig}^q\Big(\bar{b}_{k+1,i_k} \\ & - \sum_{c\in\mathcal{N}_{\text{out}}(k+1)} w_{c,k+1} + \sum_{p\in\mathcal{N}_{\text{in}}(k+1),p\neq i_k} w_{k+1,p}\Big) \\ =& -2\gamma_3 V_1^{\frac{1+q}{2}}.\end{aligned}$$

According to [26, Theorem 4.2], we have $V_1 \equiv 0$, i.e., $\bar{b}_{k+1,i_k} \equiv \bar{b}^*_{k+1,i_k}$, $\forall t \geq T_1(k)$, where $T_1(k) : E_1 \to \mathbb{R}^+$ is the settling-time function given by

$$T_1(k) = \frac{|\bar{b}_{k+1,i_k}(0) - \bar{b}^*_{k+1,i_k}|^{1-q}}{\gamma_3(1-q)}.$$

Denote $T_1^* = \max\{T_1(k), k \in E_1\}$. Note that if $d = 1$, the root with index 1 is a hub. In this case, we have $E_1 = \mathcal{I}_{N-1}$, and the settling-time of algorithm (12) is explicitly given by $T_h^* = \max\{\frac{|\bar{b}_{k+1,i_k}(0)-\bar{b}^*_{k+1,i_k}|^{1-q}}{\gamma_3(1-q)}, k \in \mathcal{I}_{N-1}\}$.

In the case of $d > 1$, let us sequentially denote $E_s = \{k \in \mathcal{I}_{N-1} \setminus \bigcup_{l=1}^{s-1} E_l | \forall c \in \bar{\mathcal{N}}_{\text{out}}(k+1),\ c-1 \in \bigcup_{l=1}^{s-1} E_l\}$, $s = 2, \cdots, d$. Then, we have $i_k = 1$, $\forall k \in E_d$. For any $s \in \{2, \cdots, d\}$ and each $k \in E_s$, consider the same Lyapunov function $V_s = V_1$. By (12)-(14), the time derivative of V_s is

$$\begin{aligned}\dot{V}_s =& -2\gamma_3(\bar{b}_{k+1,i_k} - \bar{b}^*_{k+1,i_k})\text{sig}^q\Big(\bar{b}_{k+1,i_k} \\ & + \sum_{p\in\mathcal{N}_{\text{in}}(k+1),p\neq i_k} w_{k+1,p} - \sum_{c\in\bar{\mathcal{N}}_{\text{out}}(k+1)} \bar{b}_{c,k+1} \\ & - \sum_{c\in\mathcal{N}_{\text{out}}(k+1)\setminus\bar{\mathcal{N}}_{\text{out}}(k+1)} w_{c,k+1}\Big) \\ =& -2\gamma_3(\bar{b}_{k+1,i_k} - \bar{b}^*_{k+1,i_k})\text{sig}^q\Big(\bar{b}_{k+1,i_k} \\ & - \bar{b}^*_{k+1,i_k} - \sum_{c\in\bar{\mathcal{N}}_{\text{out}}(k+1)} (\bar{b}_{c,k+1} - \bar{b}^*_{c,k+1})\Big) \\ =& -2\gamma_3(\bar{b}_{k+1,i_k} - \bar{b}^*_{k+1,i_k})\text{sig}^q\Big(\bar{b}_{k+1,i_k} \\ & - \bar{b}^*_{k+1,i_k} - \sum_{k'\triangleq c-1, c\in\bar{\mathcal{N}}_{\text{out}}(k+1)} (\bar{b}_{k'+1,i_{k'}} - \bar{b}^*_{k'+1,i_{k'}})\Big).\end{aligned}$$

Combined with the definition of E_s, we have

$$\dot{V}_s = -2\gamma_3 V_s^{\frac{1+q}{2}},\ \forall t \geq T^*_{s-1}$$

where $T^*_s = \max\{T_s(k), k \in E_s\}$ and $T_s(k) : E_s \to \mathbb{R}^+$ is

$$T_s(k) = T^*_{s-1} + \frac{|\bar{b}_{k+1,i_k}(T^*_{s-1}) - \bar{b}^*_{k+1,i_k}|^{1-q}}{\gamma_3(1-q)}.$$

Thus, $V_s \equiv 0$, i.e., $\bar{b}_{k+1,i_k} \equiv \bar{b}^*_{k+1,i_k}$, $\forall t \geq T_s(k)$.

Till now, we know that $\bar{b}_{k+1,i_k} \equiv \bar{b}^*_{k+1,i_k}$, for all $k \in \bigcup_{s=1}^d E_s$, and all $t \geq T^*_d$. It is clear that all E_s are nonempty, disjoint, and $\bigcup_{s=1}^d E_s = \mathcal{I}_{N-1}$, i.e., $\forall k \in \mathcal{I}_{N-1}$, there exists a unique $s \in \{1, \cdots, d\}$ such that $k \in E_s$. Then, we can conclude that $\bar{b} \equiv \bar{b}^*$, $\forall t \geq T^*_d$. This completes the proof. ∎

Remark 8: For solving (1) over an unbalanced graph, the left eigenvector associated with the zero eigenvalue of the global network Laplacian is required in [9]. The approach proposed here eliminates the dependence on such information.

Remark 9: In state-of-the-art weight-balancing literature [12]–[15], all the edge weights are adjusted, and the convergence follows an infinite time for real weight balancing [12], [14], and follows a finite time for integer weight balancing [13]–[15]. Differently from these works, our perspective is to fix all weights in a digraph except those in a DST, leading to a balanced digraph in finite time with real weights.

With reference to finite-time convergence, it is worth noticing that it is lost upon selecting $q = 1$: in this case, (12b) degenerates into

$$\dot{\bar{b}}_{k+1,i_k} = -\gamma_3\Big(\sum_{p\in\mathcal{N}_{\text{in}}(k+1)} \beta_{k+1,p} - \sum_{c\in\mathcal{N}_{\text{out}}(k+1)} \beta_{c,k+1}\Big). \tag{15}$$

Theorem 3: Let Assumption 2 hold. Then, the balancing law (12a)+(15) with $\gamma_3 \in \mathbb{R}^+$ drives $\bar{b}$ to $\bar{b}^*$ asymptotically. In the special case that the root of $\bar{\mathcal{G}}$ is a hub, the convergence is exponential with rate γ_3, i.e., $\|\bar{b}(t) - \bar{b}^*\| = \|\bar{b}(0) - \bar{b}^*\| e^{-\gamma_3 t}$.

Proof. The proof is similar to that of Theorem 2, combined with LaSalle's invariance principle [25, Theorem 2.2]. The details are omitted due to the space limits. ∎

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TAC.2021.3071651, IEEE Transactions on Automatic Control

V. An illustrative example

Consider a network of 6 agents interacting on the strongly connected digraph $\mathcal{G}^*$ of Fig. 1(d). The local objective functions are defined over $z = (z_1, z_2)^T \in \mathbb{R}^2$ as

$$
\begin{aligned}
f_1(z) &= z_1^2 + z_1 z_2 + 3z_2^2 + 5\sin(z_1),\\
f_2(z) &= (z_1+1)^2 + z_1 z_2 + 3(z_2-1)^2 + 10\cos(z_2+1),\\
f_3(z) &= (z_1-2)^2 + z_1 z_2 + 3(z_2+2)^2 + 20\arctan(2z_1 z_2),\\
f_4(z) &= (z_1+3)^2 + z_1 z_2 + 3(z_2-3)^2 - 5\sin(z_1),\\
f_5(z) &= (z_1-4)^2 + z_1 z_2 + 3(z_2+4)^2 - 10\cos(z_2+1),\\
f_6(z) &= (z_1+5)^2 + z_1 z_2 + 3(z_2-5)^2 - 20\arctan(2z_1 z_2).
\end{aligned}
$$

It can be easily checked that Assumption 1 is satisfied and $\mathcal{G}^*$ is weight-balanced. Despite all the local functions being non-convex, the resulting global function F is strictly convex.

To verify Theorem 1, let us select the DST $\bar{\mathcal{G}}^*$ highlighted in red in Fig. 1(d), and $\gamma_1 = \gamma_2 = 0.5$ in the algorithm (4)-(5). The initial $x_i(0)$ and $y_i(0)$ are chosen from a Gaussian distribution with standard deviation 5. The gradients of the global function $\nabla F(\cdot)$ evaluated at each $x_i(t) = (x_{i1}(t), x_{i2}(t))^T$, $i \in \mathcal{V}$, are given in Fig. 2 in a logarithmic scale, showing that the agents cooperatively solve the global optimization problem (with $\nabla F(x_i) \to 0$). Fig. 2 also shows that the weights in the DST $\bar{\mathcal{G}}^*$ converge to some constants. For comparison, we implement a modified[4] algorithm based on [9, Eqn. (3)]:

$$
\begin{aligned}
\dot{x}_i =& -\gamma_1 \nabla f_i(x_i) - \alpha_i(t)\xi_i - \sum_{j\in\mathcal{N}_{\text{in}}(i)} w_{ij}(y_i - y_j)\\
\dot{y}_i =& \alpha_i(t)\xi_i, \quad \dot{\alpha}_i = \gamma_2 \|\xi_i\|^2
\end{aligned}
\tag{16}
$$

where $\xi_i = \sum_{j\in\mathcal{N}_{\text{in}}(i)} w_{ij}(x_i - x_j)$. With the same other parameters as before and $\alpha_i(0) = 1$, the results are shown in Fig. 3. As compared to Fig. 2, one can see that allowing only certain coupling gains along a DST to be adaptive can have the benefit of reducing oscillations and improving the convergence performance.

In order to highlight the necessity of an adaptive strategy and to support Remark 6 at the same time, the gradients $\nabla F(x_i)$ under four different pairs of γ_1 and γ_2 in (4)-(5) are provided in Fig. 4, where the *nonadaptive* strategy (upper left) with $\gamma_2 = 0$ fails to solve the optimization problem.

If the agents interact over the unbalanced digraph of Fig. 1(a), one can first perform the weight-balancing process (12) to get the balanced digraph $\mathcal{G}^*$. Let us select the DST $\bar{\mathcal{G}}$ highlighted in green in Fig. 1(b), and $\gamma_3 = q = 0.1$ in algorithm (12). As shown in Fig. 5, the balancing weights $\bar{b}_{k+1,i_k}(t)$ in $\bar{\mathcal{G}}$ converge to the corresponding $\bar{b}^*_{k+1,i_k}$, $\forall k \in \mathcal{I}_5$, (see Fig. 1(c)) in finite time, as guaranteed by Theorem 2.

To verify Theorem 3, let us balance the digraph Fig. 1(a) with (12a)+(15) instead of (12). Select $\gamma_3 = 0.1$, the balancing weights converge asymptotically, as shown in Fig. 6.

[4]In [9, Eqn. (3)], the gain for ξ_i is the sum of the adaptive variable α_i and a variable $\beta_i = \|\xi_i\|^2$, where the role of β_i is to keep α_i from shifting to high values at the beginning. Here, β_i can be safely removed and γ_2 can be moved inside $\dot{\alpha}_i$. These modifications make comparisons more direct.

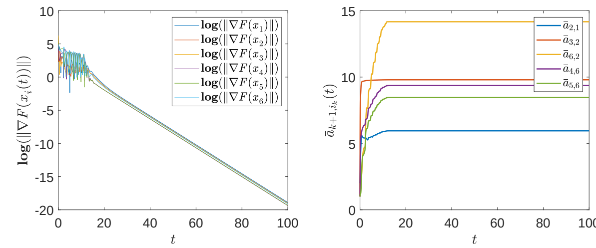

Fig. 2. Global gradient $\nabla F(\cdot)$ evaluated at $x_i(t)$ (in a logarithmic scale), $i \in \mathcal{V}$ (left). Adaptive coupling gains $\bar{a}_{k+1,i_k}(t)$, $k \in \mathcal{I}_5$, in $\bar{\mathcal{G}}^*$ (right). Here, $\gamma_1 = \gamma_2 = 0.5$.

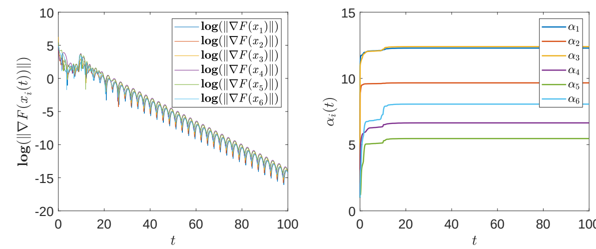

Fig. 3. Optimization and adaptation using (16) (based on [9, Eqn. (3)]).

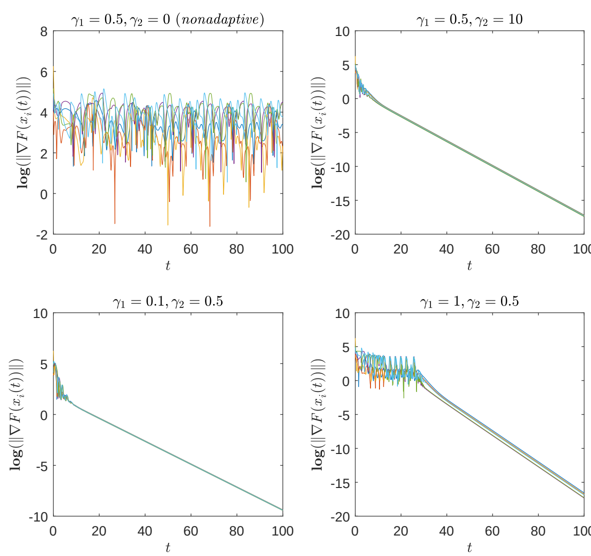

Fig. 4. Global gradient $\nabla F(\cdot)$ evaluated at $x_i(t)$ (in a logarithmic scale), $i \in \mathcal{V}$, for different parameters γ_1 and γ_2, as comparisons with Fig. 2 (left).

VI. Conclusions

A directed-spanning-tree based adaptive framework has been derived for distributed optimization of a summation-separable cost function with possibly non-convex local functions over strongly connected digraphs. Firstly, an adaptive algorithm has been proposed for weight-balanced digraphs. Secondly, the framework has been extended to unbalanced digraphs by including a novel finite-time weight balancing algorithm. We have shown that this framework allows to remove the knowledge of global Laplacian matrix required in the literature. Future research directions include extensions

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TAC.2021.3071651, IEEE Transactions on Automatic Control

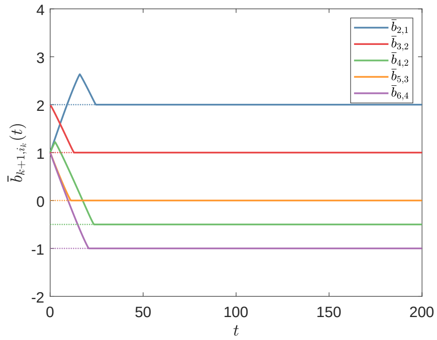

Fig. 5. Time-varying balancing weights $\bar{b}_{k+1,i_k}(t)$, $k \in \mathcal{I}_5$, in $\bar{\mathcal{G}}$. The dashed lines are the corresponding weights $\bar{b}^*_{k+1,i_k}$ in $\bar{\mathcal{G}}^*_{\text{pseudo}}$. Here, $\gamma_3 = q = 0.1$.

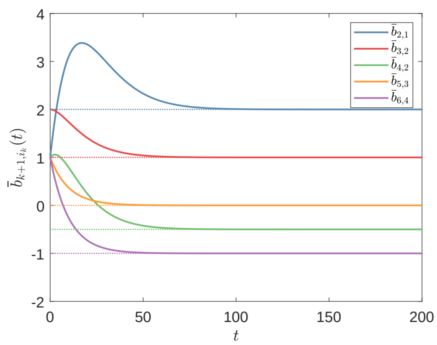

Fig. 6. Asymptotically balancing with (12a)+(15). Here, $\gamma_3 = 0.1$.

to constrained optimization [2] or to non-convex optimization even with non-convex global function [22]. Obtaining discrete-time counterparts of the proposed algorithms could also be an interesting open research direction.

Appendix
Proof of Lemma 4

The main idea of the proof is to utilize the positive definiteness of the matrix Q^{s} so as to entrust the stability of (10) to the adaptive coupling weights $\bar{a}_{k+1,i_k}$, $k \in \mathcal{I}_{N-1}$. Let us consider the following Lyapunov function:

$$V = \underbrace{\frac{1}{2}\bar{\mu}^T\bar{\mu} + \sum_{k=1}^{N-1}\frac{1}{2\gamma_2}(\bar{a}_{k+1,i_k}(t) - \phi_{k+1,i_k})^2}_{V_\mu} + \frac{3\lambda_{\max}(Q^TQ)}{\lambda_{\min}(Q^{\text{s}})}\cdot\underbrace{\frac{1}{2}(\bar{\mu}+\bar{\nu})^T(\bar{\mu}+\bar{\nu})}_{V_\nu} \tag{17}$$

where $Q^{\text{s}} > 0$ is guaranteed by statement 4) of Lemma 1, and $\phi_{k+1,i_k} \in \mathbb{R}^+$, $k \in \mathcal{I}_{N-1}$, are to be decided later.

The time derivative of V_μ along the trajectory of (10) is

$$\dot{V}_\mu = -\gamma_1\bar{\mu}^T(\Xi\otimes\mathbf{I}_n)h - \bar{\mu}^T(Q^{\mathcal{A}}(t)\otimes\mathbf{I}_n)\bar{\mu} - \bar{\mu}^T(Q\otimes\mathbf{I}_n)\bar{\nu} + \sum_{k=1}^{N-1}(\bar{a}_{k+1,i_k} - \phi_{k+1,i_k})(\bar{\mu}_k - \sum_{j+1\in\bar{\mathcal{N}}_{\text{out}}(k+1)}\bar{\mu}_j)^T\bar{\mu}_k. \tag{18}$$

From (11), one has

$$\begin{aligned}&\sum_{k=1}^{N-1}\bar{a}_{k+1,i_k}(\bar{\mu}_k - \sum_{j+1\in\bar{\mathcal{N}}_{\text{out}}(k+1)}\bar{\mu}_j)^T\bar{\mu}_k\\ &= \sum_{k=1}^{N-1}(\bar{Q}^{\mathcal{A}}_{kk}(t)\bar{\mu}_k + \sum_{j=1,j\neq k}^{N-1}\bar{Q}^{\mathcal{A}}_{jk}(t)\bar{\mu}_j)^T\bar{\mu}_k\\ &= \sum_{k=1}^{N-1}\sum_{j=1}^{N-1}\bar{Q}^{\mathcal{A}}_{jk}(t)\bar{\mu}_j^T\bar{\mu}_k = \bar{\mu}^T(\bar{Q}^{\mathcal{A}}(t)\otimes\mathbf{I}_n)\bar{\mu}.\end{aligned} \tag{19}$$

Let us define $\Phi \in \mathbb{R}^{(N-1)\times(N-1)}$ as

$$\Phi_{kj} = \begin{cases}\phi_{j+1,i_j}, & \text{if } j = k,\\ -\phi_{j+1,i_j}, & \text{if } j = i_k - 1,\\ 0, & \text{otherwise.}\end{cases} \tag{20}$$

Then, it follows from Assumption 1 and (18)-(20) that

$$\begin{aligned}\dot{V}_\mu =& -\gamma_1\bar{\mu}^T(\Xi\otimes\mathbf{I}_n)h - \bar{\mu}^T(Q^{\mathcal{A}}(t)\otimes\mathbf{I}_n)\bar{\mu}\\ &- \bar{\mu}^T(Q\otimes\mathbf{I}_n)\bar{\nu} + \bar{\mu}^T((\bar{Q}^{\mathcal{A}}(t) - \Phi)\otimes\mathbf{I}_n)\bar{\mu}\\ \leq& -\gamma_1\lambda_{\min}(\Upsilon)\bar{\mu}^T\bar{\mu} - \gamma_1\bar{\mu}^T(\Xi\otimes\mathbf{I}_n)h'\\ &- \bar{\mu}^T((\tilde{Q}+\Phi)\otimes\mathbf{I}_n)\bar{\mu} - \bar{\mu}^T(Q\otimes\mathbf{I}_n)\bar{\nu}\end{aligned} \tag{21}$$

where $h' = \psi(\mu+\tilde{x}) - \psi(\tilde{x})$ with $\psi(x) = \text{col}(\psi_1(x_1), \cdots, \psi_N(x_N))$. Using Young's inequality, we have

$$\begin{aligned}-\gamma_1\bar{\mu}^T(\Xi\otimes\mathbf{I}_n)h' &\leq \frac{\bar{\mu}^T\bar{\mu}}{2} + \frac{\gamma_1^2h'^T(\Xi^T\Xi\otimes\mathbf{I}_n)h'}{2}\\ &\leq \frac{\bar{\mu}^T\bar{\mu}}{2} + \frac{\gamma_1^2\lambda_{\max}(\Xi^T\Xi)h'^Th'}{2}\end{aligned} \tag{22}$$

$$\begin{aligned}-\bar{\mu}^T(Q\otimes\mathbf{I}_n)\bar{\nu} &\leq \frac{\bar{\mu}^T\bar{\mu}}{2} + \frac{\bar{\nu}^T(Q^TQ\otimes\mathbf{I}_n)\bar{\nu}}{2}\\ &\leq \frac{\bar{\mu}^T\bar{\mu}}{2} + \frac{\lambda_{\max}(Q^TQ)\bar{\nu}^T\bar{\nu}}{2}.\end{aligned} \tag{23}$$

Since all ψ_i are bounded by K, we have $\|\psi(x)\| \leq \sqrt{N}K$, for all $x \in \mathbb{R}^{Nn}$. Then,

$$h'^Th' \leq (\|\psi(\mu+\tilde{x})\| + \|\psi(\tilde{x})\|)^2 \leq 4NK^2. \tag{24}$$

It follows from (21)-(24) that

$$\begin{aligned}\dot{V}_\mu \leq& -\bar{\mu}^T((\tilde{Q}+\Phi+\gamma_1\lambda_{\min}(\Upsilon)\mathbf{I}_{N-1})\otimes\mathbf{I}_n)\bar{\mu} + \bar{\mu}^T\bar{\mu}\\ &+ \frac{\lambda_{\max}(Q^TQ)}{2}\bar{\nu}^T\bar{\nu} + 2NK^2\gamma_1^2\lambda_{\max}(\Xi^T\Xi).\end{aligned} \tag{25}$$

The time derivative of V_ν can be obtained as

$$\begin{aligned}\dot{V}_\nu =& -\gamma_1\bar{\mu}^T(\Xi\otimes\mathbf{I}_n)h - \bar{\mu}^T(Q\otimes\mathbf{I}_n)\bar{\nu}\\ &- \gamma_1\bar{\nu}^T(\Xi\otimes\mathbf{I}_n)h - \bar{\nu}^T(Q\otimes\mathbf{I}_n)\bar{\nu}\\ \leq& \frac{\bar{\mu}^T\bar{\mu}}{2} + 2NK^2\gamma_1^2\lambda_{\max}(\Xi^T\Xi) + \frac{\lambda_{\max}(Q^TQ)\bar{\mu}^T\bar{\mu}}{\lambda_{\min}(Q^{\text{s}})}\\ &+ \frac{\lambda_{\min}(Q^{\text{s}})\bar{\nu}^T\bar{\nu}}{4} + \frac{\lambda_{\min}(Q^{\text{s}})\bar{\nu}^T\bar{\nu}}{2}\\ &+ \frac{2NK^2\gamma_1^2\lambda_{\max}(\Xi^T\Xi)}{\lambda_{\min}(Q^{\text{s}})} - \lambda_{\min}(Q^{\text{s}})\bar{\nu}^T\bar{\nu}\\ =& \frac{\lambda_{\min}(Q^{\text{s}}) + 2\lambda_{\max}(Q^TQ)}{2\lambda_{\min}(Q^{\text{s}})}\bar{\mu}^T\bar{\mu} - \frac{\lambda_{\min}(Q^{\text{s}})}{4}\bar{\nu}^T\bar{\nu}\\ &+ \frac{2NK^2\gamma_1^2\lambda_{\max}(\Xi^T\Xi)(1+\lambda_{\min}(Q^{\text{s}}))}{\lambda_{\min}(Q^{\text{s}})}\end{aligned} \tag{26}$$

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TAC.2021.3071651, IEEE Transactions on Automatic Control

where Young's inequality and the positive definiteness of Q^{s} have been used to get the inequality.

From (25) and (26), $\dot{V}$ in (17) is upper bounded by

$$\begin{aligned}\dot{V} \leq & -\bar{\mu}^T((\tilde{Q} + \Phi + \gamma_1 \lambda_{\min}(\Upsilon)\mathbf{I}_{N-1}) \otimes \mathbf{I}_n)\bar{\mu} \\ & - \frac{\lambda_{\max}(Q^T Q)}{4}\bar{\nu}^T\bar{\nu} + \eta_1 \bar{\mu}^T\bar{\mu} + \eta_2\end{aligned}$$

where $\eta_1 = 1 + \frac{3\lambda_{\max}(Q^TQ)\lambda_{\min}(Q^{\mathrm{s}}) + 6\lambda_{\max}(Q^TQ)^2}{2\lambda_{\min}(Q^{\mathrm{s}})^2}$, and $\eta_2 = 2NK^2\gamma_1^2\lambda_{\max}(\Xi^T\Xi)(1 + \frac{3\lambda_{\max}(Q^TQ)(1+\lambda_{\min}(Q^{\mathrm{s}}))}{\lambda_{\min}(Q^{\mathrm{s}})^2})$.

Let $\delta \in \mathbb{R}^+$ be an arbitrarily small scalar. Then, when $\bar{\mu}^T\bar{\mu} \geq \delta$, there always exists a sufficiently large $\eta \in \mathbb{R}^+$ such that $\eta\bar{\mu}^T\bar{\mu} \geq \eta_1\bar{\mu}^T\bar{\mu} + \eta_2$. In this case, we have

$$\begin{aligned}\dot{V} \leq & -\bar{\mu}^T((\Phi^{\mathrm{s}} - \eta\mathbf{I}_{N-1} + \tilde{Q}^{\mathrm{s}} + \gamma_1\lambda_{\min}(\Upsilon)\mathbf{I}_{N-1}) \otimes \mathbf{I}_n)\bar{\mu} \\ & - \frac{\lambda_{\max}(Q^TQ)}{4}\bar{\nu}^T\bar{\nu}.\end{aligned}$$

A natural next step is to find some appropriate ϕ_{k+1,i_k} such that Φ^{s} is sufficiently positive definite to stabilize the system. First, we show that $\Phi^{\mathrm{s}} - \eta\mathbf{I}_{N-1} > 0$ where

$$\Phi^{\mathrm{s}} = \begin{pmatrix} \phi_{2,i_1} & \frac{1}{2}\phi_{21} & \cdots & \frac{1}{2}\phi_{N-2,1} & \frac{1}{2}\phi_{N-1,1} \\ \frac{1}{2}\phi_{21} & \phi_{3,i_2} & \cdots & \cdots & \frac{1}{2}\phi_{N-1,2} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \frac{1}{2}\phi_{N-2,1} & \vdots & \cdots & \phi_{N-1,i_{N-2}} & \frac{1}{2}\phi_{N-1,N-2} \\ \frac{1}{2}\phi_{N-1,1} & \frac{1}{2}\phi_{N-1,2} & \cdots & \frac{1}{2}\phi_{N-1,N-2} & \phi_{N,i_{N-1}} \end{pmatrix}.$$

To see this, let us denote $\Omega_1 = \begin{pmatrix} \phi_{2,i_1} - \eta \end{pmatrix}$, and $\Omega_k = \begin{pmatrix} \Omega_{k-1} & \varphi_k \\ \varphi_k^T & \phi_{k+1,i_k} - \eta \end{pmatrix}$, where $\varphi_k = \frac{1}{2}(\phi_{k1}, \phi_{k2}, \cdots, \phi_{k,k-1})^T$, $k = 2, \cdots, N-1$. Clearly, $\Omega_1 > 0$ by choosing $\phi_{2,i_1} > \eta$. Now suppose $\Omega_{k-1} > 0$, $k \geq 2$. Notice by (20) that $|\phi_{kj}| \leq |\phi_{j+1,i_j}|$, $\forall j \in \mathcal{I}_{k-1}$. Then, one has $\varphi_k^T\Omega_{k-1}^{-1}\varphi_k \leq \frac{\sum_{j=2}^k \phi_{j,i_{j-1}}^2}{4\lambda_{\min}(\Omega_{k-1})}$. By choosing $\phi_{k+1,i_k} > \eta + \frac{\sum_{j=2}^k \phi_{j,i_{j-1}}^2}{4\lambda_{\min}(\Omega_{k-1})}$, one has $\Omega_k > 0$ according to the Schur complement [28, Chapter 2.1]. By mathematical induction, $\Phi^{\mathrm{s}} - \eta\mathbf{I}_{N-1} = \Omega_{N-1}$ is positive definite.

Moreover, since $\tilde{Q}$ and Υ are fixed, one can always choose sufficiently large ϕ_{k+1,i_k} such that $\lambda_{\min}(\Phi^{\mathrm{s}} - \eta\mathbf{I}_{N-1} + \tilde{Q}^{\mathrm{s}}) > -\gamma_1\lambda_{\min}(\Upsilon)$. Then, we have $\dot{V} \leq 0$, for all $\bar{\mu}^T\bar{\mu} \geq \delta$ and all $\bar{\nu} \in \mathbb{R}^{(N-1)n}$. As a by-product, the trajectories of (10) are bounded. Note that since $\dot{V}$ is continuous, by LaSalle's invariance principle [25], we can conclude that $\bar{\mu}$ converges to the residual set $\mathcal{S} = \{\bar{\mu} | \|\bar{\mu}\|^2 \leq \delta\}$, and that $\bar{\nu}$ and $\bar{a}_{k+1,i_k}$ converge to some corresponding values such that $\dot{V} = 0$. Since δ can be arbitrarily small, it is guaranteed that both $\bar{\mu}$ and $\bar{\nu}$ converge to zero, and that the weights $\bar{a}_{k+1,i_k}$, $k \in \mathcal{I}_{N-1}$, converge to some finite constant values.

References

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

[2] Q. Liu, S. Yang, and Y. Hong, "Constrained consensus algorithms with fixed step size for distributed convex optimization over multiagent networks," *IEEE Trans. Autom. Control*, vol. 62, no. 8, pp. 4259–4265, 2017.

[3] P. Lin, W. Ren, and J. A. Farrell, "Distributed continuous-time optimization: nonuniform gradient gains, finite-time convergence, and convex constraint set," *IEEE Trans. Autom. Control*, vol. 62, no. 5, pp. 2239–2253, 2017.

[4] B. Gharesifard and J. Cortés, "Distributed continuous-time convex optimization on weight-balanced digraphs," *IEEE Trans. Autom. Control*, vol. 59, no. 3, pp. 781–786, 2013.

[5] S. S. Kia, J. Cortés, and S. Martínez, "Distributed convex optimization via continuous-time coordination algorithms with discrete-time communication," *Automatica*, vol. 55, pp. 254–264, 2015.

[6] R. Xin and U. A. Khan, "A linear algorithm for optimization over directed graphs with geometric convergence," *IEEE Control Syst. Lett.*, vol. 2, no. 3, pp. 315–320, 2018.

[7] S. Pu, W. Shi, J. Xu, and A. Nedić, "A push-pull gradient method for distributed optimization in networks," in *Proc. Conf. Decis. Control*. IEEE, 2018, pp. 3385–3390.

[8] B. Touri and B. Gharesifard, "A modified saddle-point dynamics for distributed convex optimization on general directed graphs," *IEEE Trans. Autom. Control*, vol. 65, no. 7, pp. 3098–3103, 2019.

[9] Z. Li, Z. Ding, J. Sun, and Z. Li, "Distributed adaptive convex optimization on directed graphs via continuous-time algorithms," *IEEE Trans. Autom. Control*, vol. 63, no. 5, pp. 1434–1441, 2017.

[10] D. Varagnolo, F. Zanella, A. Cenedese, G. Pillonetto, and L. Schenato, "Newton-raphson consensus for distributed convex optimization," *IEEE Trans. Autom. Control*, vol. 61, no. 4, pp. 994–1009, 2016.

[11] X. Wang, G. Wang, and S. Li, "Distributed finite-time optimization for integrator chain multiagent systems with disturbances," *IEEE Trans. Autom. Control*, vol. 65, no. 12, pp. 5296–5311, 2020.

[12] A. Priolo, A. Gasparri, E. Montijano, and C. Sagues, "A decentralized algorithm for balancing a strongly connected weighted digraph," in *Proc. Amer. Control Conf.* IEEE, 2013, pp. 6547–6552.

[13] B. Gharesifard and J. Cortés, "Distributed strategies for making a digraph weight-balanced," in *Proc. Annu. Allerton Conf. Commun. Control Comput.*, 2009, pp. 771–777.

[14] A. I. Rikos, T. Charalambous, and C. N. Hadjicostis, "Distributed weight balancing over digraphs," *IEEE Trans. Control Netw. Syst.*, vol. 1, no. 2, pp. 190–201, 2014.

[15] A. I. Rikos and C. N. Hadjicostis, "Distributed integer weight balancing in the presence of time delays in directed graphs," *IEEE Trans. Control Netw. Syst.*, vol. 5, no. 3, pp. 1300–1309, 2018.

[16] W. Yu, J. Lu, X. Yu, and G. Chen, "Distributed adaptive control for synchronization in directed complex networks," *SIAM J. Control Optim.*, vol. 53, no. 5, pp. 2980–3005, 2015.

[17] Z. Yu, D. Huang, H. Jiang, C. Hu, and W. Yu, "Distributed consensus for multiagent systems via directed spanning tree based adaptive control," *SIAM J. Control Optim.*, vol. 56, no. 3, pp. 2189–2217, 2018.

[18] D. Yue, S. Baldi, J. Cao, Q. Li, and B. De Schutter, "A directed spanning tree adaptive control solution to time-varying formations," *IEEE Trans. Control Netw. Syst.*, to be published, doi: 10.1109/TCNS.2021.3050332.

[19] W. Ren and R. W. Beard, *Distributed Consensus in Multi-Vehicle Cooperative Control*. Springer, 2008.

[20] C. W. Wu and L. O. Chua, "Synchronization in an array of linearly coupled dynamical systems," *IEEE Trans. Circuits Syst. I, Reg. Papers,*, vol. 42, no. 8, pp. 430–447, 1995.

[21] F. Bullo, J. Cortés, and S. Martinez, *Distributed Control of Robotic Networks*. Princeton University Press, 2009.

[22] T. Tatarenko and B. Touri, "Non-convex distributed optimization," *IEEE Trans. Autom. Control*, vol. 62, no. 8, pp. 3744–3757, 2017.

[23] P. Humblet, "A distributed algorithm for minimum weight directed spanning trees," *IEEE Trans. Commun.*, vol. 31, no. 6, pp. 756–762, 1983.

[24] P. Tseng, "On the rate of convergence of a partially asynchronous gradient projection algorithm," *SIAM J. Optim.*, vol. 1, no. 4, pp. 603–619, 1991.

[25] M. Krstić, I. Kanellakopoulos, and P. V. Kokotović, *Nonlinear and Adaptive Control Design*. Wiley New York, 1995.

[26] S. P. Bhat and D. S. Bernstein, "Finite-time stability of continuous autonomous systems," *SIAM J. Control Optim.*, vol. 38, no. 3, pp. 751–766, 2000.

[27] A. Stanoev and D. Smilkov, "Consensus theory in networked systems," in *Consensus and Synchronization in Complex Networks*. Springer, 2013, pp. 1–22.

[28] S. Boyd, L. El Ghaoui, E. Feron, and V. Balakrishnan, *Linear Matrix Inequalities in System and Control Theory*. SIAM, 1994, vol. 15.