IEEE CONTROL SYSTEMS LETTERS, VOL. 7, 2023

Nonconvex Distributed Optimization via Lasalle and Singular Perturbations

Guido Carnevale, *Graduate Student Member, IEEE*, and Giuseppe Notarstefano, *Member, IEEE*

***Abstract*—In this letter we address nonconvex distributed consensus optimization, a popular framework for distributed big-data analytics and learning. We consider the Gradient Tracking algorithm and, by resorting to an elegant system theoretical analysis, we show that agent estimates asymptotically reach consensus to a stationary point. We take advantage of suitable coordinates to write the Gradient Tracking as the interconnection of a *fast* dynamics and a *slow* one. To use a singular perturbation analysis, we separately study two auxiliary subsystems called *boundary layer* and *reduced* systems, respectively. We provide a Lyapunov function for the boundary layer system and use Lasalle-based arguments to show that trajectories of the reduced system converge to the set of stationary points. Finally, a customized version of a Lasalle's Invariance Principle for singularly perturbed systems is proved to show the convergence properties of the Gradient Tracking.**

***Index Terms*—Distributed control, control of networks, optimization, optimization algorithms.**

I. Introduction

DISTRIBUTED optimization has received significant attention from several scientific communities, see, e.g., [1]–[3] for an overview about problems and methods in this area. In particular, many tasks addressed in big-data analytics and deep learning can be posed as distributed consensus optimization problems, where network agents aim to minimize the sum of local cost functions depending on a common decision variable. Early references [4]–[6] address these problems by combining the gradient method with a consensus mechanism. Despite their fast convergence rate, these methods cannot achieve the exact solution due to the partial knowledge of the global objective function gradient. Exact convergence is achieved by the so-called *Gradient Tracking* algorithm, whose algorithmic structure relies on a "tracking action" based on dynamic average consensus (see [7], [8]) to locally reconstruct the global objective function gradient. The properties of the Gradient Tracking and its variants have been studied under different problem assumptions, see [9]–[16]. The big-data analytics and deep learning contexts motivate interest in the case of nonconvex objective functions. To the best of authors' knowledge, [17] is the only work studying the Gradient Tracking in case of nonconvex objective function over digraphs. In particular, it is shown that every limit point of the sequence generated by the Gradient Tracking is a stationary point of the given problem. Moreover, when the objective function satisfies the Kurdyka-Lojasiewicz property at any of its stationary points, the convergence of the entire sequence is proven. Our work tackles the same scenario but takes on a system theoretical perspective. Other works have shown the advantages of system theoretical tools for distributed optimization in the convex case as, e.g., [18]–[23]. In this letter the convergence of the Gradient Tracking to the set of stationary points is proven by leveraging the elegance of system theory and thus avoiding a long series of nontrivial inequalities needed to show the decreasing property of a Lyapunov-like function, [17]. Moreover, our work paves the way to (i) analyze the Gradient Tracking behavior in more general scenarios (like, e.g., the ones with error and/or disturbances due to, e.g., compressed communication [24], gradients unavailable [25], privacy mechanisms [26]) by exploiting additional well-established system theoretical tools (see, e.g., [27] where distributed optimization algorithm are systematically analyzed using tools from robust control), and (ii) develop algorithm extensions based on advanced control theoretic tools (see, e.g., [28] where event triggered versions of the Gradient Tracking are given in the convex case). In particular, we use the coordinates introduced in [23] to study a causal, state-space form of the algorithm. Following the approach in [29], we reformulate the system as the interconnection of a *fast dynamics* and a *slow* one thus obtaining a so-called *singularly perturbed system*. By taking advantage of this interpretation of the scheme, we separately study the identified subsystems. In particular, we build an auxiliary system called *boundary layer system* associated to the fast dynamics, and the so-called *reduced system* related to the slow one. For the boundary layer system, we provide a Lyapunov function independent of the slow state, thus proving the global exponential stability of the origin uniformly. Then, we use Lasalle-based arguments to show that trajectories of the reduced system globally converge to the set of problem stationary points. Finally, we merge these two results to assess the convergence of the

Manuscript received 21 March 2022; revised 31 May 2022; accepted 15 June 2022. Date of publication 1 July 2022; date of current version 15 July 2022. This work was supported by the European Research Council (ERC) through the European Union's Horizon 2020 Research and Innovation Programme under Grant 638992-OPT4SMART. Recommended by Senior Editor F. Dabbene. *(Corresponding author: Guido Carnevale.)*

The authors are with the Department of Electrical, Electronic and Information Engineering, Alma Mater Studiorum—Universitàdi Bologna, 40126 Bologna, Italy (e-mail: guido.carnevale@unibo.it; giuseppe.notarstefano@unibo.it).

Digital Object Identifier 10.1109/LCSYS.2022.3187918

original interconnection. This last step relies on a (generic) Lasalle's invariance principle for a specific class of singularly perturbed systems. To the best of the authors' knowledge, such a result is new and represents an additional contribution of the work. Also, in the strongly convex case, the same approach is used to show the linear convergence by simply applying the Polyak-Lojasiewicz inequality.

Section II sets up the problem and reformulates the Gradient Tracking. In Section III, a Lasalle's Invariance Principle for singularly perturbed systems is provided, and, in Section IV, we study the Gradient Tracking by resorting to singular perturbations. Section V shows a numerical simulation.

Notation: The identity matrix in $\mathbb{R}^{m\times m}$ is I_m. 1_N denotes the vector of N ones, while $\mathbf{1} := 1_N \otimes I_d$ with $\otimes$ being the Kronecker product. Dimensions are omitted whenever they are clear from the context. The vertical concatenation of the column vectors $v_1, \dots, v_N$ is $\text{COL}(v_1, \dots, v_N)$. Given a function $g : \mathbb{R}^n \to \mathbb{R}^m$, we define $\ker\{g(\cdot)\} := \{x \in \mathbb{R}^n \mid g(x) = 0\}$.

II. Problem Set-Up and Gradient Tracking Reformulation as a Singularly Perturbed System

A. Preliminaries

In this letter, we consider a network of N agents that aim at solving optimization problems in the form

$$\min_{\mathrm{x}\in\mathbb{R}^d} \sum_{i=1}^{N} f_i(\mathrm{x}), \tag{1}$$

with each function $f_i : \mathbb{R}^d \to \mathbb{R}$ known to agent i only. As formalized in the next assumption, we do not require the convexity of f_i, thus making this letter attractive for complex settings as, e.g., the ones involving big-data and deep learning (where nonconvex cost functions are often used).

Assumption 1 (Objective function): For all $i \in \{1, \dots, N\}$, $f_i : \mathbb{R}^d \to \mathbb{R}$ is of class $\mathcal{C}^1$ and has β-Lipschitz continuous gradient, for some $\beta > 0$. Moreover, $f(\mathrm{x}) := \sum_{i=1}^{N} f_i(\mathrm{x})$ is radially unbounded.

The communication among the N agents is modeled by a directed graph $\mathcal{G} = (\{1, \dots, N\}, \mathcal{E})$, with $\mathcal{E} \subset \{1, \dots, N\} \times \{1, \dots, N\}$ such that i can receive information from agent j only if the edge $(i, j) \in \mathcal{E}$. The set of neighbors of i is $\mathcal{N}_i := \{j \in \{1, \dots, N\} \mid (i, j) \in \mathcal{E}\}$. We associate to the graph $\mathcal{G}$ a weighted adjacency matrix $\mathcal{W}_{\mathcal{G}} \in \mathbb{R}^{N\times N}$ whose entries match the graph, i.e., $w_{ij} > 0$ whenever $(i, j) \in \mathcal{E}$ and $w_{ij} = 0$ otherwise. We point out that $(i, i) \in \mathcal{E}$ for all $i \in \{1, \dots, N\}$, so that to $w_{ii} > 0$ for all $i \in \{1, \dots, N\}$.

Assumption 2 (Network): The directed graph $\mathcal{G}$ is strongly connected and the matrix $\mathcal{W}_{\mathcal{G}}$ is doubly stochastic.

A popular method to address problem (1) is the Gradient Tracking algorithm. At each iteration $t \in \mathbb{N}$, each agent i maintains an estimate $\mathrm{x}_i^t \in \mathbb{R}^d$ of the solution of problem (1) and an auxiliary state $\mathrm{s}_i^t \in \mathbb{R}^d$ that are updated as

$$\mathrm{x}_i^{t+1} = \sum_{j\in\mathcal{N}_i} w_{ij}\mathrm{x}_j^t - \gamma \mathrm{s}_i^t \tag{2a}$$

$$\mathrm{s}_i^{t+1} = \sum_{j\in\mathcal{N}_i} w_{ij}\mathrm{s}_j^t + \nabla f_i(\mathrm{x}_i^{t+1}) - \nabla f_i(\mathrm{x}_i^t), \tag{2b}$$

where $\gamma > 0$ is a step-size, while each w_{ij} is the (i, j)-th entry of the adjacency matrix $\mathcal{W}_{\mathcal{G}}$ matching the graph $\mathcal{G}$. The initialization $\mathrm{s}_i^0 = \nabla f_i(\mathrm{x}_i^0)$ is required for all $i \in \{1, \dots, N\}$.

Following [23], we define $\mathrm{z}_i^t = \gamma(\mathrm{s}_i^t - \nabla f_i(\mathrm{x}_i^t))$ and introduce the following causal (still distributed) version of (2)

$$\mathrm{x}_i^{t+1} = \sum_{j\in\mathcal{N}_i} w_{ij}\mathrm{x}_j^t - \mathrm{z}_i^t - \gamma\nabla f_i(\mathrm{x}_i^t) \tag{3a}$$

$$\mathrm{z}_i^{t+1} = \sum_{j\in\mathcal{N}_i} w_{ij}\mathrm{z}_j^t - \gamma\nabla f_i(\mathrm{x}_i^t) + \gamma\sum_{j\in\mathcal{N}_i} w_{ij}\nabla f_j(\mathrm{x}_j^t). \tag{3b}$$

In an aggregate form Algorithm (3) reads as

$$\mathrm{x}^{t+1} = \mathcal{W}\mathrm{x}^t - \mathrm{z}^t - \gamma\nabla F(\mathrm{x}^t) \tag{4a}$$

$$\mathrm{z}^{t+1} = \mathcal{W}\mathrm{z}^t - \gamma(I_{Nd} - \mathcal{W})\nabla F(\mathrm{x}^t), \tag{4b}$$

where $\mathrm{x}^t := \text{COL}(\mathrm{x}_1^t, \dots, \mathrm{x}_N^t)$, $\mathrm{z}^t := \text{COL}(\mathrm{z}_1^t, \dots, \mathrm{z}_N^t)$, $\nabla F(\mathrm{x}^t) := \text{COL}(\nabla f_1(\mathrm{x}_1^t), \dots, \nabla f_N(\mathrm{x}_N^t))$, $\mathcal{W} := \mathcal{W}_{\mathcal{G}} \otimes I_d$. As shown in [23], in these new coordinates, the initialization of the auxiliary state reads as $\mathbf{1}^\top\mathrm{z}^0 = 0$.

B. Gradient Tracking as a Singularly Perturbed System

The aim of this section is to provide an equivalent reformulation of the Gradient Tracking. Let $R \in \mathbb{R}^{Nd\times(N-1)d}$ be such that $R^\top R = I$ and $R^\top\mathbf{1} = 0$, and define

$$\begin{bmatrix}\bar{\mathrm{x}}^t\\ \mathrm{x}_\perp^t\end{bmatrix} := \begin{bmatrix}\frac{\mathbf{1}^\top}{N}\\ R^\top\end{bmatrix}\mathrm{x}^t, \quad \begin{bmatrix}\bar{\mathrm{z}}^t\\ \mathrm{z}_\perp^t\end{bmatrix} := \begin{bmatrix}\frac{\mathbf{1}^\top}{N}\\ R^\top\end{bmatrix}\mathrm{z}^t. \tag{5}$$

Thus, we can rewrite (4) as

$$\begin{bmatrix}\bar{\mathrm{x}}^{t+1}\\ \mathrm{x}_\perp^{t+1}\\ \bar{\mathrm{z}}^{t+1}\\ \mathrm{z}_\perp^{t+1}\end{bmatrix} = \begin{bmatrix} I_d & 0 & \frac{\mathbf{1}}{N} & 0\\ 0 & R^\top\mathcal{W}R & 0 & -I_{(N-1)d}\\ 0 & 0 & I_d & 0\\ 0 & 0 & R^\top\mathcal{W}\mathbf{1} & R^\top\mathcal{W}R\end{bmatrix}\begin{bmatrix}\bar{\mathrm{x}}^t\\ \mathrm{x}_\perp^t\\ \bar{\mathrm{z}}^t\\ \mathrm{z}_\perp^t\end{bmatrix} + \gamma\begin{bmatrix}-\frac{\mathbf{1}^\top}{N}\\ -R^\top\\ R^\top(\mathcal{W} - I_{Nd})\\ 0\end{bmatrix}\nabla F(\mathbf{1}\bar{\mathrm{x}}^t + R\mathrm{x}_\perp^t). \tag{6}$$

The initialization $\mathbf{1}^\top\mathrm{z}^0 = 0$ guarantees that $\bar{\mathrm{z}}^t \equiv 0$ for all $t \geq 0$ so that we can neglect the dynamics of $\bar{\mathrm{z}}$ and study

$$\bar{\mathrm{x}}^{t+1} = \bar{\mathrm{x}}^t - \frac{\gamma}{N}\mathbf{1}^\top\nabla F\big(\mathbf{1}\bar{\mathrm{x}}^t + R\mathrm{x}_\perp^t\big) \tag{7a}$$

$$\begin{bmatrix}\mathrm{x}_\perp^{t+1}\\ \mathrm{z}_\perp^{t+1}\end{bmatrix} = A\begin{bmatrix}\mathrm{x}_\perp^t\\ \mathrm{z}_\perp^t\end{bmatrix} + \gamma B\nabla F\big(\mathbf{1}\bar{\mathrm{x}}^t + R\mathrm{x}_\perp^t\big), \tag{7b}$$

with

$$A := \begin{bmatrix}R^\top\mathcal{W}R & -I_{(N-1)d}\\ 0 & R^\top\mathcal{W}R\end{bmatrix}, B := \begin{bmatrix}-R^\top\\ R^\top(\mathcal{W} - I_{Nd})\end{bmatrix}. \tag{8}$$

System (7) fits the class of singularly perturbed systems (see, e.g., [30]) given by the interconnection between a slow dynamics, which in our case is (7a), and a fast one represented by (7b) (see Fig. 1 for a schematic representation). We point out that, numerically, it clearly appears that the convergence of subsystem (7b) is faster than the one of (7a).

For the sake of completeness, we provide some notation of singularly perturbed systems that will be used in the analysis. We denote boundary layer system the one obtained by "freezing" $\bar{\mathrm{x}} \in \mathbb{R}^d$ within (7b). As we will see later, such a system exhibits an equilibrium parametrized in $\bar{\mathrm{x}}$, say $h(\bar{\mathrm{x}}, \gamma)$. We denote as reduced system the one obtained by considering (7a) with $\text{COL}(\mathrm{x}_\perp^t, \mathrm{z}_\perp^t) = h(\bar{\mathrm{x}}^t, \gamma)$ for any $t \geq 0$.

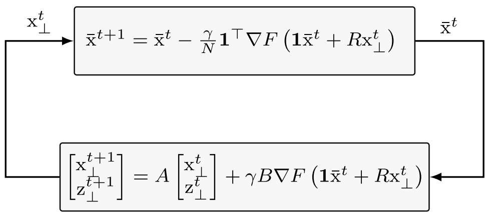

Fig. 1. Schematic representation of system (7).

III. A Lasalle's Invariance Principle for Singularly Perturbed Systems

In this section we provide a Lasalle's Invariance Principle for (discrete-time) singularly perturbed systems, which extends the Lyapunov theorem in, e.g., [30] (see [31, Ch. 11] for the continuous-time case).

Definition 1: Consider $\mathcal{M} \subseteq \mathbb{R}^n$ and $\mathrm{x}^{t+1} = T(\mathrm{x}^t)$, with $T : \mathbb{R}^n \to \mathbb{R}^n$ and $\mathrm{x} \in \mathbb{R}^n$. Let $T(\mathcal{M}) := \{\mathrm{y} \in \mathbb{R}^n | \mathrm{y} = T(\mathrm{x}) \text{ for some } \mathrm{x} \in \mathcal{M}\}$. $\mathcal{M}$ is invariant if $T(\mathcal{M}) \equiv \mathcal{M}$.

Theorem 1: Consider the system

$$\bar{\mathrm{x}}^{t+1} = \bar{\mathrm{x}}^t + \gamma\phi(\bar{\mathrm{x}}^t, \zeta^t) \tag{9a}$$
$$\zeta^{t+1} = g(\zeta^t, \bar{\mathrm{x}}^t, \gamma), \tag{9b}$$

with $\bar{\mathrm{x}}^t \in \mathbb{R}^n$, $\zeta^t \in \mathbb{R}^m$, $\phi : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^n$, $g : \mathbb{R}^m \times \mathbb{R}^n \times \mathbb{R} \to \mathbb{R}^m$, and $\gamma > 0$. Assume that ϕ and g are Lipschitz continuous in $\bar{\mathrm{x}}$ and ζ with parameters $L_1 > 0$ and $L_g(\gamma) > 0$, respectively, where L_g is continuous. Assume that there exists $h : \mathbb{R}^n \times \mathbb{R} \to \mathbb{R}^m$ such that $g(h(\bar{\mathrm{x}}, \gamma), \bar{\mathrm{x}}, \gamma) = h(\bar{\mathrm{x}}, \gamma)$ for any $\bar{\mathrm{x}} \in \mathbb{R}^n$ and that h is Lipschitz continuous in $\bar{\mathrm{x}}$ with parameter $L_h(\gamma) > 0$, where L_h is continuous. Let

$$\bar{\mathrm{x}}^{t+1} = \bar{\mathrm{x}}^t + \gamma\phi(\bar{\mathrm{x}}^t, h(\bar{\mathrm{x}}^t, \gamma)) \tag{10}$$

be the reduced system and

$$\psi^{t+1} = g(\psi^t + h(\bar{\mathrm{x}}, \gamma), \bar{\mathrm{x}}, \gamma) - h(\bar{\mathrm{x}}, \gamma) \tag{11}$$

be the boundary layer system with $\psi^t \in \mathbb{R}^m$. Assume that there exists $\bar{\gamma}_1 > 0$ such that, for any $\gamma \in (0, \bar{\gamma}_1)$, there exists a Lyapunov function $W : \mathbb{R}^m \to \mathbb{R}$ such that

$$b_1\|\psi\|^2 \leq W(\psi) \leq b_2\|\psi\|^2 \tag{12a}$$
$$W(g(\psi + h(\bar{\mathrm{x}}, \gamma), \bar{\mathrm{x}}, \gamma) - h(\bar{\mathrm{x}}, \gamma)) - W(\psi) \leq -b_3\|\psi\|^2 \tag{12b}$$
$$|W(\psi_1) - W(\psi_2)| \leq b_4\|\psi_1 - \psi_2\|(\|\psi_1\| + \|\psi_2\|), \tag{12c}$$

for any $\psi, \psi_1, \psi_2 \in \mathbb{R}^m$, $\bar{\mathrm{x}} \in \mathbb{R}^n$, and some $b_1, b_2, b_3, b_4 > 0$. Further, assume there exists $\bar{\gamma}_2 > 0$ and a radially unbounded function $U : \mathbb{R}^n \to \mathbb{R}$ such that

$$U(\bar{\mathrm{x}} + \gamma\phi(\bar{\mathrm{x}}, h(\bar{\mathrm{x}}, \gamma)) - U(\bar{\mathrm{x}}) \leq -c_1\|\phi(\bar{\mathrm{x}}, h(\bar{\mathrm{x}}, \gamma))\|^2 \tag{13a}$$
$$\begin{aligned} U(\bar{\mathrm{x}}_1 + \bar{\mathrm{x}}_2) - U(\bar{\mathrm{x}}_1 + \bar{\mathrm{x}}_3) &\leq c_2\|\phi(\bar{\mathrm{x}}_1, h(\bar{\mathrm{x}}_1, \gamma))\|\|\bar{\mathrm{x}}_2 - \bar{\mathrm{x}}_3\| \\ &\quad + c_3\left(\|\bar{\mathrm{x}}_2\|^2 + \|\bar{\mathrm{x}}_3\|^2\right), \end{aligned} \tag{13b}$$

for any $\gamma \in (0, \bar{\gamma}_2)$, $\bar{\mathrm{x}}, \bar{\mathrm{x}}_1, \bar{\mathrm{x}}_2, \bar{\mathrm{x}}_3 \in \mathbb{R}^n$, and some $c_1, c_2, c_3 > 0$. Then, there exists $\bar{\gamma} \in (0, \min\{\bar{\gamma}_1, \bar{\gamma}_2\})$ such that, for all $\gamma \in (0, \bar{\gamma})$, any trajectory of system (9) satisfies

$$\lim_{t\to\infty} \inf_{\xi\in\mathcal{M}} \left\| \begin{bmatrix}\bar{\mathrm{x}}^t\\ \zeta^t\end{bmatrix} - \begin{bmatrix}\xi\\ h(\xi, \gamma)\end{bmatrix} \right\| = 0,$$

where $\mathcal{M} \subseteq \ker\{\phi(\cdot, h(\cdot, \gamma))\} \subseteq \mathbb{R}^n$ denotes the largest invariant set for (10) contained within $\ker\{\phi(\cdot, h(\cdot, \gamma))\}$.

Proof: The proof is given in the Appendix. ■

IV. Singular Perturbation Analysis of the Gradient Tracking

Once the Gradient Tracking has been posed in the singularly perturbed form (7), we can separately study two auxiliary schemes associated to the subsystems (7a) and (7b). The *boundary layer system* is obtained by freezing the state of the slow dynamics in the fast one (7b). Notice that

$$h(\bar{\mathrm{x}}, \gamma) := \gamma\begin{bmatrix}0\\ -R^\top\nabla F(\mathbf{1}\bar{\mathrm{x}})\end{bmatrix} \tag{14}$$

is an equilibrium of system (7b) for any "frozen" $\bar{\mathrm{x}}$. Now, we introduce the error coordinates of the fast dynamics with respect to $h(\bar{\mathrm{x}}, \gamma)$. Let $\psi^t := \mathrm{COL}(\mathrm{x}_\perp^t, \mathrm{z}_\perp^t) - h(\bar{\mathrm{x}}, \gamma)$, we can write the boundary layer system associated to (7a) as

$$\psi^{t+1} = A\psi^t + \gamma B u_{\mathrm{bl}}^t(\bar{\mathrm{x}}), \tag{15}$$

where

$$u_{\mathrm{bl}}^t(\bar{\mathrm{x}}) := \nabla F\big(\mathbf{1}\bar{\mathrm{x}} + \big[R \;\; 0\big]\psi^t\big) - \nabla F(\mathbf{1}\bar{\mathrm{x}}). \tag{16}$$

Remark 1: From Assumption 2 the matrix $\mathcal{W}$ has 1 as an eigenvalue with multiplicity d, while all the remaining ones lie within the open unit circle. The left and right eigenvectors associated to 1 belong to the span of $\mathbf{1}^\top$ and $\mathbf{1}$, respectively. Thus, $R^\top\mathcal{W}R$ is Schur. Then, the matrix A, being up-triangular with two Schur matrices on the diagonal blocks, is Schur too. See [23] for a detailed discussion.

Exponential stability of (15) uniformly in $\bar{\mathrm{x}}$ is given now.

Lemma 1: Consider system (15). Let $m := 2(N-1)d$. Then, there exists $\bar{\gamma}_1 > 0$ and a Lyapunov function $W : \mathbb{R}^m \to \mathbb{R}$ such that, for any $\gamma \in (0, \bar{\gamma}_1)$, it holds

$$b_1\|\psi\|^2 \leq W(\psi) \leq b_2\|\psi\|^2 \tag{17a}$$
$$W(A\psi + \gamma u_{\mathrm{bl}}^t(\bar{\mathrm{x}})) - W(\psi) \leq -b_3\|\psi\|^2 \tag{17b}$$
$$|W(\psi_1) - W(\psi_2)| \leq b_4\|\psi_1 - \psi_2\|(\|\psi_1\| + \|\psi_2\|), \tag{17c}$$

for any $\psi, \psi_1, \psi_2 \in \mathbb{R}^m$, $\bar{\mathrm{x}} \in \mathbb{R}^n$, and some $b_1, b_2, b_3, b_4 > 0$.

Proof: Pick any $Q \in \mathbb{R}^{m\times m}$, $Q = Q^\top > 0$. Being A Schur (see Remark 1), there exists $P = P^\top > 0$ so that

$$A^\top PA - P = -Q. \tag{18}$$

Pick such P to define $W : \mathbb{R}^m \to \mathbb{R}$ as

$$W(\psi^t) := (\psi^t)^\top P\psi^t,$$

which clearly satisfies (17a) and (17c). Further, along (15), $\Delta W(\psi^t) := W(\psi^{t+1}) - W(\psi^t)$ is given by

$$\begin{aligned} \Delta W(\psi^t) &= (\psi^t)^\top(A^\top PA - P)\psi^t + 2\gamma(\psi^t)^\top A^\top PBu_{\mathrm{bl}}^t(\bar{\mathrm{x}}) \\ &\quad + \gamma^2(u_{\mathrm{bl}}^t(\bar{\mathrm{x}}))^\top B^\top PBu_{\mathrm{bl}}^t(\bar{\mathrm{x}}) \\ &\overset{(a)}{\leq} -(\psi^t)^\top Q\psi^t + 2\gamma\left\|A^\top PB\right\|\left\|\psi^t\right\|\left\|u_{\mathrm{bl}}^t(\bar{\mathrm{x}})\right\| \\ &\quad + \gamma^2\left\|B^\top PB\right\|\left\|u_{\mathrm{bl}}^t(\bar{\mathrm{x}})\right\|^2, \end{aligned} \tag{19}$$

where in (a) we have used the result (18) and the Cauchy-Schwarz inequality. Being each ∇f_i Lipschitz continuous (cf. Assumption 1), we bound $u_{\mathrm{bl}}^t(\bar{\mathrm{x}})$ (defined in (16)) as

$$\left\|u_{\mathrm{bl}}^t(\bar{\mathrm{x}})\right\| \leq \beta\left\|\big[R \quad 0\big]\right\|\left\|\psi^t\right\|, \tag{20}$$

for any $\bar{x} \in \mathbb{R}^d$. Thus, we can use (20) to bound (19) as

$$\Delta W(\psi^t) \leq -(q - \gamma c_1 - \gamma^2 c_2)\left\|\psi^t\right\|^2, \tag{21}$$

where q is the (positive) smallest eigenvalue of Q, while $c_1 := 2\beta\|A^\top PB\|\|[R \quad 0]\|$, $c_2 := \beta^2\|B^\top PB\|\|[R \quad 0]\|^2$. Thus, there exists $\bar{\gamma}_1 > 0$ such that $q - \gamma c_1 - \gamma^2 c_2 > 0$ for any $\gamma \in (0, \bar{\gamma}_1)$ so that (17b) holds and the proof follows. ■

Now, we consider $\text{COL}(x_\perp^t, z_\perp^t) = h(\bar{x}^t, \gamma)$ for all $t \geq 0$ within (7a) obtaining the so-called reduced system as

$$\bar{x}^{t+1} = \bar{x}^t - \frac{\gamma}{N}\mathbf{1}^\top \nabla F\big(\mathbf{1}\bar{x}^t + \begin{bmatrix} R & 0 \end{bmatrix} h(\bar{x}^t, \gamma)\big). \tag{22}$$

By exploiting h (cf. (14)) and ∇F, we write system (22) as

$$\bar{x}^{t+1} = \bar{x}^t - \frac{\gamma}{N}\nabla f(\bar{x}^t). \tag{23}$$

The next lemma uses the (radially unbounded) function f to show the convergence of system (23) to the set $X_{\text{staz}} := \{\bar{x} \in \mathbb{R}^d \mid \nabla f(\bar{x}) = 0\}$ of stationary points of (1).

Lemma 2: Consider system (23) Then, there exists $\bar{\gamma}_2 > 0$ such that, for any $\gamma \in (0, \bar{\gamma}_2)$, it holds

$$f\left(\bar{x} - \frac{\gamma}{N}\nabla f(\bar{x})\right) - f(\bar{x}) \leq -d_1\|\nabla f(\bar{x})\|^2 \tag{24a}$$

$$f(x_1 + x_2) - f(x_1 + x_3) \leq d_2\|\nabla f(\bar{x}_1)\|\|x_2 - x_3\| + d_3\left(\|x_2\|^2 + \|x_3\|^2\right), \tag{24b}$$

for any $x, x_1, x_2, x_3 \in \mathbb{R}^d$, and some $d_1, d_2, d_3 > 0$.

Proof: In light of Assumption 1, f has Lipschitz continuous gradient. Thus, we apply the Descent Lemma (cf. [32, Proposition 6.1.2]) to write

$$f(\bar{x}^{t+1}) - f(\bar{x}^t) \leq -\tfrac{\gamma}{N}\left\|\nabla f(\bar{x}^t)\right\|^2 + \gamma^2\tfrac{\beta}{2N^2}\left\|\nabla f(\bar{x}^t)\right\|^2. \tag{25}$$

Thus, for any $\gamma \in (0, \bar{\gamma}_2)$ with $\bar{\gamma}_2 := \frac{\beta}{2N}$, the inequality (25) ensures that (24a) is satisfied. With same arguments, also the inequality (24b) with $d_2 = 1$ and $d_3 = \frac{\beta}{2}$ can be shown. ■

Once these preliminary results have been provided, we can use them in the next theorem to state the convergence properties of the Gradient Tracking distributed algorithm.

Theorem 2: Consider the Gradient Tracking given in (4). Let Assumptions 1 and 2 hold. Then, for any initial condition $(x^0, z^0) \in \mathbb{R}^{2Nd}$ such that $\mathbf{1}^\top z^0 = 0$, there exists $\bar{\gamma} > 0$ such that, for any $\gamma \in (0, \bar{\gamma})$, it holds

$$\lim_{t\to\infty}\inf_{\xi \in X_{\text{staz}}}\left\|x_i^t - \xi\right\| = 0, \quad \forall i \in \{1, \ldots, N\}.$$

Proof: The proof relies on Theorem 1. Indeed, Lemma 1 and Lemma 2 provide the functions W and $U \equiv f$ satisfying conditions (12) and (13), respectively. Further, Assumption 1 guarantees the radial unboundedness of U and the regularity properties required for ϕ, g, and h. Hence, by Theorem 1, there exists $\bar{\gamma} > 0$ such that, for any $\gamma \in (0, \bar{\gamma})$, any trajectory of system (7) satisfies

$$\lim_{t\to\infty}\inf_{\xi \in \mathcal{M}'}\left\|\begin{bmatrix} \bar{x}^t \\ x_\perp^t \\ z_\perp^t \end{bmatrix} - \begin{bmatrix} \xi \\ h(\xi, \gamma) \end{bmatrix}\right\|, \tag{26}$$

where $\mathcal{M}' \subseteq \ker\{\nabla f(\cdot)\} \subseteq \mathbb{R}^d$ denotes the largest invariant set for system (23) contained within $\ker\{\nabla f(\cdot)\}$. The proof follows by noting that $\mathcal{M}' \equiv \ker\{\nabla f(\cdot)\} \equiv X_{\text{staz}}$. ■

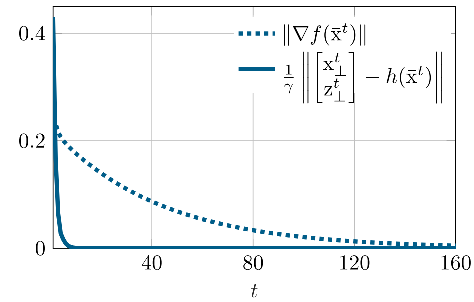

Fig. 2. Convergence of the fast and slow subsystem.

Next Proposition addresses the strongly convex case.

Proposition 1: Let f be μ-strongly convex with minimizer $x^\star \in \mathbb{R}^d$ and suppose the assumptions of Theorem 2 hold. Then, there exist $\bar{\gamma}, a_1, a_2 > 0$ such that for any $\gamma \in (0, \bar{\gamma})$

$$\left\|x_i^t - x^\star\right\| \leq a_1\exp(-a_2 t), \quad \forall i \in \{1, \ldots, N\}.$$

Proof: Assume w.l.o.g. that $f(x^\star) = 0$. Being f strongly convex, the Polyak-Lojasiewicz inequality holds, namely

$$-\|\nabla f(x)\|^2 \leq -2\mu(f(x) - f(x^\star)) = -2\mu f(x), \tag{27}$$

for any $x \in \mathbb{R}^d$. By combining (27) and (25) (cf. proof of Lemma 2), we claim that, for any $\gamma \in (0, \bar{\gamma}_2)$, it holds

$$\underbrace{U(\bar{x}^{t+1}) - U(\bar{x}^t)}_{f(\bar{x}^{t+1}) - f(\bar{x}^t)} \leq \underbrace{-d_4 U(\bar{x}^t)}_{-2d_1\mu f(\bar{x}^t)}.$$

for some $d_4 > 0$ ensuring that $x^\star$ is globally exponentially stable for system (10). Further Lemma 1 provides a Lyapunov function for the global exponential stability of the origin for (15). Hence, by [30, Corolllary 2]), there exists $\bar{\gamma} > 0$ such that, for any $\gamma \in (0, \bar{\gamma})$, $\text{COL}(x^\star, h(x^\star))$ is globally exponentially stable for (7) and the proof follows. ■

V. Numerical Simulation: Target Localization

This section validates our theoretical findings with a numerical simulation about the target localization problem given in [10, Sec. IV.A]. A network of $N = 10$ agents aims to locate a common target through some distance measurements. Each agent i is located at $\omega_i \in \mathbb{R}^d$ and has a noisy measurement $\phi_i > 0$ of the target squared distance. The target position is estimated by solving an instance of problem (1) with $f_i(x) = \phi_i - \|x - \omega_i\|^2$. We consider an Erdős-Rényi directed graph with parameter 0.6. We empirically tuned the step-size parameter as $\gamma = 0.01$ to guarantee the algorithm effectiveness. Fig. 2 separately shows the convergence of the fast and slow subsystems identified in (7) and graphically highlights their different rates.

VI. Conclusion

We studied the Gradient Tracking for nonconvex distributed optimization with system theory tools. By resorting to suitable coordinates, we posed the scheme in a singularly perturbed form. The latter allowed us to elegantly prove the convergence of the estimates to the set of problem stationary points.

Appendix

A. Proof of Theorem 1

To show convergence properties of (9) we use the sum of the functions W and U satisfying (12) and (13), respectively. We point out that these two functions have been used on (10) and (11) that are different from the subsystems of (9). We start by defining $h_\gamma(\bar{\mathrm{x}}) := h(\bar{\mathrm{x}}, \gamma)$ and

$$L_2 := \sup_{\gamma \in [0, \bar{\gamma}_3]} \{L_g(\gamma)\}, \quad L_3 := \sup_{\gamma \in [0, \bar{\gamma}_3]} \{L_h(\gamma)\},$$

where $\bar{\gamma}_3 := \max\{\bar{\gamma}_1, \bar{\gamma}_2\}$ and both L_2 and L_3 are finite in light of the continuity of g and h, respectively. Thus, the global Lipschitz properties of g and h with parameters $L_g(\gamma)$ and $L_h(\gamma)$ lead to the Lipschitz property of g and h_γ with parameters L_2 and L_3 in the interval $[0, \bar{\gamma}_3]$. With this result at hand, define $\psi^t := \zeta^t - h_\gamma(\bar{\mathrm{x}}^t)$, and rewrite (9) as

$$\bar{\mathrm{x}}^{t+1} = \bar{\mathrm{x}}^t + \gamma\phi(\bar{\mathrm{x}}^t, \psi^t + h_\gamma(\bar{\mathrm{x}}^t)) \tag{28a}$$

$$\psi^{t+1} = g(\psi^t + h_\gamma(\bar{\mathrm{x}}^t), \bar{\mathrm{x}}^t, \gamma) - h_\gamma(\bar{\mathrm{x}}^{t+1}). \tag{28b}$$

Pick the function U satisfying (13). Thus, by evaluating $\Delta U(\bar{\mathrm{x}}^t) := U(\bar{\mathrm{x}}^{t+1}) - U(\bar{\mathrm{x}}^t)$ along (28a), we get

$$\begin{aligned}
\Delta U(\bar{\mathrm{x}}^t) &= U(\bar{\mathrm{x}}^t + \gamma\phi(\bar{\mathrm{x}}^t, \psi^t + h_\gamma(\bar{\mathrm{x}}^t))) - U(\bar{\mathrm{x}}^t)\\
&\overset{(a)}{=} U(\bar{\mathrm{x}}^t + \gamma\phi(\bar{\mathrm{x}}^t, h_\gamma(\bar{\mathrm{x}}^t))) - U(\bar{\mathrm{x}}^t)\\
&\quad + U(\bar{\mathrm{x}}^t + \gamma\phi(\bar{\mathrm{x}}^t, \psi^t + h_\gamma(\bar{\mathrm{x}}^t))) - U(\bar{\mathrm{x}}^t + \gamma\phi(\bar{\mathrm{x}}^t, h_\gamma(\bar{\mathrm{x}}^t)))\\
&\overset{(b)}{\le} -c_1\left\|\phi(\bar{\mathrm{x}}^t, h_\gamma(\bar{\mathrm{x}}^t))\right\|^2\\
&\quad + U(\bar{\mathrm{x}}^t + \gamma\phi(\bar{\mathrm{x}}^t, \psi^t + h_\gamma(\bar{\mathrm{x}}^t))) - U(\bar{\mathrm{x}}^t + \gamma\phi(\bar{\mathrm{x}}^t, h_\gamma(\bar{\mathrm{x}}^t)))\\
&\overset{(c)}{\le} -c_1\left\|\phi(\bar{\mathrm{x}}^t, h_\gamma(\bar{\mathrm{x}}^t))\right\|^2\\
&\quad + \gamma c_2\left\|\phi(\bar{\mathrm{x}}^t, h_\gamma(\bar{\mathrm{x}}^t))\right\|\left\|\phi(\bar{\mathrm{x}}^t, \psi^t + h_\gamma(\bar{\mathrm{x}}^t)) - \phi(\bar{\mathrm{x}}^t, h_\gamma(\bar{\mathrm{x}}^t))\right\|\\
&\quad + \gamma^2 c_3\Big(\left\|\phi(\bar{\mathrm{x}}^t, \psi^t + h_\gamma(\bar{\mathrm{x}}^t))\right\|^2 + \left\|\phi(\bar{\mathrm{x}}^t, h_\gamma(\bar{\mathrm{x}}^t))\right\|^2\Big),
\end{aligned} \tag{29}$$

where in (a) we add and subtract $U(\bar{\mathrm{x}}^t + \gamma\phi(\bar{\mathrm{x}}^t, h_\gamma(\bar{\mathrm{x}}^t)))$, in (b) we use (13a) to bound $U(\bar{\mathrm{x}}^t + \gamma\phi(\bar{\mathrm{x}}^t, h_\gamma(\bar{\mathrm{x}}^t))) - U(\bar{\mathrm{x}}^t)$, and in (c) we use (13b) to bound $U(\bar{\mathrm{x}}^t + \gamma\phi(\bar{\mathrm{x}}^t, \psi^t + h_\gamma(\bar{\mathrm{x}}^t))) - U(\bar{\mathrm{x}}^t + \gamma\phi(\bar{\mathrm{x}}^t, h_\gamma(\bar{\mathrm{x}}^t)))$. By adding and subtracting $\phi(\bar{\mathrm{x}}^t, h_\gamma(\bar{\mathrm{x}}^t))$ in $\left\|\phi(\bar{\mathrm{x}}^t, \psi^t + h_\gamma(\bar{\mathrm{x}}^t))\right\|^2$, (29) becomes

$$\begin{aligned}
\Delta U(\bar{\mathrm{x}}^t) &\le -c_1\left\|\phi(\bar{\mathrm{x}}^t, h_\gamma(\bar{\mathrm{x}}^t))\right\|^2\\
&\quad + \gamma c_2\left\|\phi(\bar{\mathrm{x}}^t, h_\gamma(\bar{\mathrm{x}}^t))\right\|\left\|\phi(\bar{\mathrm{x}}^t, \psi^t + h_\gamma(\bar{\mathrm{x}}^t)) - \phi(\bar{\mathrm{x}}^t, h_\gamma(\bar{\mathrm{x}}^t))\right\|\\
&\quad + \gamma^2 c_3\left\|\phi(\bar{\mathrm{x}}^t, \psi^t + h_\gamma(\bar{\mathrm{x}}^t)) - \phi(\bar{\mathrm{x}}^t, h_\gamma(\bar{\mathrm{x}}^t)) + \phi(\bar{\mathrm{x}}^t, h_\gamma(\bar{\mathrm{x}}^t))\right\|^2\\
&\quad + \gamma^2 c_3\left\|\phi(\bar{\mathrm{x}}^t, h_\gamma(\bar{\mathrm{x}}^t))\right\|^2\\
&\overset{(a)}{\le} -c_1\left\|\phi(\bar{\mathrm{x}}^t, h_\gamma(\bar{\mathrm{x}}^t))\right\|^2 + \gamma c_2 L_1\left\|\phi(\bar{\mathrm{x}}^t, h_\gamma(\bar{\mathrm{x}}^t))\right\|\left\|\psi^t\right\|\\
&\quad + \gamma^2 c_3 2\left\|\phi(\bar{\mathrm{x}}^t, h_\gamma(\bar{\mathrm{x}}^t))\right\|^2 + \gamma^2 c_3 L_1^2\left\|\psi^t\right\|^2\\
&\quad + \gamma^2 c_3 2L_1\left\|\psi^t\right\|\left\|\phi(\bar{\mathrm{x}}^t, h_\gamma(\bar{\mathrm{x}}^t))\right\|,
\end{aligned} \tag{30}$$

where (a) exploits the square norm and the fact that ϕ is Lipschitz. Now, pick W satisfying (12). By evaluating $\Delta W(\psi^t) := W(\psi^{t+1}) - W(\psi^t)$, we get

$$\begin{aligned}
\Delta W(\psi^t) &= W(g(\psi^t + h_\gamma(\bar{\mathrm{x}}^t), \bar{\mathrm{x}}^t, \gamma) - h_\gamma(\bar{\mathrm{x}}^{t+1})) - W(\psi^t)\\
&\overset{(a)}{=} W(g(\psi^t + h_\gamma(\bar{\mathrm{x}}^t), \bar{\mathrm{x}}^t, \gamma) - h_\gamma(\bar{\mathrm{x}}^t)) - W(\psi^t)\\
&\quad + W(g(\psi^t + h_\gamma(\bar{\mathrm{x}}^t), \bar{\mathrm{x}}^t, \gamma) - h_\gamma(\bar{\mathrm{x}}^{t+1}))\\
&\quad - W(g(\psi^t + h_\gamma(\bar{\mathrm{x}}^t), \bar{\mathrm{x}}^t, \gamma) - h_\gamma(\bar{\mathrm{x}}^t))\\
&\overset{(b)}{\le} -b_3\left\|\psi^t\right\|^2 + \tilde{W}(\psi^t, \bar{\mathrm{x}}^t),
\end{aligned} \tag{31}$$

where in (a) we add and subtract the term $W(g(\psi^t + h_\gamma(\bar{\mathrm{x}}^t), \bar{\mathrm{x}}^t, \gamma) - h_\gamma(\bar{\mathrm{x}}^t))$ and in (b) we bound the term $W(g(\psi^t + h_\gamma(\bar{\mathrm{x}}^t), \bar{\mathrm{x}}^t) - h_\gamma(\bar{\mathrm{x}}^t)) - W(\psi^t)$ by applying the result (12b) (which holds for any $\gamma \in (0, \bar{\gamma}_1)$) and introduce

$$\begin{aligned}
\tilde{W}(\psi^t, \bar{\mathrm{x}}^t) := W(g(\psi^t + h_\gamma(\bar{\mathrm{x}}^t), \bar{\mathrm{x}}^t, \gamma) - h_\gamma(\bar{\mathrm{x}}^{t+1}))\\
- W(g(\psi^t + h_\gamma(\bar{\mathrm{x}}^t), \bar{\mathrm{x}}^t, \gamma) - h_\gamma(\bar{\mathrm{x}}^t)).
\end{aligned}$$

By using (12c), we bound the above term as

$$\begin{aligned}
&\tilde{W}(\psi^t, \bar{\mathrm{x}}^t)\\
&\le b_4\left\|h_\gamma(\bar{\mathrm{x}}^{t+1}) - h_\gamma(\bar{\mathrm{x}}^t)\right\|\left\|g(\psi^t + h_\gamma(\bar{\mathrm{x}}^t), \bar{\mathrm{x}}^t, \gamma) - h_\gamma(\bar{\mathrm{x}}^{t+1})\right\|\\
&\quad + b_4\left\|h_\gamma(\bar{\mathrm{x}}^{t+1}) - h_\gamma(\bar{\mathrm{x}}^t)\right\|\left\|g(\psi^t + h_\gamma(\bar{\mathrm{x}}^t), \bar{\mathrm{x}}^t, \gamma) - h_\gamma(\bar{\mathrm{x}}^t)\right\|\\
&\overset{(a)}{\le} b_4\left\|h_\gamma(\bar{\mathrm{x}}^{t+1}) - h_\gamma(\bar{\mathrm{x}}^t)\right\|^2\\
&\quad + b_4 2\left\|h_\gamma(\bar{\mathrm{x}}^{t+1}) - h_\gamma(\bar{\mathrm{x}}^t)\right\|\left\|g(\psi^t + h_\gamma(\bar{\mathrm{x}}^t), \bar{\mathrm{x}}^t, \gamma) - h_\gamma(\bar{\mathrm{x}}^t)\right\|\\
&\overset{(b)}{\le} b_4\left\|h_\gamma(\bar{\mathrm{x}}^{t+1}) - h_\gamma(\bar{\mathrm{x}}^t)\right\|^2\\
&\quad + b_4 2\left\|h_\gamma(\bar{\mathrm{x}}^{t+1}) - h_\gamma(\bar{\mathrm{x}}^t)\right\|\left\|\Delta g(\psi^t, \bar{\mathrm{x}}^t, \gamma)\right\|\\
&\overset{(c)}{\le} \gamma^2 b_4 L_3^2\left\|\phi(\bar{\mathrm{x}}^t, \psi^t + h_\gamma(\bar{\mathrm{x}}^t))\right\|^2\\
&\quad + \gamma b_4 2L_3 L_2\left\|\phi(\bar{\mathrm{x}}^t, \psi^t + h_\gamma(\bar{\mathrm{x}}^t))\right\|\left\|\psi^t\right\|,
\end{aligned} \tag{32}$$

where in (a) we add and subtract within the second norm $h_\gamma(\bar{\mathrm{x}})$ and use the triangle inequality, in (b) we add within the last norm $g(h_\gamma(\bar{\mathrm{x}}^t), \bar{\mathrm{x}}^t, \gamma) - h(\bar{\mathrm{x}}^t) = 0$ and introduce $\Delta g(\psi^t, \bar{\mathrm{x}}^t, \gamma) := g(\psi^t + h_\gamma(\bar{\mathrm{x}}^t), \bar{\mathrm{x}}^t, \gamma) - g(h_\gamma(\bar{\mathrm{x}}^t), \bar{\mathrm{x}}^t, \gamma)$, and in (c) we exploit the Lipschitz continuity of h_γ and g. Now, add and subtract $\phi(\bar{\mathrm{x}}^t, h_\gamma(\bar{\mathrm{x}}^t))$ in $\left\|\phi(\bar{\mathrm{x}}^t, \psi^t + h_\gamma(\bar{\mathrm{x}}^t))\right\|^2$ and $\left\|\phi(\bar{\mathrm{x}}^t, \psi^t + h_\gamma(\bar{\mathrm{x}}^t))\right\|$, use the triangle inequality and the Lipschitz property of ϕ to bound (32) as

$$\begin{aligned}
\tilde{W}(\psi^t, \bar{\mathrm{x}}^t) &= \gamma^2 b_4 L_3^2 L_1^2\left\|\psi^t\right\|^2 + \gamma^2 b_4 L_3^2\left\|\phi(\bar{\mathrm{x}}^t, h_\gamma(\bar{\mathrm{x}}^t))\right\|^2\\
&\quad + \gamma^2 b_4 2L_3^2 L_1\left\|\psi^t\right\|\left\|\phi(\bar{\mathrm{x}}^t, h_\gamma(\bar{\mathrm{x}}^t))\right\| + \gamma b_4 2L_3 L_2 L_1\left\|\psi^t\right\|^2\\
&\quad + \gamma b_4 2L_3 L_2\left\|\phi(\bar{\mathrm{x}}^t, h_\gamma(\bar{\mathrm{x}}^t))\right\|\left\|\psi^t\right\|.
\end{aligned} \tag{33}$$

Thus, we can use (33) to bound (31) as

$$\begin{aligned}
\Delta W(\psi^t) &\le -b_3\left\|\psi^t\right\|^2 + \gamma^2 b_4 L_3^2 L_1^2\left\|\psi^t\right\|^2\\
&\quad + \gamma^2 b_4 L_3^2\left\|\phi(\bar{\mathrm{x}}^t, h_\gamma(\bar{\mathrm{x}}^t))\right\|^2\\
&\quad + \gamma^2 b_4 2L_3^2 L_1\left\|\psi^t\right\|\left\|\phi(\bar{\mathrm{x}}^t, h_\gamma(\bar{\mathrm{x}}^t))\right\| + \gamma b_4 2L_3 L_2 L_1\left\|\psi^t\right\|^2\\
&\quad + \gamma b_4 2L_3 L_2\left\|\phi(\bar{\mathrm{x}}^t, h_\gamma(\bar{\mathrm{x}}^t))\right\|\left\|\psi^t\right\|.
\end{aligned} \tag{34}$$

Now, define $V : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$ as

$$V(\bar{\mathrm{x}}^t, \psi^t) = U(\bar{\mathrm{x}}^t) + W(\psi^t).$$

Thus, by evaluating $\Delta V(\bar{\mathrm{x}}^t, \psi^t) := V(\bar{\mathrm{x}}^{t+1}, \psi^{t+1}) - V(\bar{\mathrm{x}}^t, \psi^t) = \Delta U(\bar{\mathrm{x}}^t) + \Delta W(\psi^t)$ along the trajectories of (28), we can use the results (30) and (34) to write

$$\Delta V(\bar{\mathrm{x}}^t, \psi^t) \le -\begin{bmatrix}\left\|\phi(\bar{\mathrm{x}}^t, h_\gamma(\bar{\mathrm{x}}^t))\right\|\\ \left\|\psi^t\right\|\end{bmatrix}^\top H \begin{bmatrix}\left\|\phi(\bar{\mathrm{x}}^t, h_\gamma(\bar{\mathrm{x}}^t))\right\|\\ \left\|\psi^t\right\|\end{bmatrix}, \tag{35}$$

where $H \in \mathbb{R}^{2\times 2}$ denotes the symmetric matrix

$$H := \begin{bmatrix} c_1 - \gamma^2 k_1 & -\gamma k_2 - \gamma^2 k_3\\ -\gamma k_2 - \gamma^2 k_3 & b_3 - \gamma k_4 - \gamma^2 k_5\end{bmatrix},$$

in which the notation has been shortened through the constants

$$k_1 := b_4 L_3^2 + c_3 2,\, k_2 := \frac{c_2 L_1 + b_4 2 L_2 L_3}{2},\, k_3 := \frac{c_3 2 L_1 + b_4 2 L_1 L_3^2}{2}$$
$$k_4 := b_4 2 L_1 L_2 L_3,\, k_5 := c_3 L_1^2 + b_4 L_1^2 L_3^2.$$

Being $H = H^\top$, by Sylvester Criterion, $H > 0$ if and only if

$$\begin{cases} c_1 > p_1(\gamma) \\ c_1 b_3 > p_2(\gamma), \end{cases} \tag{36}$$

where we have introduced the polynomials

$$p_1(\gamma) := \gamma^2 k_1 \tag{37a}$$
$$p_2(\gamma) := \gamma c_1(k_4 + \gamma k_5) + \gamma^2 k_1(b_3 - \gamma k_4 - \gamma^2 k_5) + (\gamma k_2 + \gamma^2 k_3)^2. \tag{37b}$$

We notice that $\lim_{\gamma\to 0} p_1(\gamma) = \lim_{\gamma\to 0} p_2(\gamma) = 0$. Thus, there exists $\bar{\gamma} \in (0, \min\{\bar{\gamma}_1, \bar{\gamma}_2\})$ such that, for any $\gamma \in (0, \bar{\gamma})$, the conditions in (36) hold leading to the positiveness of H. Hence (35) ensures that $\Delta V(\bar{\mathrm{x}}^t, \psi^t) \leq 0$ for any $\bar{\mathrm{x}}^t \in \mathbb{R}^n$ and any $\psi^t \in \mathbb{R}^m$. In particular, the right-hand side of (35) is null when $\bar{\mathrm{x}}^t \in E'$, where $E' \subseteq \mathbb{R}^{n+m}$ reads as

$$E' := \{(\bar{\mathrm{x}}, \psi) \in \mathbb{R}^{n+m} \mid \bar{\mathrm{x}} \in \ker\{\phi(\cdot, h_\gamma(\cdot))\}, \psi = 0\}. \tag{38}$$

Thus, we apply the Lasalle Invariance Principle (cf. [33, Th. 3.7]) to conclude that, for any $\gamma \in (0, \bar{\gamma})$, any trajectory of system (28) approaches

$$\lim_{t\to\infty} \inf_{\xi' \in \mathcal{M}'} \left\| \begin{bmatrix} \bar{\mathrm{x}}^t \\ \psi^t \end{bmatrix} - \xi' \right\| = 0, \tag{39}$$

where $\mathcal{M}' \subseteq E'$ denotes the largest invariant set for system (28) contained within the subspace E defined in (38). The proof follows by noticing that (i) $\mathcal{M}' \equiv E'$, and that (ii), turning out to the coordinates $(\bar{\mathrm{x}}, \zeta)$, the result (39) implies that, for any $\gamma \in (0, \bar{\gamma})$, any trajectory of (9) converges to $\mathcal{M} := \{(\bar{\mathrm{x}}, \zeta) \in \mathbb{R}^{n+m} \mid \bar{\mathrm{x}} \in \ker\{\phi(\cdot, h_\gamma(\cdot))\}, \zeta = h_\gamma(\bar{\mathrm{x}})\}$.

References

[1] A. Nedić and J. Liu, "Distributed optimization for control," *Annu. Rev. Control Robot. Auton. Syst.*, vol. 1, pp. 77–103, May 2018.

[2] G. Notarstefano, I. Notarnicola, and A. Camisa, "Distributed optimization for smart cyber-physical networks," *Found. Trends Syst. Control*, vol. 7, no. 3, pp. 253–383, 2019.

[3] T. Yang *et al.*, "A survey of distributed optimization," *Annu. Rev. Control*, vol. 47, pp. 278–305, Jun. 2019.

[4] B. Johansson, T. Keviczky, M. Johansson, and K. H. Johansson, "Subgradient methods and consensus algorithms for solving convex optimization problems," in *Proc. 47th IEEE Conf. Decis. Control*, 2008, pp. 4185–4190.

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

[6] A. Nedić, A. Ozdaglar, and P. A. Parrilo, "Constrained consensus and optimization in multi-agent networks," *IEEE Trans. Autom. Control*, vol. 55, no. 4, pp. 922–938, Apr. 2010.

[7] M. Zhu and S. Martínez, "Discrete-time dynamic average consensus," *Automatica*, vol. 46, no. 2, pp. 322–329, 2010.

[8] S. S. Kia, B. Van Scoy, J. Cortes, R. A. Freeman, K. M. Lynch, and S. Martinez, "Tutorial on dynamic average consensus: The problem, its applications, and the algorithms," *IEEE Control Syst. Mag.*, vol. 39, no. 3, pp. 40–72, Jun. 2019.

[9] 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, Apr. 2016.

[10] P. Di Lorenzo and G. Scutari, "NEXT: In-network nonconvex optimization," *IEEE Trans. Signal Inf. Process. Netw.*, vol. 2, no. 2, pp. 120–136, Jun. 2016.

[11] A. Nedić, A. Olshevsky, and W. Shi, "Achieving geometric convergence for distributed optimization over time-varying graphs," *SIAM J. Optim.*, vol. 27, no. 4, pp. 2597–2633, 2017.

[12] G. Qu and N. Li, "Harnessing smoothness to accelerate distributed optimization," *IEEE Trans. Control Netw. Syst.*, vol. 5, no. 3, pp. 1245–1260, Sep. 2018.

[13] J. Xu, S. Zhu, Y. C. Soh, and L. Xie, "Convergence of asynchronous distributed gradient methods over stochastic networks," *IEEE Trans. Autom. Control*, vol. 63, no. 2, pp. 434–448, Feb. 2018.

[14] C. Xi, R. Xin, and U. A. Khan, "ADD-OPT: Accelerated distributed directed optimization," *IEEE Trans. Autom. Control*, vol. 63, no. 5, pp. 1329–1339, May 2018.

[15] 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, Jul. 2018.

[16] G. Scutari and Y. Sun, "Distributed nonconvex constrained optimization over time-varying digraphs," *Math. Program.*, vol. 176, nos. 1–2, pp. 497–544, 2019.

[17] A. Daneshmand, G. Scutari, and V. Kungurtsev, "Second-order guarantees of distributed gradient algorithms," *SIAM J. Optim.*, vol. 30, no. 4, pp. 3029–3068, 2020.

[18] J. Wang and N. Elia, "Control approach to distributed optimization," in *Proc. 48th Annu. Allerton Conf. Commun. Control Comput. (Allerton)*, 2010, pp. 557–561.

[19] J. Wang and N. Elia, "A control perspective for centralized and distributed convex optimization," in *Proc. 50th IEEE Conf. Decis. Control Eur. Control Conf.*, 2011, pp. 3800–3805.

[20] T. Hatanaka, N. Chopra, T. Ishizaki, and N. Li, "Passivity-based distributed optimization with communication delays using PI consensus algorithm," *IEEE Trans. Autom. Control*, vol. 63, no. 12, pp. 4421–4428, Dec. 2018.

[21] C.-Y. Chang, M. Colombino, J. Cortés, and E. Dall'Anese, "Saddle-flow dynamics for distributed feedback-based optimization," *IEEE Control Syst. Lett.*, vol. 3, no. 4, pp. 948–953, Oct. 2019.

[22] A. Sundararajan, B. Van Scoy, and L. Lessard, "A canonical form for first-order distributed optimization algorithms," in *Proc. Amer. Control Conf. (ACC)*, 2019, pp. 4075–4080.

[23] M. Bin, I. Notarnicola, L. Marconi, and G. Notarstefano, "A system theoretical perspective to gradient-tracking algorithms for distributed quadratic optimization," in *Proc. IEEE Conf. Decis. Control (CDC)*, 2019, pp. 2994–2999.

[24] T. T. Doan, S. T. Maguluri, and J. Romberg, "Fast convergence rates of distributed subgradient methods with adaptive quantization," *IEEE Trans. Autom. Control*, vol. 66, no. 5, pp. 2191–2205, May 2021.

[25] Y. Tang, J. Zhang, and N. Li, "Distributed zero-order algorithms for nonconvex multiagent optimization," *IEEE Trans. Control Netw. Syst.*, vol. 8, no. 1, pp. 269–281, Mar. 2021.

[26] Y. Lu and M. Zhu, "Privacy preserving distributed optimization using homomorphic encryption," *Automatica*, vol. 96, pp. 314–325, Oct. 2018.

[27] A. Sundararajan, B. Van Scoy, and L. Lessard, "Analysis and design of first-order distributed optimization algorithms over time-varying graphs," *IEEE Trans. Control Netw. Syst.*, vol. 7, no. 4, pp. 1597–1608, Dec. 2020.

[28] G. Carnevale, I. Notarnicola, L. Marconi, and G. Notarstefano, "Triggered gradient tracking for asynchronous distributed optimization," 2022, *arXiv:2203.02210*.

[29] N. Mimmo, G. Carnevale, A. Testa, and G. Notarstefano, "Extremum seeking tracking for derivative-free distributed optimization," 2021, *arXiv:2110.04234*.

[30] R. Bouyekhf and A. El Moudni, "On analysis of discrete singularly perturbed non-linear systems: Application to the study of stability properties," *J. Franklin Inst.*, vol. 334, no. 2, pp. 199–212, 1997.

[31] H. K. Khalil, *Nonlinear Systems*. Upper Saddle River, NJ, USA: Prentice-Hall, 2002.

[32] D. P. Bertsekas and A. Scientific, *Convex Optimization Algorithms*. Belmont, CA, USA: Athena Sci., 2015.

[33] T. Ge, W. Lin, and J. Feng, "Invariance principles allowing of non-Lyapunov functions for estimating attractor of discrete dynamical systems," *IEEE Trans. Autom. Control*, vol. 57, no. 2, pp. 500–505, Feb. 2012.