# Distributed Optimization of Linear Multi-Agent Systems via Feedback-DGD

Amir Mehrnoosh and Gianluca Bianchin

***Abstract*— Feedback optimization is an increasingly popular control paradigm to optimize dynamical systems, accounting for control objectives that concern the system's operation at the steady-state. Existing feedback optimization techniques heavily rely on centralized system and controller architectures, and thus suffer from scalability and privacy issues when systems become large-scale. In this paper, we propose a distributed architecture for feedback optimization inspired by distributed gradient descent, whereby each agent updates its local control variable by combining the average of its neighbors with a local negative gradient step. Under convexity and smoothness assumptions for the cost, we establish convergence of the control method to a critical optimization point. By reinforcing the assumptions to restricted strong convexity, we show that our algorithm converges linearly to a neighborhood of the optimal point, where the size of the neighborhood depends on the choice of the stepsize. Simulations corroborate the theoretical results.**

***Index Terms* – Optimization algorithms, feedback optimization, distributed control, multi-agent systems.**

## I. Introduction

Optimal steady-state regulation is concerned with the problem of controlling a dynamical systems to an optimal steady-state point, as characterized by a mathematical optimization problem [1]. The classical approach to tackle this goal relies on the principle of separation between planning and control, whereby the optimization problem is solved beforehand (offline) to determine optimal system states, which are then inputed as references to controllers responsible for regulating the system to these states. Remarkably, a key assumption in this approach is that disturbances are known beforehand and fed to the optimization solver; this allows the optimization to be solved with high precision to generate the required reference states. Unfortunately, in most control applications, disturbances are unknown. Often, the main objective of a control system is to ensure optimality in the face of unmeasurable disturbances or imprecise system knowledge. Notably, classical batch optimization algorithms fail [2] when are approximately known or vary after the optimization has been solved because disturbances may perturb optimal steady states.

Recently, several authors have studied optimal steady-state regulation problems in a centralized setting. A list of representative works on this topic (necessarily incomplete) includes [3]–[9]. See also the recent developments using zeroth order algorithms and data-driven approaches [10]–[12]. Feedback optimization controllers have gained popularity thanks to their capability to regulate physical systems to optimal steady-state points while rejecting constant [4] or time-varying disturbances [7], [8]. The central idea consists of adapting numerical optimization algorithms to operate as feedback controllers. This is achieved by using an inexact gradient evaluated using real-time measurements to update control inputs without requiring the exact plant model and disturbances. This feature endows feedback optimization with the versatility to handle various scenarios. All of these methods are designed to be implemented in a centralized architecture, and thus suffer from scalability issues when systems become large-scale, as well as privacy concerns when cost functions or feedback signals need to be maintained private. This work departs from this existing literature by focusing on the problem of optimal steady-state regulation for systems with a distributed architecture. This connects our work with the body of literature on distributed optimization. Distributed gradient descent (DGD) was proposed in [13], studied in [14], a diminishing stepsize was used in [15] to ensure exact convergence; see also [16]–[19]. Other distributed optimization algorithms have been explored in recent years; we refer to [20] for a comprehensive discussion. Particularly related to our problem are the works [21]–[24]. Compare to [21], we do not require a two-layer control architecture and tracking controllers; in contrast to [22], we do not approximate the system's sensitivity matrix by its diagonal elements, ignoring the coupling between subsystems, which leads to a loss in accuracy; [23] focuses on systems modeled as a static linear map, while in this work we account for dynamics; finally, with respect to [24], we account for performance metrics that depend on a vector quantity as opposed to a scalar aggregate one.

This work features three main contributions. First, we propose a distributed architecture for the optimal steady-state regulation problem, and a distributed control algorithm to address this problem. Our algorithm is inspired by distributed optimization approaches and combines a gradient-descent step with a consensus operation to simultaneously solve an optimization and seek an agreement between the agents. Second, we present proof of convergence to a fixed point for the controller-system state. Our technical arguments provide guidelines on how to choose the (sufficiently small) controller stepsize to guarantee convergence of the controlled system. Third, we provide an explicit bound for the control error. Precisely, we show that under restricted strong convexity assumptions, the controller state converges linearly to a neighborhood of the optimal point. In line with the existing literature [14], the size of such a neighborhood depends on

The authors are with ICTEAM Institute and the Department of Mathematical Engineering at UCLouvain, Belgium. A. Mehrnoosh is supported by F.R.S.-FNRS. {amir.mehrnoosh, gianluca.bianchin}@uclouvain.be.

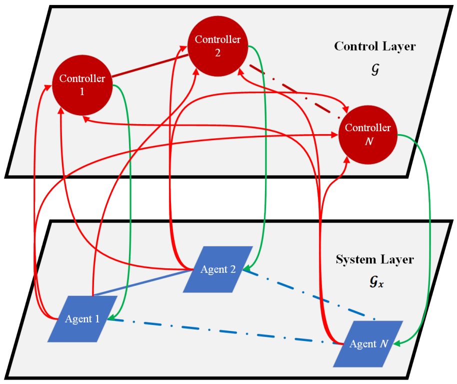



Fig. 1. Distributed system architecture considered in this work (cf. (2)). Each local controller actuates the corresponding subsystem (green lines), by using global feedback information (red lines), see (10).

the choice of the controller stepsize. Intuitively, convergence cannot be exact since distributed controllers need to average between moving in a direction that decreases the gradient while maintaining an agreement with the other controllers.

The rest of the article is organized as follows. Section II formalizes the problem of distributed optimal steady-state regulation. Section III illustrates the proposed controller. Section IV presents the two main contributions of this work, being the convergence analysis and error bound for the proposed controller. Section V validates the findings through numerical simulations, and Section VI concludes the paper.

**Notation.** For a symmetric matrix $W$, we denote its eigenvalues by $\lambda_1(W) \geq \lambda_2(W) \geq \cdots \geq \lambda_N(W)$. We assume that the mixing matrix $W$ is symmetric and doubly stochastic. The eigenvalues of $W$ are real and sorted in a nonincreasing order $1 = \lambda_1(W) \geq \lambda_2(W) \geq \cdots \geq \lambda_N(W) > -1$. Finally, we let

$$\beta := \max\{|\lambda_2(W)|, |\lambda_N(W)|\}. \tag{1}$$

## II. Problem Setting

### A. System to control

Consider a structured dynamical system composed of $N$ subsystems $\mathcal{V}_x = \{1, \ldots, N\}$; we describe the physical couplings between the subsystems using a directed graph (called system graph) $\mathcal{G}_x = (\mathcal{V}_x, \mathcal{E}_x)$, where $\mathcal{E}_x \subseteq \mathcal{V}_x \times \mathcal{V}_x$. See Fig. 1-(System Layer). Each subsystem $i \in \mathcal{V}$ is described by a local state $x_i^k \in \mathbb{R}^{n_i}$, updating as:

$$x_i^{k+1} = \sum_{j \in \mathcal{N}_i} A_{ij} x_j^k + B_i u_i^k + E_i w_i, \qquad k \in \mathbb{Z}_{\geq 0}, \tag{2}$$

where $\mathcal{N}_i$ denotes the set of subsystems collaborating with $i$, $u_i^k \in \mathbb{R}^{m_i}$ is the local control decision, and $A_{ij} \in \mathbb{R}^{n_i \times n_j}, B \in \mathbb{R}^{n_i \times m_i}$. In (2), $w_i \in \mathbb{R}^{r_i}$ with $E_i \in \mathbb{R}^{n_i \times r_i}$, models a constant unknown disturbance acting on subsystem $i$. We will denote by $n := \sum_i n_i$, $m := \sum_i m_i$, and $r := \sum_i r_i$. Further, we assume that the system's state is observed by means of the output signal:

$$y^k = \sum_{i=1}^{N} C_i x_i^k + D_i u_i^k, \tag{3}$$

where $C_i \in \mathbb{R}^{p \times n_i}$ and $D_i \in \mathbb{R}^{p \times m_i}$.

*Remark 1: **(Output model)*** The output model (3) implicitly requires that each subsystem is capable of measuring (or estimating) the *global* system state. Although alternative output models could be considered (e.g., were each subsystem $i$ measures $y_i^k = C_i x_i^k + D_i u_i^k$), the model (3) is common to many practical applications. For example, consider a swarm of drones where each drone monitors the surrounding environment through an areal camera view (which overlaps with that of its neighbors), or industrial systems, where a Supervisory Control and Data Acquisition (SCADA) system collects and processes all sensor measurements. □

In vector form, (2) reads as:

$$\begin{aligned} x^{k+1} &= Ax^k + Bu^k + Ew, \\ y^k &= Cx^k + Du^k, \end{aligned} \tag{4}$$

where $x^k = (x_1^k, \ldots, x_N^k) \in \mathbb{R}^n$ is the vector of states, $u^k = (u_1^k, \ldots, u_N^k) \in \mathbb{R}^m$ the vector of inputs, $w = (w_1, \ldots, w_N) \in \mathbb{R}^r$ the vector of disturbances, $A = [A_{ij}]$, $B = [B_{ij}]$, $E = [E_{ij}]$, $C = [C_1, \ldots, C_N]$, and $D = [D_1, \ldots, D_N]$ are block matrices. In what follows, we let

$$\begin{aligned} G(z) &= C(zI - A)^{-1}B + D, \\ H(z) &= C(zI - A)^{-1}E, \end{aligned} \tag{5}$$

denote the transfer functions from $u$ to $y$ and from $w$ to $y$, respectively. The definition (5) is intended for the values of $z \in \mathbb{C}$ for which the inverse is defined. We will adopt the compact notation $G := G(1), H := H(1)$ to express the steady-state response of the system to constant inputs; formally, when $u_k = \bar{u}, w_k = \bar{w}$ for all $k \in \mathbb{Z}_{\geq 0}$, we have:

$$\lim_{k \to \infty} y_k = G\bar{u} + H\bar{w}.$$

While we make no assumption on $\mathcal{G}_x$, we assume the following for (2).

*Assumption 1: **(Stability and control properties of system)*** The system (2) is asymptotically stable. Precisely, given $Q \succ 0$, we let $P \succ 0$ be such that $A^\mathsf{T} PA - P = -Q$. Moreover, (4) is controllable and observable. □

Controllability and observability are standard assumptions to guarantee that control problems are well-defined. Moreover, because our goal here is to tackle advanced control objectives aiming at optimizing the system operation (see (6) shortly below), we assume that the system has been prestabilized as in Assumption 1; the latter can be achieved using well-established static state feedback techniques [25].

### B. Distributed structure of the controller

We consider a setting where the system is controlled by a group of distributed controllers, each co-located with a local subsystem and actuating the corresponding local control variable (see Fig. 1). The combination of a local subsystem and the corresponding controller will hereafter be called an agent. To control the system, the controllers collaborate with each other; we describe their interaction topology by adopting an undirected graph (hereafter called control graph) $\mathcal{G} = (\mathcal{V}_u, \mathcal{E}_u)$ where $\mathcal{V}_u = \mathcal{V}_x$ and $\mathcal{E}_u \subseteq \mathcal{V}_u \times \mathcal{V}_u$. See Fig. 1-(Control Layer). A pair of controllers can collaboratively compute a control law if and only if they are connected by a link in $\mathcal{E}_u$. Recall that a graph is connected if there exists a path between any two nodes.

*Assumption 2:* ***(Connectivity of the control graph)*** The graph $\mathcal{G}$ is connected. Therefore, there exists a weight matrix $W = [w_{ij}] \in \mathbb{R}^{N\times N}$, associated with the communication graph which is a symmetric and doubly stochastic matrix with $\beta < 1$ (see (1)). $\square$

### C. Control objectives as an optimization problem

We study a control problem where the ensemble of controllers seeks to collaboratively compute an input that solves

$$\underset{u\in\mathbb{R}^m}{\text{minimize}} \quad \sum_{i=1}^{N} \Phi_i(u, Gu + Hw), \tag{6}$$

where $\Phi_i : \mathbb{R}^m \times \mathbb{R}^p \to \mathbb{R}, i \in \mathcal{V}_x$. The optimization problem (6) describes a setting where the group of controllers wants to determine a control input that optimizes (as measured by the cost $\sum_{i=1}^{N} \Phi_i(\cdot,\cdot)$) the system at steady-state (captured by the dependence of the cost on the steady-state output $Gu + Hw$). Moreover, the cost in (6) has a separable structure, describing that each $\Phi_i(\cdot,\cdot)$ is held locally by agent $i$. We also remark that optimization problem (6) is *parametrized* by $w$; as such, its solutions cannot be computed using standard optimization solvers, since the disturbance $w$ is unknown and unmeasurable.

*Remark 2:* ***(Practical relevance of*** (6)***)*** Allowing the local costs $\Phi_i(\cdot,\cdot)$ in (6) to depend on the global system input $u$ and on the global steady-state output $Gu + Hw$ allows our framework to describe a variety of problems where each agent has a personalized metric of system performance (namely, $\Phi_i(\cdot,\cdot)$), but altogether the group seeks to strike a balance between these heterogeneous objectives (by minimizing their sum, as in (6)). Examples of this setting are energy systems, where each subsystem may measures system optimality using a different performance measure.

Moreover, notice that a special case of (6) is:

$$\underset{u\in\mathbb{R}^m}{\text{minimize}} \quad \sum_{i=1}^{N} \tilde{\Phi}_i(u_i, Gu + Hw), \tag{7}$$

where $\tilde{\Phi}_i : \mathbb{R}^{m_i} \times \mathbb{R}^p \to \mathbb{R}$ now depends only on the local actuation variable $u_i$ (instead of the global one). We stress that our framework is general enough to account for this setting as a special case. This formulation describes, for example, problems where the ensemble of agents would like to optimize the global system operation (as described by $Gu + Hw$), while minimizing the local control effort. Returning to the swarm of drones example (see Remark 1), each drone may seek to reduce its local power consumption, while ensuring that the entire swarm reaches a desired configuration, which is measured by the global $y$. $\square$

We will denote in compact form

$$\Phi(u, y) := \sum_{i=1}^{N} \Phi_i(u, y).$$

*Assumption 3:* ***(Lipschitz and convexity of the cost)*** For all $i$, $(u, y) \mapsto \Phi_i(u, y)$ is proper closed convex, lower bounded, and Lipschitz differentiable with constant $L_{\Phi_i}$. $\square$

Assumption 3 is standard in optimization. This assumption allows us to derive the following inequality[1]:

$$\|\Pi^\mathsf{T}(\nabla\Phi(u, y) - \nabla\Phi(u', y'))\| \leq L_\Phi \left\| \begin{bmatrix} u \\ y \end{bmatrix} - \begin{bmatrix} u' \\ y' \end{bmatrix} \right\|, \tag{8}$$

where $\Pi^\mathsf{T} := \begin{bmatrix} I_m & G^\mathsf{T} \end{bmatrix}$, which holds for all $y, y' \in \mathbb{R}^n$, $u, u' \in \mathbb{R}^m$, and $L_\Phi := \|\Pi\| \sum_i L_{\Phi_i}$. In what follows, we denote the set of optimizers of (6) by

$$\mathcal{A}^* := \{(u^*, x^*) \ : \ (u^*, x^*) \text{ is a first-order optimizer of (6)}\},$$

and we assume that this set is nonempty and closed.

## III. Distributed Controller Design

A centralized algorithm to solve the steady-state regulation problem (6) has been studied in [12] and continuous-time counterparts [4], [7], [8]. In [12], the authors propose a gradient-type controller

$$u^{k+1} = u^k - \eta\Pi^\mathsf{T}\nabla\Phi(u^k, y^k), \tag{9}$$

where $\eta > 0$ denotes a scalar stepsize, being a design parameter. The controller (9) implements a gradient-type iteration to solve the optimization (6), modified by replacing the true gradient $\Pi^\mathsf{T}\nabla\Phi(u^k, Gu^k + Hw)$ with a measurement-based version $\Pi^\mathsf{T}\nabla\Phi(u^k, y^k)$, which avoids the need to measure $w$. Unfortunately, (9) is inapplicable to our setting since

(i) (9) does not respect the distributed nature of the controller considered here (cf. Section II-B), and
(ii) (9) requires centralized knowledge of the gradients $\{\nabla\Phi_i\}_{i\in\mathcal{V}_u}$, which is impractical in our case since each $\Phi_i$ is known only locally.

With this motivation, we next propose a distributed version of (9) that can be implemented in our control architecture (cf. Fig. 1). We propose an algorithm where each agent $i \in \mathcal{V}_u$ holds a local copy $u_{(i)}^k \in \mathbb{R}^m$ of $u^k$, and updates it as:

$$u_{(i)}^{k+1} = \sum_{j=1}^{N} w_{ij} u_{(j)}^k - \eta\Pi^\mathsf{T}\nabla\Phi_i(u_{(i)}^k, y^k). \tag{10}$$

In this control model, each agent $i$ updates its local state $u_{(i)}$ by performing two steps, (i) it computes a weighted average of its neighbors' states $\sum_{j=1}^{N} w_{ij} u_{(j)}^k$ to seek a consensus between the agents, and (ii) it applies $-\Pi^\mathsf{T}\nabla\Phi_i(u_{(i)}^k, y^k)$ to decrease $\Phi_i(u_{(i)}^k, y^k)$. We remark that this control law is distributed in the sense that each agent $i$ requires only knowledge of the local $\nabla\Phi_i$. Note that each agent needs to know the steady-state map (i.e., $G$) and measure the global output feedback signal $y^k$. In line with the literature on distributed optimization, we call (10) Feedback Distributed Gradient Descent (FDGD) algorithm.

[1]The notation $\nabla\Phi(u, y)$ indicates $\nabla\Phi(u, y) = (\nabla_u\Phi(u, y), \nabla_y\Phi(u, y)) \in \mathbb{R}^{m+p}$.

*Remark 3:* ***(Relationship with distributed optimization algorithm)*** The algorithm (10) is inspired from the DGD algorithm for distributed optimization [14]. Although other algorithms could be considered (e.g., EXTRA, gradient tracking [20], etc.), we leave an investigation of these approaches as the scope of future works. Finally, we also observe that DGD is the preferable method in certain circumstances, such as under drifts in the network topology [26]. □

## IV. Convergence Analysis and Error Bounds

In this section, we study the convergence of (10) when applied to control the system (4). In the remainder, we employ the following notations of stacked vectors: $u^k_{(1:N)} := (u^k_{(1)}, u^k_{(2)}, \dots, u^k_{(N)}) \in \mathbb{R}^{mN}$ and

$$\gamma(u^k_{(1:N)}, y^k) := \begin{bmatrix} \Pi^\top \nabla\Phi_1(u^k_{(1)}, y^k) \\ \vdots \\ \Pi^\top \nabla\Phi_N(u^k_{(N)}, y^k) \end{bmatrix} \in \mathbb{R}^{mN}.$$

In vector form, the system (4) controlled by (10) reads as:

$$x^{k+1} = Ax^k + BSu^k_{(1:N)} + Ew, \tag{11a}$$

$$y^k = Cx^k + DSu^k_{(1:N)},$$

$$u^{k+1}_{(i)} = \sum_{j\in\mathcal{N}_i} w_{ij} u^k_{(j)} - \eta\Pi^\top \nabla\Phi_i(u^k_{(i)}, y^k), \quad i \in \mathcal{V}_u, \tag{11b}$$

where $S \in \mathbb{R}^{m\times mN}$ is given by:

$$S = \mathrm{diag}([I_{m_1}, 0, \dots], [0, I_{m_2}, 0, \dots], \dots [0, \dots, 0, I_{m_N}]).$$

### A. Asymptotic convergence

The following result shows that, under a suitable choice of the stepsize $\eta$, the state of (11) converges asymptotically.

*Theorem 4.1:* ***(Convergence of the state sequences)*** Let Assumptions 1-3 hold, $W$ be such that $\beta < 1$, and the stepsize $\eta \le \bar\eta := \min\{\eta_1, \eta_2, \eta_3\}$, where

$$\eta_1 = \frac{1 - 2\mu + \lambda_N(W)}{L_\Phi}, \qquad \eta_2 = \frac{\mu}{\lambda_1(P) L_h^2}, \tag{12}$$

$$\eta_3 = \frac{\mu\lambda_n(Q)}{\frac{L_\Phi^2}{4} + L_h^2(\|A^\top P\|^2 + \lambda_n(Q)\lambda_1(P)) + L_h L_\Phi \|A^\top P\|},$$

with $\mu$ an arbitrary constant, $0 < \mu \le 1 - \frac{(1-\lambda_N(W)) + \eta L_\Phi}{2}$, and $L_h = \|(I - A)^{-1} BS\|$. Then, the sequences $x^k$, $u^k_{(i)}$ generated by (11) converges. □

*Proof:* We will prove this claim by using La Salle's Invariance Principle [25, Thm 4.4]. For clarity of presentation, the proof is organized into four main steps.

*1) Change of variables and storage function.* Let $h(u) = (I - A)^{-1} BSu + (I - A)^{-1} Ew$, and consider the new coordinate $\tilde{x}^k = x^k - h(u^k_{(1:N)})$, which shifts the equilibrium point of (11a) to the origin. Inspired by singular-perturbation reasoning [25, Sec. 11], consider the storage function:

$$U(u_{(1:N)}, \tilde{x}) := \frac{d}{\eta} V_u(u_{(1:N)}) + (1 - d)V_x(\tilde{x}), \tag{13}$$

for each $\tilde{x} \in \mathbb{R}^n, u_{(1:N)} \in \mathbb{R}^{mN}$, where $d \in (0, 1)$ and

$$V_u(u_{(1:N)}) = -\frac{1}{2}\sum_{i,j=1}^N w_{ij} u_{(i)}^\top u_{(j)} + \sum_{i=1}^N \left(\frac{1}{2}\|u_{(i)}\|^2 + \eta\Phi_i(u_{(i)}, Gu_{(i)} + Hw)\right),$$

$$V_x(\tilde{x}) = \tilde{x}^\top P\tilde{x}. \tag{14}$$

Notice that $V_u$ is Lipschitz differentiable with constant $L_{V_u} \le (1 - \lambda_N(W)) + \eta L_\Phi$ and it is convex (since all $\Phi_i$ are convex and $\sum_{i=1}^N \|u_{(i)}\|^2 - \sum_{i,j=1}^N w_{ij} u_{(i)}^\top u_{(j)}$ is also convex due to $\lambda_1(W) = 1$). Next, we introduce the quantity

$$\tilde{V}_u(u_{(1:N)}, \tilde{x}) = -\frac{1}{2}\sum_{i,j=1}^N w_{ij} u_{(i)}^\top u_{(j)} + \sum_{i=1}^N \left(\frac{1}{2}\|u_{(i)}\|^2 + \eta\Phi_i(u_{(i)}, C\tilde{x} + GSu_{(i)} + Hw)\right), \tag{15}$$

and

$$F_c(u_{(1:N)}, \tilde{x}) := \begin{bmatrix} \nabla_{u_{(1)}} \tilde{V}_u(u_{(1:N)}, \tilde{x}) \\ \vdots \\ \nabla_{u_{(N)}} \tilde{V}_u(u_{(1:N)}, \tilde{x}) \end{bmatrix}.$$

With this notation, (11b) and (15) can be re-expressed as:

$$u^{k+1}_{(1:N)} = u^k_{(1:N)} - F_c(u^k_{(1:N)}, \tilde{x}^k),$$

$$V_u(u_{(1:N)}) = \tilde{V}_u(u_{(1:N)}, 0), \tag{16}$$

by using $y^k = C(\tilde{x}^k + h(u^k_{(1:N)}) + DSu^k_{(1:N)}$.

*2) Bounding the variation of $V_u(\cdot)$.* We have:

$$\begin{aligned} V_u(u^{k+1}_{(1:N)}) &\le V_u(u^k_{(1:N)}) + \nabla V_u(u^k_{(1:N)})^\top (u^{k+1}_{(1:N)} - u^k_{(1:N)}) + \frac{L_{V_u}}{2}\|u^{k+1}_{(1:N)} - u^k_{(1:N)}\|^2 \\ &= V_u(u^k_{(1:N)}) - \nabla V_u(u^k_{(1:N)})^\top F_c(u^k_{(1:N)}, \tilde{x}^k) + \frac{L_{V_u}}{2}\|F_c(u^k_{(1:N)}, \tilde{x}^k)\|^2 \\ &\le V_u(u^k_{(1:N)}) - \|F_c(u^k_{(1:N)}, \tilde{x}^k)\|^2 + \|F_c(u^k_{(1:N)}, \tilde{x}^k)\|\|\nabla V_u(u^k_{(1:N)})^\top - F_c(u^k_{(1:N)}, \tilde{x}^k)\| + \frac{L_{V_u}}{2}\|F_c(u^k_{(1:N)}, \tilde{x}^k)\|^2 \\ &\le V_u(u^k_{(1:N)}) - \left(1 - \frac{L_{V_u}}{2}\right)\|F_c(u^k_{(1:N)}, \tilde{x}^k)\|^2 + \|F_c(u^k_{(1:N)}, \tilde{x}^k)\|\|\nabla V_u(u^k_{(1:N)})^\top - F_c(u^k_{(1:N)}, x^k)\| \\ &\le V_u(u^k_{(1:N)}) - \mu\|F_c(u^k_{(1:N)}, x^k)\|^2 + \eta L_\Phi \|F_c(u^k_{(1:N)}, x^k)\|\|\tilde{x}^k\| \end{aligned} \tag{17}$$

where the first row follows by convexity and Lipschitz differentiability of $V_u$, the second row follows by application of (16), the third row follows by adding and subtracting $F_c(u^k_{(1:N)}, \tilde{x}^k)$ to $\nabla V_u(u^k_{(1:N)})$ and using the triangle inequality, and the last row follows by Lipschitz differentiability of $\Phi$ and definition of $\mu$.

*3) Bounding the variation of $V_x(\cdot)$.* In the variables $\tilde{x}^k$, the plant dynamics (11a) read as:

$$\tilde{x}^{k+1} = A\tilde{x}^k + h(u^k_{(1:N)}) - h(u^{k+1}_{(1:N)}).$$

We have:

$$\begin{aligned}
V_x(\tilde{x}^{k+1}) &= (\tilde{x}^k)^\top A^\top P A \tilde{x}^k \\
&\quad + 2(\tilde{x}^k)^\top A^\top P(h(u^k_{(1:N)}) - h(u^{k+1}_{(1:N)})) \\
&\quad + (h(u^k_{(1:N)}) - h(u^{k+1}_{(1:N)}))^\top P (h(u^k_{(1:N)}) - h(u^{k+1}_{(1:N)})) \\
&\le V_x(\tilde{x}^k) - (\tilde{x}^k)^\top Q \tilde{x}^k + \lambda_1(P)\|h(u^k_{(1:N)}) - h(u^{k+1}_{(1:N)})\|^2 \\
&\quad + 2\|A^\top P\|\|\tilde{x}^k\|\|h(u^k_{(1:N)}) - h(u^{k+1}_{(1:N)})\| \\
&\le V_x(\tilde{x}^k) - \lambda_n(Q)\|\tilde{x}^k\|^2 + \lambda_1(P)L_h^2\|u^k_{(1:N)} - u^{k+1}_{(1:N)}\|^2 \\
&\quad + 2L_h\|A^\top P\|\|\tilde{x}^k\|\|u^k_{(1:N)} - u^{k+1}_{(1:N)}\| \\
&= V_x(\tilde{x}^k) - \lambda_n(Q)\|\tilde{x}^k\|^2 + \lambda_1(P)L_h^2\|F_c(u^k_{(1:N)}, x^k)\|^2 \\
&\quad + 2L_h\|A^\top P\|\|\tilde{x}^k\|\|F_c(u^k_{(1:N)}, x^k)\|, \qquad (18)
\end{aligned}$$

where the second row follows from Assumption 1, the third row from Lipschitz continuity of $h$, and the last row by application of (16).

*4) Combining the bounds.* By combining (17) and (18), we conclude $U(u^{k+1}_{(1:N)}, \tilde{x}^{k+1}) - U(u^k_{(1:N)}, \tilde{x}^k) \le -\xi(u^k_{(1:N)}, \tilde{x}^k)^\top \Lambda \xi(u^k_{(1:N)}, \tilde{x}^k)$, where

$$\xi(u^k_{(1:N)}, \tilde{x}^k) = \begin{bmatrix} \|F_c(u^k_{(1:N)}, \tilde{x}^k)\| \\ \|\tilde{x}^k\| \end{bmatrix}$$

and

$$\Lambda = \begin{bmatrix} d\frac{\mu}{\eta} - (1-d)\lambda_1(P)L_h^2 & \alpha \\ \alpha & (1-d)\lambda_n(Q) \end{bmatrix},$$

where $\alpha := -\frac{1}{2}(dL_\Phi + 2(1-d)L_h\|A^\top P\|)$. It is guaranteed that $\Lambda \succ 0$ when $\eta \in (0, \bar{\eta}), \bar{\eta} = \min\{\bar{\eta}_1, \bar{\eta}_2\}$, where

$$\begin{aligned}
\bar{\eta}_2 &= \frac{d\mu}{(1-d)\lambda_1(P)L_h^2}, \\
\bar{\eta}_3 &= \frac{d(1-d)\mu\lambda_n(Q)}{\frac{d^2L_\Phi^2}{4} + (1-d)^2L_h^2\rho + d(1-d)L_hL_\Phi\|A^\top P\|},
\end{aligned}$$

with $\rho := \|A^\top P\|^2 + \lambda_n(Q)\lambda_1(P)$. By making the following choice for the free variable $d$ :

$$d = \frac{L_h^2(\|A^\top P\|^2 + \lambda_n(Q)\lambda_1(P)) - \frac{L_\Phi L_h}{2}\sqrt{\rho}}{L_h^2(\|A^\top P\|^2 + \lambda_n(Q)\lambda_1(P)) - \frac{L_\Phi^2}{4}},$$

we obtain the largest value of $\bar{\eta}_3$ that guarantees decrease of the storage function. For the sake of simplicity, we choose $d = 0.5$, yielding the choice $\eta_2$ and $\eta_3$ in (12). To conclude, since $U(u^k_{(1:N)}, \tilde{x}^k)$ strongly decreases along the trajectories of (11) and its minimum is the point $(\|F_c(u^k_{(1:N)}, \tilde{x}^k)\|, \|\tilde{x}^k\|) = (0, 0)$, by La Salle's invariance principle the claim follows. ∎

Theorem 4.1 shows that under a sufficiently small choice of the stepsize $\eta$, the sequences $x^k, u^k_{(i)}$ of the system and controller states, converge asymptotically to a fixed point, respectively. Notice that this convergence claim is not straightforward, since the proposed controller (10) incorporates two simultaneous steps, a consensus one and a gradient step. As such, this update may oscillate or fail to converge when $\eta$ is chosen inadequately. The upper bounds $\eta_1, \eta_2, \eta_3$ depend on the various parameters of the system (4) and optimization problem (6). Importantly, notice that the imposed bounds on $\mu$ guarantee that $\bar{\eta} > 0$ is well-defined. Finally, we notice that a sufficiently small choice for $\eta$ also guarantees that there a choice of $\mu$ that satisfies $0 < \mu \le 1 - \frac{(1-\lambda_N(W)) + \eta L_\Phi}{2}$ (e.g., $\eta L_\Phi < 1 + \lambda_N(W)$).

Before proceeding, we present an instrumental result that will be used in the remainder. Based on the definition of $u^k_{(1:N)}$ and $\gamma(u^k_{(1:N)}, y^k)$, we rewrite (10) as:

$$u^{k+1}_{(1:N)} = (W \otimes I)u^k_{(1:N)} - \eta\,\gamma(u^k_{(1:N)}, y^k), \qquad (19)$$

where $\otimes$ denotes the Kronecker product.

*Proposition 4.2:* ***(Bounded gradient)*** Let the assumptions of Theorem 4.1 hold. Moreover, assume that the initial conditions of the controller satisfy $u^0_{(i)} = 0,\ \forall i \in \mathcal{V}$. Then, for all $k \in \mathbb{Z}_{\ge 0}$, (19) satisfies:

$$\|\gamma(u^k_{(1:N)}, y^k)\| \le \sigma, \qquad (20)$$

where $\sigma := \sqrt{2L_\Phi\left(\sum_{i=1}^N \Phi_i(u^0_{(i)}, y^0) - \Phi^{\text{opt}}\right)}$ Here, $\Phi^{\text{opt}} = \sum_{i=1}^N \Phi_i^{\text{opt}}$, with $\Phi_i^{\text{opt}} = \Phi_i(u^{\text{opt}}_{(i)}, y^{\text{opt}})$ and $(u^{\text{opt}}_{(i)}, y^{\text{opt}}) = \arg\min_{u,y} \Phi_i(u, y)$.

*Proof:* To prove (20), we rewrite

$$\begin{aligned}
\sum_{i=1}^N \Phi_i(u^k_{(i)}, y^k) &\le \eta^{-1}\tilde{V}(u^k_{(1:N)}, \tilde{x}^k) \qquad (21) \\
&\le \eta^{-1}\tilde{V}(u^0_{(1:N)}, \tilde{x}^0) \quad = \sum_{i=1}^N \Phi_i(u^0_{(i)}, y^0)
\end{aligned}$$

where the first inequality follows from (15) and by using $\sum_{i=1}^N \|u_{(i)}\|^2 - \sum_{i,j=1}^N w_{ij}u_{(i)}^\top u_{(j)} \ge 0$ (which holds since $\beta < 1$), the second inequality holds from $\tilde{V}_u(u^{k+1}_{(1:N)}, \tilde{x}^{k+1}) \le \tilde{V}_u(u^k_{(1:N)}, \tilde{x}^k)$ (which follows by iterating the steps in (17) applied to $\tilde{V}_u(u^{k+1}_{(1:N)}, \tilde{x}^{k+1})$ instead of $V_u(u^{k+1}_{(1:N)})$), and the last identity follows from (14) using $u^0_{(1:N)} = 0$.

Moreover, recall that for any differentiable convex function $g$ with minimizer $u^*, y^*$, and Lipschitz constant $L_g$, we have $g(u_a, y_a) \ge g(u_b, y_b) + \nabla g^\top(u_b, y_b)\begin{bmatrix} u_a - u_b & y_a - y_b \end{bmatrix}^\top + \frac{1}{2L_g}\|\nabla g(u_a, y_a) - \nabla g(u_b, y_b)\|^2$ and $\nabla g(u^*, y^*) = 0$. Then, $\|\nabla g(u, y)\|^2 \le 2L_g(g(u, y) - g^*)$, where $g^* := g(u^*, y^*)$. Using this inequality and (21), we obtain

$$\begin{aligned}
\|\gamma(u^k_{(1:N)}, y^k)\|^2 &= \sum_{i=1}^N \|\Pi^\top \nabla\Phi_i(u^k_{(i)}, y^k)\|^2 \qquad (22) \\
&\le \sum_{i=1}^N 2L_\Phi(\Phi_i(u^k_{(i)}, y^k) - \Phi_i^o) \\
&\le 2L_\Phi\left(\sum_{i=1}^N \Phi_i(u^0_{(i)}, y^0) - \Phi^{\text{opt}}\right),
\end{aligned}$$

where $\Phi^{\text{opt}} = \sum_{i=1}^{N} \Phi_i^{\text{opt}}$, with $\Phi_i^{\text{opt}} = \Phi_i(u_{(i)}^{\text{opt}}, y^{\text{opt}})$ and $(u_{(i)}^{\text{opt}}, y^{\text{opt}}) = \arg\min_{u,y} \Phi_i(u, y)$. Note that $u_{(i)}^{\text{opt}}, y^{\text{opt}}$ exist because of Assumption 3 and 2. ■

Proposition 4.2 ensures that the sequence of gradients $\gamma(u_{(1:N)}^k, y^k)$ is uniformly bounded. This result will be key in the subsequent section when characterizing the control error. Interestingly, unlike [16]–[18] that assume bounded gradient, in our analysis our choice of stepsize ensures that gradient remains bounded. We conclude by noting that when the initial conditions $u_{(1:N)}^0$ are nonzero, a uniform bound of the form (20) can still be proven by adjusting the step (22), $\sigma$ needs to be modified to account for additional error terms.

### B. Control error bounds

While Theorem 4.1 certifies that the states sequences converge asymptotically, it remains to quantify explicitly the controller performance. This is the focus of this section.

In line with the existing literature [14], to establish a liner rate of convergence, we will restrict our focus on cost functions that are restricted strongly convex; recall that $f : \operatorname{dom} f \to \mathbb{R}$ is restricted strongly convex [27] with modulus $\nu_f$ if

$$(\nabla f(z) - \nabla f(z^*))^\mathsf{T}(z - z^*) \geq \nu_f \|z - z^*\|^2, \tag{23}$$

for all $z \in \operatorname{dom} f, z^* = \operatorname{Proj}_{\mathcal{Z}^*}(z)$, where $\operatorname{Proj}_{\mathcal{Z}^*}(z)$ is the projection of $z$ onto the solution set $\mathcal{Z}^*$ such that $\nabla f(z^*) = 0$. The following result is instrumental.

*Lemma 4.3:* [27, Lemma 6] Suppose that $f$ is restricted strongly convex with modulus $\nu_f$ and $\nabla f$ is Lipschitz continuous with constant $L_f$. Then, we have

$$\begin{aligned}(z - z^*)^\mathsf{T}&(\nabla f(z) - \nabla f(z^*)) \\ &\geq c_1 \|\nabla f(z) - \nabla f(z^*)\|^2 + c_2\|z - z^*\|^2,\end{aligned} \tag{24}$$

where $z^*$ is as in (23). Moreover, for any $\theta \in [0, 1]$,

$$c_1 = \frac{\theta}{L_f}, \qquad c_2 = (1 - \theta)\nu_f. \tag{25}$$

*Remark 4:* Notice that, if $f$ is strongly convex with modulus $\nu_f$, then it is also restricted strong convexity with the same modulus [27]. In this case, (24) holds with

$$c_1 = \frac{1}{\nu_f + L_f}, \qquad c_2 = \frac{\nu_f L_f}{\nu_f + L_f}. \tag{26}$$

□

The following is the second main result of this paper.

*Theorem 4.4:* ***(Control error bounds)*** Let the assumptions of Proposition (4.2) hold. If $(u, y) \mapsto \Phi(u, y)$ is restricted strongly convex with modulus $\nu_\Phi$, then, for (11) it holds:

$$\|u_{(i)}^k - u^{*k}\| \leq c_3^k \|u_{(i)}^0 - u^{*0}\| + \frac{c_4}{\sqrt{1 - c_3^2}} + \frac{\eta\sigma}{1 - \beta}, \tag{27}$$

where

$$c_3^2 = 1 - \eta c_2 + \eta\delta - \eta^2\delta c_2, \quad c_4^2 = \eta^3(\eta + \delta^{-1})\frac{L_\Phi^2 \sigma^2}{(1 - \beta)^2},$$

$(u^{*k}, x^{*k}) := \operatorname{Proj}_{\mathcal{A}^*}(\bar{u}^k, x^k)$, $\sigma$ is as in (20), $\delta > 0$ is an arbitrary constant, and $c_1$ and $c_2$ are as in (25) with $\nu_f = \nu_\Phi/N$. Moreover, if $\Phi(u, y)$ is strongly convex, then $c_1$ and $c_2$ are as in (26). □

*Proof:* We begin by proving (27). It will be convenient to measure the control error relative to the average controller state: $\bar{u}^k := \frac{1}{N}\sum_{i=1}^{N} u_{(i)}^k$. We have

$$\|u_{(i)}^k - u^{*k}\| \leq \|u_{(i)}^k - \bar{u}^k\| + \|\bar{u}^k - u^{*k}\|. \tag{28}$$

For presentation purposes, the proof is organized into two main steps.

*1) Bound for* $\|u_{(i)}^k - \bar{u}^k\|$. By expanding (19) in time:

$$u_{(1:N)}^k = -\eta \sum_{s=0}^{k-1} (W^{k-1-s} \otimes I)\gamma(u_{(1:N)}^s, y^s). \tag{29}$$

Next, let $\bar{u}_{(1:N)}^k = (\bar{u}^k, \cdots, \bar{u}^k) \in \mathbb{R}^{mN}$, and notice that $\bar{u}_{(1:N)}^k = \frac{1}{N}((1_N 1_N^\mathsf{T}) \otimes I)u_{(1:N)}^k$. As a result,

$$\begin{aligned}
\|u_{(i)}^k - \bar{u}^k\| &\leq \|u_{(1:N)}^k - \bar{u}_{(1:N)}^k\| \\
&= \|u_{(1:N)}^k - \frac{1}{N}((1_N 1_N^\mathsf{T}) \otimes I)u_{(1:N)}^k\| \\
&= \| - \eta \sum_{s=0}^{k-1}(W^{k-1-s} \otimes I)\gamma(u_{(1:N)}^s, y^s) \\
&+ \eta \sum_{s=0}^{k-1} \frac{1}{N}((1_N 1_N^\mathsf{T} W^{k-1-s}) \otimes I)\gamma(u_{(1:N)}^s, y^s)\| \\
&= \| - \eta \sum_{s=0}^{k-1}(W^{k-1-s} \otimes I)\gamma(u_{(1:N)}^s, y^s) \\
&+ \eta \sum_{s=0}^{k-1} \frac{1}{N}((1_N 1_N^\mathsf{T}) \otimes I)\gamma(u_{(1:N)}^s, y^s)\| \\
&= \eta\| \sum_{s=0}^{k-1} \left( \left( W^{k-1-s} - \frac{1}{N} 1_N 1_N^\mathsf{T} \right) \otimes I \right) \gamma(u_{(1:N)}^s, y^s)\| \\
&\leq \eta \sum_{s=0}^{k-1} \|W^{k-1-s} - \frac{1}{N} 1_N 1_N^\mathsf{T}\| \|\gamma(u_{(1:N)}^s, y^s)\| \\
&= \eta \sum_{s=0}^{k-1} \beta^{k-1-s} \|\gamma(u_{(1:N)}^s, y^s)\|,
\end{aligned} \tag{30}$$

where the fourth row holds because $W$ is doubly stochastic. From $\|\gamma(u_{(1:N)}^k, y^k)\| \leq \sigma$ and $\beta < 1$, it follows that

$$\begin{aligned}
\|u_{(i)}^k - \bar{u}^k\| &\leq \eta \sum_{s=0}^{k-1} \beta^{k-1-s} \|\gamma(u_{(1:N)}^s, y^s)\| \leq \sum_{s=0}^{k-1} \beta^{k-1-s}\sigma \\
&\leq \frac{\eta\sigma}{1 - \beta}.
\end{aligned} \tag{31}$$

*2) Bound for* $\|\bar{u}^k - u^{*k}\|$. We will denote in compact form:

$$\bar{e}^k := \bar{u}^k - u^{*k}.$$

To bound this term, let

$$g(u_{(1:N)}^k, y^k) = \frac{1}{N} \sum_{i=1}^{N} \Pi^\mathsf{T} \nabla \Phi_i(u_{(i)}^k, y^k),$$

$$\bar{g}(u_{(1:N)}^k, y^k) = \frac{1}{N} \sum_{i=1}^{N} \Pi^\mathsf{T} \nabla \Phi_i(\bar{u}^k, y^k).$$

We are interested in $g(u^k_{(1:N)}, y^k)$ because $-\eta g(u^k_{(1:N)}, y^k)$ updates $\bar{u}^k$. To see this, by taking the average of (10) over $i$ and noticing $W = [w_{ij}]$ is doubly stochastic, we obtain

$$\begin{aligned}\bar{u}^{k+1} &= \frac{1}{N}\sum_{i=1}^{N} u^{k+1}_{(i)} \\ &= \frac{1}{N}\sum_{i,j=1}^{N} w_{ij}u^k_{(j)} - \frac{\eta}{N}\sum_{i=1}^{N}\Pi^\mathsf{T}\nabla\Phi_i(u^k_{(i)}, y^k) \\ &= \bar{u}^k - \eta g(u^k_{(1:N)}, y^k). \end{aligned} \tag{32}$$

Before proceeding notice that the following bound holds:

$$\begin{aligned}\|\Pi^\mathsf{T}(\nabla\Phi_i(u^k_{(i)}, y^k) - \nabla\Phi_i(\bar{u}^k, y^k))\| &\leq L_\Phi\|u^k_{(i)} - \bar{u}^k\| \\ &\leq \frac{\eta\sigma L_\Phi}{1-\beta}\end{aligned}$$

by Assumptions 3, 2, and (8), and where the last inequality follows from (31). Moreover, we also have:

$$\begin{aligned}&\|g(u^k_{(1:N)}, y^k) - \bar{g}(u^k_{(1:N)}, y^k)\| \\ &= \|\frac{1}{N}\sum_{i=1}^{N}\Pi^\mathsf{T}(\nabla\Phi_i(u^k_{(i)}, y^k) - \nabla\Phi_i(\bar{u}^k, y^k))\| \\ &\leq \frac{1}{N}L_\Phi\sum_{i=1}^{N}\|u^k_{(i)} - \bar{u}^k\| \\ &\leq \frac{\eta\sigma L_\Phi}{1-\beta}. \end{aligned} \tag{33}$$

Recalling that $(u^{*k+1}, x^{*k+1}) = \mathrm{Proj}_{\mathcal{A}^*}(\bar{u}^{k+1}, x^{k+1})$ and $\bar{e}^{k+1} = \bar{u}^{k+1} - u^{*k+1}$, we have

$$\begin{aligned}&\|\bar{e}^{k+1}\|^2 \leq \|\bar{u}^{k+1} - u^{*k}\|^2 \\ &= \|\bar{u}^k - u^{*k} - \eta g(u^k_{(1:N)}, y^k)\|^2 \\ &= \|\bar{e}^k - \eta\bar{g}(u^k_{(1:N)}, y^k) + \eta(\bar{g}(u^k_{(1:N)}, y^k) - g(u^k_{(1:N)}, y^k))\|^2 \\ &= \|\bar{e}^k - \eta\bar{g}(u^k_{(1:N)}, y^k)\|^2 + \eta^2\|\bar{g}(u^k_{(1:N)}, y^k) - g(u^k_{(1:N)}, y^k)\|^2 \\ &+ 2\eta(\bar{g}(u^k_{(1:N)}, y^k) - g(u^k_{(1:N)}, y^k))^\mathsf{T}(\bar{e}^k - \eta\bar{g}(u^k_{(1:N)}, y^k)) \\ &\leq (1+\eta\delta)\|\bar{e}^k - \eta\bar{g}(u^k_{(1:N)}, y^k)\|^2 \\ &\quad + \eta(\eta+\delta^{-1})\|\bar{g}(u^k_{(1:N)}, y^k) - g(u^k_{(1:N)}, y^k)\|^2.\end{aligned}$$

The first inequality holds since $u^*_{k+1}$ is the projection of $\bar{u}_{k+1}$ onto the optimality set, and thus for any other optimizer $\hat{u}^*_{k+1}$ we have $|\hat{u}^*_{k+1} - \bar{u}_{k+1}| \geq |u^*_{k+1} - \bar{u}_{k+1}|$. The last inequality follows from $\pm 2a^\mathsf{T}b \leq \delta^{-1}\|a\|^2 + \delta\|b\|^2$ for any $\sigma \geq 0$. Next, we shall bound $\|\bar{e}^k - \eta\bar{g}(u^k_{(1:N)}, y^k)\|^2$. Applying Lemma (4.3), we have

$$\begin{aligned}&\|\bar{e}^k - \eta\bar{g}(u^k_{(1:N)}, y^k)\|^2 = \|\bar{e}^k\|^2 + \eta^2\|\bar{g}(u^k_{(1:N)}, y^k)\|^2 \\ &- 2\eta\bar{e}^{k\mathsf{T}}\bar{g}(u^k_{(1:N)}, y^k) \leq \|\bar{e}^k\|^2 + \eta^2\|\bar{g}(u^k_{(1:N)}, y^k)\|^2 \\ &- \eta c_1\|\bar{g}(u^k_{(1:N)}, y^k)\|^2 - \eta c_2\|\bar{e}^k\|^2 \\ &= (1-\eta c_2)\|\bar{e}^k\|^2 + \eta(\eta - c_1)\|\bar{g}(u^k_{(1:N)}, y^k)\|^2.\end{aligned}$$

We shall pick $\eta \leq c_1$ so that $\eta(\eta - c_1)\|\bar{g}(u^k_{(1:N)}, y^k)\|^2 \leq 0$. Then, from the last two inequality arrays, we have

$$\begin{aligned}\|\bar{e}^{k+1}\|^2 &\leq (1+\eta\delta)(1-\eta c_2)\|\bar{e}^k\|^2 \\ &\quad + \eta(\eta+\delta^{-1})\|\bar{g}(u^k_{(1:N)}, y^k) - g(u^k_{(1:N)}, y^k)\|^2 \\ &\leq (1 - \eta c_2 + \eta\delta - \eta^2\delta c_2)\|\bar{e}^k\|^2 \\ &+ \eta^3(\eta+\delta^{-1})\frac{L^2_\Phi\sigma^2}{(1-\beta)^2}.\end{aligned}$$

Where the second inequality follows from (33). Note that if $\Phi$ is restricted strongly convex, then $c_1c_2 = \frac{\theta(1-\theta)\nu_\Phi}{L_\Phi} < 1$ because $\theta \in [0,1]$ and $\nu_\Phi < L_\Phi$; if $\Phi$ is strongly convex, then $c_1c_2 = \frac{\mu_\Phi L_\Phi}{(\mu_\Phi + L_\Phi)^2} < 1$. Therefore, we have $c_1 < 1/c_2$. When $\eta < c_1$, $(1+\eta\delta)(1-\eta c_2) > 0$.

Using

$$\|\bar{e}^k\|^2 \leq c_3^k\|\bar{e}^0\|^2 + \frac{1-c_3^{2k}}{1-c_3^2}c_4^2 \leq c_3^k\|\bar{e}^0\|^2\frac{c_4^2}{1-c_3^2},$$

we get

$$\|\bar{e}^k\| \leq c_3^k\|\bar{e}^0\| + \frac{c_4}{\sqrt{1-c_3^2}}. \tag{34}$$

The claim thus follows by combining (31) and (34) ■

Theorem 4.4 shows that the local agents states converge geometrically until reaching a neighborhood of the optimal solution. The size of this neighborhood depends on two quantities: $\frac{\eta\sigma}{1-\beta}$, which measures the asymptotic error due to an inexact agreement (namely, $\|u^k_{(i)} - \bar{u}^k\|$ where $\bar{u}^k := \frac{1}{N}\sum_{i=1}^{N} u^k_{(i)}$), and $\frac{c_4}{\sqrt{1-c_3^2}}$, which quantifies the asymptotic error between the average and the optimizer (namely, $\|\bar{u}^k - u^{*k}\|$). We conclude with the following remark, which relates $\frac{c_4}{\sqrt{1-c_3^2}}$ explicitly with $\eta$ and $\beta$.

*Remark 5:* ***(Refinement of bound** (27)**)*** In (27), by choosing $\delta = \frac{c_2}{2(1-\eta c_2)}$, we have $c_3 = \sqrt{1 - \frac{\eta c_2}{2}} \in (0,1)$ and

$$\begin{aligned}\frac{c_4}{\sqrt{1-c_3^2}} &= \frac{\eta L_\Phi\sigma}{1-\beta}\sqrt{\frac{\eta(\eta + \frac{2(1-\eta c_2)}{c_2})}{\frac{\eta c_2}{2}}} = \frac{\eta L_\Phi\sigma}{1-\beta}\sqrt{\frac{4}{c_2^2} - \frac{2}{c_2}\eta} \\ &\leq \frac{2\eta L_\Phi\sigma}{c_2(1-\beta)} = \mathcal{O}\left(\frac{\eta}{1-\beta}\right).\end{aligned}$$

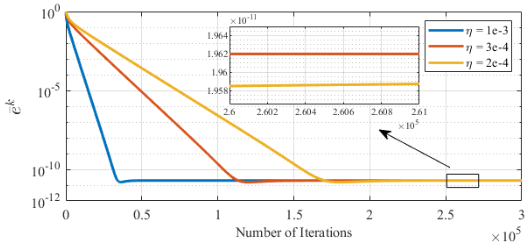
In this case, the local agent states converge geometrically to an $\mathcal{O}\left(\frac{\eta}{1-\beta} + \frac{\eta\sigma}{1-\beta}\right)$ neighborhood of the solution set $\mathcal{A}^*$ □

## V. Simulation Results

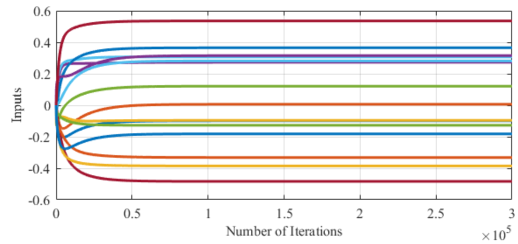
In this section, we report our numerical results. We consider a circular network consisting of $N = 15$ agents. The elements of the system matrices $B$, $C$, and $E$ are randomly drawn from the normal distribution and $A$ is chosen as a Schur stable matrix with random entries and circulant structure. We choose $n_i = 1$, $\forall i$, so that $n = N$. We choose the mixing matrix $W$ with the same circular structure as $A$ using the Metropolis weight selection. We apply (11) to:

$$\underset{u\in\mathbb{R}^m}{\text{minimize}} \ \ \frac{1}{2}\left(\|u\|^2_R + \|Gu + Hw - y^{ref}\|^2_Q\right), \tag{35}$$

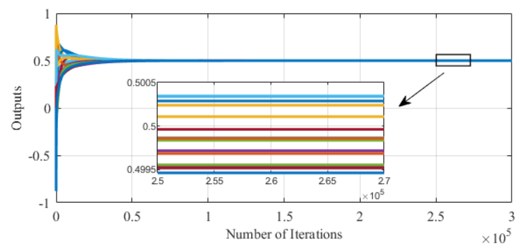
where $Q = I_p$ and $R = 0.001I_m$ are the multiplied identity matrices of the corresponding dimension. We set the desired




(a) Comparison of the proposed distributed algorithm for the problem (35) with different stepsizes.




(b) Inputs of the system with $\eta = 2 \times 10^{-4}$.




(c) Outputs of the system with $\eta = 2 \times 10^{-4}$.

Fig. 2. Error $\bar{e}^k = \frac{1}{N}\sum_{i=1}^N u_{(i)}^k - u^{*k}$, inputs, and outputs of the proposed decentralized algorithm with different stepsizes, where $u^{*k} = \text{Proj}_{\mathcal{A}^*}(\frac{1}{N}\sum_{i=1}^N u_{(i)}^k)$.

output to $y^{ref} = 0.5\mathbf{1}_p$, where $\mathbf{1}_p$ is a vector of all ones. We further scale $A$ to let $\|A\|_2 = 0.2$. The disturbance $w$ is generated from the standard uniform distribution.

Fig. 2(a) illustrates the error $\bar{e}^k$ for three different choices of stepsize. The simulations show that $\bar{e}^k$ reduces geometrically until reaching an $\mathcal{O}(\eta)$-neighborhood of the optimal point, thus validating the conclusions of Theorem 4.4. Moreover, it shows that a smaller $\eta$ causes the algorithm to converge slowly, but more accurately. Fig. 2(b) and (c), illustrate the inputs and outputs of the system. Notice that the optimizer is a point that strikes a balance between tracking $y^{ref} = 0.5\mathbf{1}_p$ and minimizing the control effort $\|u\|_R^2$.

## VI. Conclusions

We developed a distributed controller for solving optimal steady-state regulation problems with constant disturbance rejection. The distributed architecture ensures scalability and maintains the privacy of individual cost functions. We proved convergence under convexity and smoothness assumptions, and geometric convergence to a neighborhood of the optimal solution under restricted strong convexity, in line with the existing literature on distributed optimization [14]. Future work may explore local output feedback, exact convergence algorithms, constrained optimization, and nonlinear system generalizations.

## References


[1] A. Jokic, M. Lazar, and P. van den Bosch, "On constrained steady-state regulation: Dynamic KKT controllers," vol. 54, no. 9, pp. 2250–2254, 2009.

[2] E. Dall'Anese, A. Simonetto, S. Becker, and L. Madden, "Optimization and learning with information streams: Time-varying algorithms and applications," *IEEE Signal Processing Magazine*, vol. 37, no. 3, pp. 71–83, 2020.

[3] M. Colombino, E. Dall'Anese, and A. Bernstein, "Online optimization as a feedback controller: Stability and tracking," vol. 7, no. 1, pp. 422–432, 2020.

[4] A. Hauswirth, S. Bolognani, G. Hug, and F. Dörfler, "Timescale separation in autonomous optimization," vol. 66, no. 2, pp. 611–624, 2021.

[5] S. Menta, A. Hauswirth, S. Bolognani, G. Hug, and F. Dörfler, "Stability of dynamic feedback optimization with applications to power systems," in *Annual Conf. on Communication, Control, and Computing*, Oct. 2018, pp. 136–143.

[6] L. S. P. Lawrence, J. W. Simpson-Porco, and E. Mallada, "Linear-convex optimal steady-state control," 2018, to appear.

[7] G. Bianchin, J. Cortés, J. I. Poveda, and E. Dall'Anese, "Time-varying optimization of LTI systems via projected primal-dual gradient flows," vol. 9, no. 1, pp. 474–486, Mar. 2022.

[8] G. Bianchin, J. I. Poveda, and E. Dall'Anese, "Online optimization of switched LTI systems using continuous-time and hybrid accelerated gradient flows," vol. 146, p. 110579, Aug. 2022.

[9] F. Brunner, H.-B. Dürr, and C. Ebenbauer, "Feedback design for multi-agent systems: A saddle point approach," 2012, pp. 3783–3789.

[10] Z. He, S. Bolognani, J. He, F. Dörfler, and X. Guan, "Model-free nonlinear feedback optimization," 2023, in press.

[11] X. Chen, J. I. Poveda, and N. Li, "Model-free feedback constrained optimization via projected primal-dual zeroth-order dynamics," *arXiv preprint*, 2022, arXiv:2206.11123.

[12] G. Bianchin, M. Vaquero, J. Cortés, and E. Dall'Anese, "Online stochastic optimization for unknown linear systems: Data-driven synthesis and controller analysis," 2024, early Access.

[13] A. Nedić and A. Ozdaglar, "Distributed subgradient methods for multi-agent optimization," 2009, to appear.

[14] K. Yuan, Q. Ling, and W. Yin, "On the convergence of decentralized gradient descent," vol. 26, no. 3, pp. 1835–1854, 2016.

[15] A. Nedić, A. Ozdaglar, and P. A. Parrilo, "Constrained consensus and optimization in multi-agent networks," vol. 55, no. 4, pp. 922–938, 2010.

[16] J. C. Duchi, A. Agarwal, and M. J. Wainwright, "Dual averaging for distributed optimization: Convergence analysis and network scaling," vol. 57, no. 3, pp. 592–606, 2011.

[17] D. Jakovetić, J. Xavier, and J. M. Moura, "Fast distributed gradient methods," vol. 59, no. 5, pp. 1131–1146, 2014.

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

[19] C. Xi, V. S. Mai, R. Xin, E. H. Abed, and U. A. Khan, "Linear convergence in optimization over directed graphs with row-stochastic matrices," vol. 63, no. 10, pp. 3558–3565, 2018.

[20] T. Yang, X. Yi, J. Wu, Y. Yuan, D. Wu, Z. Meng, Y. Hong, H. Wang, Z. Lin, and K. H. Johansson, "A survey of distributed optimization," *Annual Reviews in Control*, vol. 47, pp. 278–305, 2019.

[21] T. Liu, Z. Qin, Y. Hong, and Z.-P. Jiang, "Distributed optimization of nonlinear multiagent systems: A small-gain approach," vol. 67, no. 2, pp. 676–691, 2021.

[22] W. Wang, Z. He, G. Belgioioso, S. Bolognani, and F. Dörfler, "Decentralized feedback optimization via sensitivity decoupling: Stability and sub-optimality," *arXiv preprint*, 2023, arXiv:2311.09408.

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

[24] G. Carnevale and G. Notarstefano, "Nonconvex distributed optimization via lasalle and singular perturbations," vol. 7, pp. 301–306, 2022.

[25] H. K. Khalil, *Nonlinear Systems*, 2nd ed., 1995.

[26] K. Yuan, W. Xu, and Q. Ling, "Can primal methods outperform primal-dual methods in decentralized dynamic optimization?" vol. 68, pp. 4466–4480, 2020.

[27] M.-J. Lai and W. Yin, "Augmented $\ell_1$ and Nuclear-norm models with a globally linearly convergent algorithm," *SIAM Journal on Imaging Sciences*, vol. 6, no. 2, pp. 1059–1091, 2013.