Distributed Feedback Optimization of Linear Multi-agent Systems

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 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 and study a distributed architecture for feedback optimization, in which each agent updates its local control state by combining the average of its neighbors with a local negative-gradient step. Under convexity and smoothness assumptions, we establish convergence of the control method to a fixed point. By reinforcing the assumptions to restricted strong convexity of the cost, 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. In fact, often the main objective of a control system is to ensure optimality in the face of unmeasurable disturbances or imprecise system knowledge. Notably, because disturbances may perturb optimal steady states, classical batch optimization algorithms fail [2] when are approximately known or vary after the optimization has been solved.

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 (or multi-agent) architecture. This connects our work with the body of literature on distributed optimization. Distributed gradient descent was proposed in [13], studied in [14], a diminishing stepsize was used in [15] to ensure exact convergence; see also [16]–[19]. Particularly related to our problem are the works on distributed optimization of nonlinear systems [20], which however relies on a separation between optimization and tracking control; [21], which relies on an inexact gradient obtained by approximating the sensitivity matrix by its diagonal elements (corresponding to a situation where the coupling between different subsystems is ignored); [22], which focuses on systems modeled as a static linear map.

This work features three main contributions. First, we propose a distributed architecture for the optimal steady-state regulation problem, and we propose 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 a 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, 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 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,

The authors are with ICTEAM Institute and the Department of Mathematical Engineering at the Université catholique de Louvain, Belgium. {amir.mehrnoosh, gianluca.bianchin}@uclouvain.be.

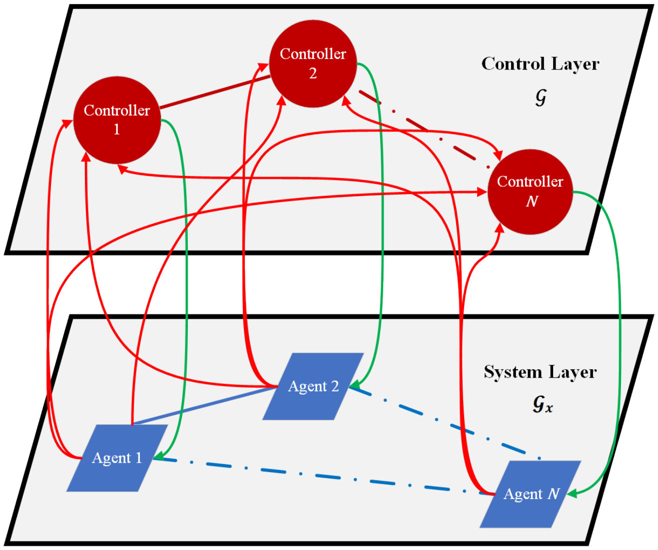


Fig. 1. Distributed system architecture considered in this work (cf. (2)). Each local controller actuates the corresponding subsystem (green lines), by using global feedback information (red lines), see (9).

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 $W = [w_{ij}]$ is symmetric and doubly stochastic. The eigenvalues of W are real and sorted in a nonincreasing order $1 = \lambda_1(W) \geq \lambda_2(W) \geq \cdots \geq \lambda_n(W) > -1$. Let the second-largest eigenvalue magnitude of W be:

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

II. Problem Setting

A. System structure and problem formulation

Consider a distributed system composed of N subsystems $\mathcal{V}_x = \{1, \dots, N\}$; we describe the physical couplings between interacting 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 $x_i^k \in \mathbb{R}^{n_i}$, which updates for $k \in \mathbb{Z}_{\geq 0}$ as:

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

where $\mathcal{N}_i$ denotes the set of in-neighbors of 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. In what follows, we let $n := \sum_i n_i$, $m := \sum_i m_i$, and $r := \sum_i r_i$. We assume that the local controllers have access to global output measurements of the system state, of the form:

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

where $C_i \in \mathbb{R}^{p_i \times n_i}$ and $D_i \in \mathbb{R}^{p_i \times m_i}$; we let $p := \sum_i p_i$.

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 $x^k = (x_1^k, \dots, x_n^k) \in \mathbb{R}^N$ is the vector of states, $u^k = (u_1^k, \dots, u_N^k) \in \mathbb{R}^m$ is the vector of inputs, $w = (w_1, \dots, w_N) \in \mathbb{R}^r$ is the vector of disturbances, and $A = [A_{ij}]$, $B = [B_{ij}]$, $C = [C_1, \dots, C_N]$, $D = [D_1, \dots, D_N]$, and $E = [E_1, \dots, E_N]$ are block matrices. In what follows, we use the notation

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

to 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. In what follows, we will adopt the compact notation

$$G := G(1), \qquad H := H(1).$$

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; when A is not Schur stable, it can be pre-stabilized using standard static state feedback techniques. Since the disturbance w is constant, at steady-state, x, u, and y have constant values $x^{\text{ss}}, u^{\text{ss}}$, and y^{ss} (by Assumption 1), with

$$y^{\text{ss}} = Gu^{\text{ss}} + Hw.$$

In this work, we study a control problem where the ensemble of controllers seeks to collaboratively optimize the system at steady-state:

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

where Φ_i is assumed to be private and locally known only by agent i. We remark that solutions to the optimization (6) cannot be computed explicitly since w is unknown. We make the following assumptions on the optimization.

Assumption 2: ***(Lipschitz and Convexity of cost)*** For all i, $(u, y) \mapsto \Phi_i(u, y)$ is proper closed convex, lower bounded, and Lipschitz differentiable with constant L_{Φ_i}. □

Assumption 2 is standard in optimization. By defining $\Pi^\mathsf{T} := \begin{bmatrix} I_m & G^\mathsf{T} \end{bmatrix}$, this assumption guarantees that there exists L_Φ such that

$$\|\Pi^\mathsf{T}(\nabla\Phi(u, y) - \nabla\Phi(u', y'))\| \leq L_\Phi \left\| \begin{bmatrix} u \\ y \end{bmatrix} - \begin{bmatrix} u' \\ y' \end{bmatrix} \right\|, \tag{7}$$

for all $y, y' \in \mathbb{R}^n$ and $u, u' \in \mathbb{R}^m$. Further, in what follows we denote the set of optimizers of (6) by:

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

We assume that this set is nonempty and closed. We will further denote in compact form:

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

To describe the collaborative nature of the controllers, we will adopt an undirected graph (hereafter called control graph) $\mathcal{G}_u = (\mathcal{V}_u, \mathcal{E}_u)$ 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 agents can collaboratively compute a control law if and only if they are connected by means of a communication link in $\mathcal{E}_u$. Recall that a graph is connected if there exists a path between any two nodes. We make the following assumption on the control graph.

Assumption 3: ***(Connectivity of the control graph)*** The graph $\mathcal{G}_u$ is connected. □

Under Assumption 3, there exists $W = [w_{ij}] \in \mathbb{R}^{n\times n}$, such that $(i,j) \notin \mathcal{E}_u$, implies $w_{ij} = 0$ and that is symmetric and doubly stochastic with $\beta < 1$ (see (1)).

III. Distributed controller design

A centralized version 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:

$$u^{k+1} = u^k - \eta \Pi^\mathsf{T} \nabla \Phi(u^k, y^k), \tag{8}$$

where $\eta > 0$ denotes a scalar stepsize, being a design parameter. The controller (8) 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 measurements-based version $\Pi^\mathsf{T}\nabla\Phi(u^k, y^k)$, which avoids the need to measure w.

Unfortunately, implementing (8) requires a centralized knowledge of the gradients $\{\nabla\Phi_i\}_{i\in\mathcal{V}_u}$, which is impractical in our case since each Φ_i is known only locally. Departing from this, we next propose a distributed version of (8) 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 $u^k_{(i)} \in \mathbb{R}^m$ of u^k, and updates it as:

$$u^{k+1}_{(i)} = \sum_{j=1}^{N} w_{ij} u^k_{(j)} - \eta \Pi^\mathsf{T} \nabla \Phi_i(u^k_{(i)}, y^k). \tag{9}$$

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 $\sum_{j=1}^n w_{ij} u^k_{(j)}$ to seek a consensus between the agents, and (ii) it applies $-\Pi^\mathsf{T}\nabla\Phi_i(u^k_{(i)}, y^k)$ to decrease $\Phi_i(u^k_{(i)}, y^k)$. We remark that this control law is distributed in the sense that each agent i requires only knowledge of the local $\nabla\Phi_i$ (rather than of all gradients $\{\nabla\Phi_i\}_{i\in\mathcal{V}_u}$). Notice also that each agent requires measures of the global output feedback signal y^k; we leave a generalization of the approach to the use of local output feedback signals as the scope of future works. See Fig. 1 for an illustration of the control architecture.

IV. Convergence analysis and error bounds

In this section, we study the convergence of (9) 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

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

In vector form, the system (4) controlled by (9) reads as:

$$x^{k+1} = Ax^k + BSu^k_{(1:N)} + Ew, \tag{10a}$$

$$\begin{aligned} y^k &= Cx^k + DSu^k_{(1:N)}, \\ u^{k+1}_{(i)} &= \sum_{j\in\mathcal{N}_i} w_{ij} u^k_{(j)} - \eta\Pi^\mathsf{T}\nabla\Phi_i(u^k_{(i)}, y^k), \quad i \in \mathcal{V}_u, \end{aligned} \tag{10b}$$

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

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

A. Asymptotic convergence

The following result shows that, under a suitable choice of the stepsize η, the state of (10) converges asymptotically.

Theorem 4.1: ***(Convergence of the state sequences)*** Let Assumptions 1-2 hold, W be such that $\beta < 1$, and the stepsize $\eta \le \bar\eta := \min\{\eta_1, \eta_2, \eta_3\}$, where

$$\eta_1 = \frac{1 - 2\mu + \lambda_n(W)}{L_\Phi}, \qquad \eta_2 = \frac{\mu}{\lambda_1(P) L_h^2}, \tag{11}$$

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

with μ 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 (10) converges. □

Proof: We will prove this claim by using La Salle's Invariance Principle [23, 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 (10a) to the origin. Inspired by singular-perturbation reasonings [23, Sec. 11], consider the storage function:

$$U(u_{(1:N)}, \tilde{x}) := \frac{d}{\eta} V_u(u_{(1:N)}) + (1-d)V_x(\tilde{x}), \tag{12}$$

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

$$\begin{aligned} V_u(u_{(1:N)}) &= -\frac{1}{2}\sum_{i,j=1}^{N} w_{ij} u^\mathsf{T}_{(i)} u_{(j)} \\ &\quad + \sum_{i=1}^{N}\left(\frac{1}{2}\|u_{(i)}\|^2 + \eta\Phi_i(u_{(i)}, Gu_{(i)} + Hw)\right), \\ V_x(\tilde{x}) &= \tilde{x}^\mathsf{T} P \tilde{x}. \end{aligned} \tag{13}$$

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

$$\tilde{V}_u(u_{(1:N)}, \tilde{x}) = -\frac{1}{2}\sum_{i,j=1}^N w_{ij} u_{(i)}^\top u_{(j)} + \sum_{i=1}^N \left(\frac{1}{2}\|u_{(i)}\|^2 + \eta \Phi_i(u_{(i)}, C\tilde{x} + GSu_{(i)} + Hw)\right), \tag{14}$$

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, (10b) and (14) can be re-expressed as:

$$\begin{aligned} u_{(1:N)}^{k+1} &= u_{(1:N)}^k - F_c(u_{(1:N)}^k, \tilde{x}^k), \\ V_u(u_{(1:N)}) &= \tilde{V}_u(u_{(1:N)}, 0), \end{aligned} \tag{15}$$

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

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

$$\begin{aligned} V_u(u_{(1:N)}^{k+1}) &\leq V_u(u_{(1:N)}^k) + \nabla V_u(u_{(1:N)}^k)^\top (u_{(1:N)}^{k+1} - u_{(1:N)}^k) \\ &\qquad + \frac{L_{V_u}}{2}\|u_{(1:N)}^{k+1} - u_{(1:N)}^k\|^2 \\ &= V_u(u_{(1:N)}^k) - \nabla V_u(u_{(1:N)}^k)^\top F_c(u_{(1:N)}^k, \tilde{x}^k) \\ &\qquad + \frac{L_{V_u}}{2}\|F_c(u_{(1:N)}^k, \tilde{x}^k)\|^2 \\ &\leq V_u(u_{(1:N)}^k) - \|F_c(u_{(1:N)}^k, \tilde{x}^k)\|^2 \\ &\qquad + \|F_c(u_{(1:N)}^k, \tilde{x}^k)\| \|\nabla V_u(u_{(1:N)}^k)^\top - F_c(u_{(1:N)}^k, \tilde{x}^k)\| \\ &\qquad + \frac{L_{V_u}}{2}\|F_c(u_{(1:N)}^k, \tilde{x}^k)\|^2 \\ &\leq V_u(u_{(1:N)}^k) - \left(1 - \frac{L_{V_u}}{2}\right)\|F_c(u_{(1:N)}^k, \tilde{x}^k)\|^2 \\ &\qquad + \|F_c(u_{(1:N)}^k, \tilde{x}^k)\| \|\nabla V_u(u_{(1:N)}^k)^\top - F_c(u_{(1:N)}^k, x^k)\| \\ &\leq V_u(u_{(1:N)}^k) - \mu \|F_c(u_{(1:N)}^k, x^k)\|^2 \\ &\qquad + \eta L_\Phi \|F_c(u_{(1:N)}^k, x^k)\| \|\tilde{x}^k\| \end{aligned} \tag{16}$$

where the first row follows by convexity and Lipschitz differentiability of V_u, the second row follows by application of (15), the third row follows by adding and subtracting $F_c(u_{(1:N)}^k, \tilde{x}^k)$ to $\nabla V_u(u_{(1:N)}^k)$ and using the triangle inequality, and the last row follows by Lipschitz differentiability of Φ and definition of μ.

3) Bounding the variation of $V_x(\cdot)$. In the variables $\tilde{x}^k$, the plant dynamics (10a) read as:

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

We have:

$$\begin{aligned} V_x(\tilde{x}^{k+1}) &= (\tilde{x}^k)^\top A^\top P A \tilde{x}^k \\ &\qquad + 2(\tilde{x}^k)^\top A^\top P(h(u_{(1:N)}^k) - h(u_{(1:N)}^{k+1})) \\ &\qquad + (h(u_{(1:N)}^k) - h(u_{(1:N)}^{k+1}))^\top P (h(u_{(1:N)}^k) - h(u_{(1:N)}^{k+1})) \\ &\leq V_x(\tilde{x}^k) - (\tilde{x}^k)^\top Q \tilde{x}^k + \lambda_1(P)\|h(u_{(1:N)}^k) - h(u_{(1:N)}^{k+1})\|^2 \\ &\qquad + 2\|A^\top P\| \|\tilde{x}^k\| \|h(u_{(1:N)}^k) - h(u_{(1:N)}^{k+1})\| \\ &\leq V_x(\tilde{x}^k) - \lambda_n(Q)\|\tilde{x}^k\|^2 + \lambda_1(P) L_h^2 \|u_{(1:N)}^k - u_{(1:N)}^{k+1}\|^2 \\ &\qquad + 2L_h \|A^\top P\| \|\tilde{x}^k\| \|u_{(1:N)}^k - u_{(1:N)}^{k+1}\| \\ &= V_x(\tilde{x}^k) - \lambda_n(Q)\|\tilde{x}^k\|^2 + \lambda_1(P) L_h^2 \|F_c(u_{(1:N)}^k, x^k)\|^2 \\ &\qquad + 2L_h \|A^\top P\| \|\tilde{x}^k\| \|F_c(u_{(1:N)}^k, x^k)\|, \end{aligned} \tag{17}$$

where the second row follows from Assumption 1, the third row from Lipschitz continuity of h, and the last row by application of (15).

4) Combining the bounds. By combining (16) and (17), we conclude $U(u_{(1:N)}^{k+1}, \tilde{x}^{k+1}) - U(u_{(1:N)}^k, \tilde{x}^k) \leq -\xi(u_{(1:N)}^k, \tilde{x}^k)^\top \Lambda \xi(u_{(1:N)}^k, \tilde{x}^k)$, where

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

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

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

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 η_2 and η_3 in (11). To conclude, since $U(u_{(1:N)}^k, \tilde{x}^k)$ strongly decreases along the trajectories of (10) and its minimum is the point$(\|F_c(u_{(1:N)}^k, \tilde{x}^k)\|, \|\tilde{x}^k\|) = (0,0)$, by La Salle's invariance principle the claim follows. ■

Theorem 4.1 shows that under a sufficiently small choice of the stepsize η, the sequences of states $x^k, u_{(i)}^k$ of, respectively, the system and controller states, converge asymptotically to a fixed point. Notice that this convergence claim is not straightforward, since the proposed controller (9) incorporates two simultaneous steps: a consensus one and a gradient step. As such, this update may oscillate or fail to converge when η is chosen inadequately. The upper bounds η_1, η_2, η_3 depend on the various parameters of the system (4) and optimization problem (6). Importantly, notice that the imposed bounds on μ guarantee that $\bar{\eta} > 0$ is well-defined.

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 (9) as:

$$u^{k+1}_{(1:N)} = (W \otimes I)u^k_{(1:N)} - \eta\,\gamma(u^k_{(1:N)}, y^k), \tag{18}$$

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 $k \in \mathbb{Z}_{\geq 0}$, (18) satisfies:

$$\|\gamma(u^k_{(1:N)}, y^k)\| \leq \sigma, \tag{19}$$

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 (19), we rewrite

$$\begin{aligned}\sum_{i=1}^N \Phi_i(u^k_{(i)}, y^k) &\leq \eta^{-1}\tilde{V}(u^k_{(1:N)}, \tilde{x}^k) \qquad\qquad (20)\\ &\leq \eta^{-1}\tilde{V}(u^0_{(1:N)}, \tilde{x}^0) \quad = \sum_{i=1}^N \Phi_i(u^0_{(i)}, y^0)\end{aligned}$$

where the first inequality follows from (14) and by using $\sum_{i=1}^N \|u_{(i)}\|^2 - \sum_{i,j=1}^N w_{ij} u_{(i)}^\mathsf{T} u_{(j)}^\mathsf{T} \geq 0$ (which holds since $\beta < 1$), the second inequality holds from $\tilde{V}_u(u^{k+1}_{(1:N)}, \tilde{x}^{k+1}) \leq \tilde{V}_u(u^k_{(1:N)}, \tilde{x}^k)$ (which follows by iterating the steps in (16) 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 (13) 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 $g(u_a, y_a) \geq g(u_b, y_b) + \nabla g^\mathsf{T}(u_b, y_b)[(u_a - u_b), (y_a - y_b)]^\mathsf{T} + \frac{1}{2L_g}\|\nabla g(u_a, y_a) - \nabla g(u_b, y_b)\|^2$ and $\nabla g(u^*, y^*) = 0$. Then, $\|\nabla g(u, y)\|^2 \leq 2L_g(g(u, y) - g^*)$, where $g^* := g(u^*, y^*)$. Using this inequality and (20), we obtain

$$\begin{aligned}\|\gamma(u^k_{(1:N)}, y^k)\|^2 &= \sum_{i=1}^N \|\Pi^\mathsf{T} \nabla\Phi_i(u^k_{(i)}, y^k)\|^2 \qquad (21)\\ &\leq \sum_{i=1}^N 2L_\Phi(\Phi_i(u^k_{(i)}, y^k) - \Phi_i^o)\\ &\leq 2L_\Phi\left(\sum_{i=1}^N \Phi_i(u^0_{(i)}, y^0) - \Phi^{\text{opt}}\right),\end{aligned}$$

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

Proposition 4.2 ensures that the sequence of gradients $\gamma(u^k_{(1:N)}, y^k)$ 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 $u^0_{(1:N)}$ are nonzero, a uniform bound of the form (19) can still be proven by adjusting the step (21), but σ 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 [24] with modulus ν_f if

$$(\nabla f(z) - \nabla f(z^*))^\mathsf{T}(z - z^*) \geq \nu_f \|z - z^*\|^2, \tag{22}$$

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: [24, Lemma 6] Suppose that f is restricted strongly convex with modulus ν_f and ∇f is Lipschitz continuous with constant L_f. Then, we have

$$\begin{aligned}&(z - z^*)^\mathsf{T}(\nabla f(z) - \nabla f(z^*)) \qquad\qquad (23)\\ &\qquad \geq c_1\|\nabla f(z) - \nabla f(z^*)\|^2 + c_2\|z - z^*\|^2,\end{aligned}$$

where z^* is as in (22). Moreover, for any $\theta \in [0, 1]$,

$$c_1 = \frac{\theta}{L_f}, \qquad c_2 = (1 - \theta)\nu_f. \tag{24}$$

Remark 1: Notice that, if f is strongly convex with modulus ν_f, then it is also restricted strong convexity with the same modulus [24]. In this case, (23) holds with

$$c_1 = \frac{1}{\nu_f + L_f}, \qquad c_2 = \frac{\nu_f L_f}{\nu_f + L_f}. \tag{25}$$

□

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 ν_Φ, then, for (10) it holds:

$$\|u^k_{(i)} - u^{*k}\| \leq c_3^k \|u^0_{(i)} - u^{*0}\| + \frac{c_4}{\sqrt{1 - c_3^2}} + \frac{\eta\sigma}{1 - \beta}, \tag{26}$$

where

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

$(u^{*k}, x^{*k}) := \operatorname{Proj}_{\mathcal{A}^*}(\bar{u}^k, x^k)$, σ is as in (19), $\delta > 0$ is an arbitrary constant, and c_1 and c_2 are as in (24) with $\nu_f = \nu_\Phi/N$. Moreover, if $\Phi(u, y)$ is strongly convex, then c_1 and c_2 are as in (25). □

Proof: We begin by proving (26). It will be convenient to measure the control error relative to the average controller state: $\bar{u}^k := \frac{1}{n}\sum_{i=1}^n u^k_{(i)}$. We have:

$$\|u^k_{(i)} - u^{*k}\| \leq \|u^k_{(i)} - \bar{u}^k\| + \|\bar{u}^k - u^{*k}\|. \tag{27}$$

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

1) Bound for $\|u_{(i)}^k - \bar{u}^k\|$. By expanding (18) in time:

$$u_{(1:N)}^k = -\eta \sum_{s=0}^{k-1} (W^{k-1-s} \otimes I)\gamma(u_{(1:N)}^s, y^s). \tag{28}$$

Next, let $\bar{u}_{(1:N)}^k = (\bar{u}^k, \cdots, \bar{u}^k) \in \mathbb{R}^{nm}$, 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,

$$\begin{aligned}
&\|u_{(i)}^k - \bar{u}^k\| \le \|u_{(1:N)}^k - \bar{u}_{(1:N)}^k\| \\
&\quad = \|u_{(1:N)}^k - \frac{1}{n}((1_n 1_n^\mathsf{T}) \otimes I)u_{(1:N)}^k\| \\
&\quad = \| - \eta \sum_{s=0}^{k-1} (W^{k-1-s} \otimes I)\gamma(u_{(1:N)}^s, y^s) \\
&\quad + \eta \sum_{s=0}^{k-1} \frac{1}{n}((1_n 1_n^\mathsf{T} W^{k-1-s}) \otimes I)\gamma(u_{(1:N)}^s, y^s)\| \\
&\quad = \| - \eta \sum_{s=0}^{k-1} (W^{k-1-s} \otimes I)\gamma(u_{(1:N)}^s, y^s) \\
&\quad + \eta \sum_{s=0}^{k-1} \frac{1}{n}((1_n 1_n^\mathsf{T}) \otimes I)\gamma(u_{(1:N)}^s, y^s)\| \\
&\quad = \eta \| \sum_{s=0}^{k-1} \left(\left(W^{k-1-s} - \frac{1}{n} 1_n 1_n^\mathsf{T} \right) \otimes I \right) \gamma(u_{(1:N)}^s, y^s)\| \\
&\quad \le \eta \sum_{s=0}^{k-1} \|W^{k-1-s} - \frac{1}{n} 1_n 1_n^\mathsf{T}\| \|\gamma(u_{(1:N)}^s, y^s)\| \\
&\quad = \eta \sum_{s=0}^{k-1} \beta^{k-1-s} \|\gamma(u_{(1:N)}^s, y^s)\|,
\end{aligned} \tag{29}$$

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

$$\begin{aligned}
\|u_{(i)}^k - \bar{u}^k\| &\le \eta \sum_{s=0}^{k-1} \beta^{k-1-s} \|\gamma(u_{(1:N)}^s, y^s)\| \le \sum_{s=0}^{k-1} \beta^{k-1-s}\sigma \\
&\le \frac{\eta\sigma}{1-\beta}.
\end{aligned} \tag{30}$$

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

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

We are interested in $g(u_{(1:N)}^k, y^k)$ because $-\eta g(u_{(1:N)}^k, y^k)$ updates $\bar{u}^k$. To see this, by taking the average of (9) over i and noticing $W = [w_{ij}]$ is doubly stochastic, we obtain

$$\begin{aligned}
\bar{u}^{k+1} &= \frac{1}{n} \sum_{i=1}^N u_{(i)}^{k+1} \\
&= \frac{1}{n} \sum_{i,j=1}^N w_{ij} u_{(j)}^k - \frac{\eta}{n} \sum_{i=1}^N \Pi^\mathsf{T} \nabla \Phi_i(u_{(i)}^k, y^k) \\
&= \bar{u}^k - \eta g(u_{(1:N)}^k, y^k).
\end{aligned} \tag{31}$$

Before proceeding notice that the following bound holds:

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

by Assumptions 2, 3, and (7), and where the last inequality follows from (30). Moreover, we also have:

$$\begin{aligned}
&\|g(u_{(1:N)}^k, y^k) - \bar{g}(u_{(1:N)}^k, y^k)\| \\
&\qquad = \|\frac{1}{n} \sum_{i=1}^N \Pi^\mathsf{T}(\nabla \Phi_i(u_{(i)}^k, y^k) - \nabla \Phi_i(\bar{u}^k, y^k))\| \\
&\qquad \le \frac{1}{n} L_\Phi \sum_{i=1}^N \|u_{(i)}^k - \bar{u}^k\| \\
&\qquad \le \frac{\eta\sigma L_\Phi}{1-\beta}.
\end{aligned} \tag{32}$$

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

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

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 $|\hat{u}_{k+1}^* - \bar{u}_{k+1}| \ge |u_{k+1}^* - \bar{u}_{k+1}|$. The last inequality follows from $\pm 2a^\mathsf{T} b \le \delta^{-1}\|a\|^2 + \delta\|b\|^2$ for any $\sigma \ge 0$. Next, we shall bound $\|\bar{e}^k - \eta \bar{g}(u_{(1:N)}^k, y^k)\|^2$. Applying Lemma (4.3), we have

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

We shall pick $\eta \le c_1$ so that $\eta(\eta - c_1)\|\bar{g}(u_{(1:N)}^k, y^k)\|^2 \le 0$.

Then, from the last two inequality arrays, we have

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

Where the second inequality follows from (32). Note that if Φ 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 Φ 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

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

we get

$$\|\bar{e}^k\| \le c_3^k\|\bar{e}^0\| + \frac{c_4}{\sqrt{1-c_3^2}}. \tag{33}$$

The claim thus follows by combining (30) and (33) ■

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 η and β.

Remark 2: ***(Refinement of bound** (26)**)*** In (26), 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

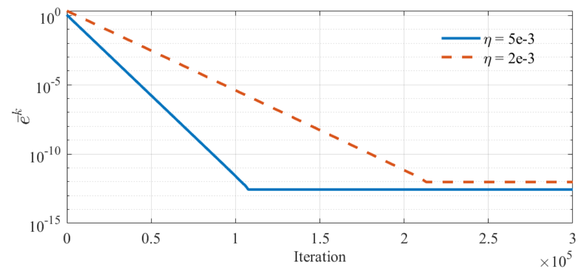
$$\begin{aligned}\frac{c_4}{\sqrt{1-c_3^2}} &= \frac{\eta L_\Phi\sigma}{1-\beta}\sqrt{\frac{\eta(\eta+\frac{2(1-\eta c_2)}{c_2})}{\frac{\eta c_2}{2}}} = \frac{\eta L_\Phi\sigma}{1-\beta}\sqrt{\frac{4}{c_2^2}-\frac{2}{c_2}\eta}\\ &\le \frac{2\eta L_\Phi\sigma}{c_2(1-\beta)} = \mathcal{O}\left(\frac{\eta}{1-\beta}\right).\end{aligned}$$

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 $\mathcal{A}^*$ □

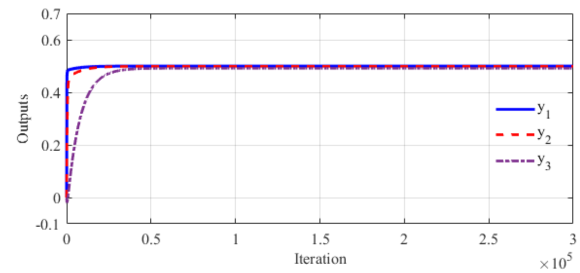
V. Simulation Results

In this section, we report our numerical results. Consider a system consisting of three agents with the dynamics of A, B, and C matrices randomly chosen from the normal distribution. As for A, we consider generating a network consisting of N agents with $\frac{N(N-1)}{2}\kappa$ edges that are uniformly randomly chosen, where $N = 3$ and $\kappa = 0.5$. We choose $n_i = 1$, $\forall i$, so that $n = N$. The same setting is used for the control layer network except that in this case, we choose the mixing matrix Wso that it is symmetric and doubly stochastic. Moreover, we make sure that $\mathcal{G}_u$ is connected. We apply (10) to the problem

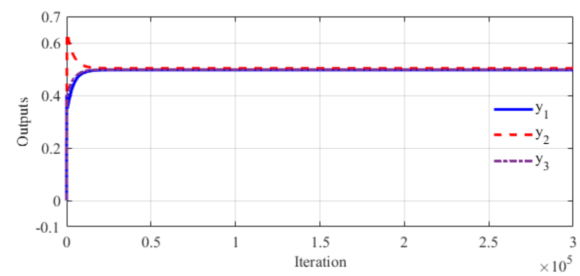
$$\underset{u\in\mathbb{R}^m}{\text{minimize}} \quad \frac{1}{2}\left(\|u\|_R^2 + \|Gu + Hw - y^{ref}\|_Q^2\right), \tag{34}$$



(a) Comparison of the proposed decentralized algorithm for the problem (34) with different fixed stepsizes.



(b) Outputs of the system with $\eta = 2\times 10^{-3}$.



(c) Outputs of the system with $\eta = 5\times 10^{-3}$.

Fig. 2. Error $\bar{e}^k = \frac{1}{n}\sum_{i=1}^n u^k_{(i)} - u^{*k}$ and outputs of the proposed decentralized algorithm with different stepsizes.

where $Q = I_m$, $R = 0.001I_p$, and $y^{ref} = 0.5I_p$. Notice that the cost function in (34) is strongly convex in this case.

Fig. 2(a) illustrates the convergence of the error $\bar{e}^k$ related to two different stepsizes. Note that $\bar{e}^k$ reduces linearly until reaching an $\mathcal{O}(\eta)$-neighborhood, thus validating the conclusions of Theorem 4.4. Moreover, it shows that a smaller η causes the algorithm to converge slower. Fig. 2(b) and (c), illustrate that the outputs of the system reach the desired output.

VI. Conclusions

We proposed a distributed controller to solve optimal steady-state regulation problems while rejecting constant disturbances. The controller follows a distributed architecture; as such, the approach scales well with the system size and can be applied in cases where the individual cost functions need to be maintained private. Under convexity and smoothness assumptions, we showed that the controller state converges; under restricted strong convexity assumptions, we showed that the controller converges geometrically to a neighborhood of the optimal solution, in line with the existing literature on distributed optimization [14]. Our work opens the opportunity for several future works, including scenarios where the output feedback signals are also local, an investigation of algorithms that can ensure exact convergence, the study of constrained optimization objectives, and a generalization to nonlinear systems.

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. 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.

[21] 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.

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

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

[24] M.-J. Lai and W. Yin, "Augmented ℓ_1 and Nuclear-norm models with a globally linearly convergent algorithm," *SIAM Journal on Imaging Sciences*, vol. 6, no. 2, pp. 1059–1091, 2013.