Optimization of Linear Multi-Agent Dynamical Systems via Feedback Distributed Gradient Descent Methods

Amir Mehrnoosh and Gianluca Bianchin

***Abstract*— Feedback optimization is a control paradigm for optimizing dynamical systems at steady-state. Existing methods rely on centralized architectures, limiting scalability and privacy in large-scale systems. We propose a distributed feedback optimization approach inspired by the Distributed Gradient Descent method, where each agent updates its control variable using local gradients and average of neighbors. Under convexity and smoothness assumptions, we establish convergence to a critical optimization point, and under restricted strong convexity, we prove linear convergence to a neighborhood of the optimum, with its size dependent on the stepsize. Simulations corroborate the theoretical results.**

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

I. Introduction

Optimal steady-state regulation aims to control dynamical systems toward an optimal steady-state defined by an optimization problem [1]. A classical approach separates planning and control, solving the optimization offline to determine optimal states, which are then used as reference inputs for controllers. However, a key assumption in this method is that disturbances are known in advance which allows high-precision solutions. In reality, disturbances are often unknown, and control systems must ensure optimality despite uncertainty. Notably, classical batch optimization methods fail [2] when disturbances are only approximately known or vary after the optimization has been solved, as they may shift the optimal steady states.

Recent studies on optimal steady-state regulation have focused on centralized architectures [3]–[9], including approaches using zeroth-order and data-driven methods [10]–[12]. Feedback optimization controllers have gained attention for their ability to regulate systems to optimal steady states while rejecting constant or time-varying disturbances [4], [7], [8]. These methods adapt numerical optimization into feedback control by estimating gradients via real-time measurements which eliminates the need for precise plant models and disturbances. However, centralized architectures face scalability and privacy challenges in large-scale systems when cost functions or feedback signals need to be maintained private. This work addresses these issues by focusing on the problem of optimal steady-state regulation for systems with a distributed architecture. Distributed gradient descent (DGD) was introduced in [13], studied in [14], a diminishing stepsize was used in [15] for exact convergence; see also [16]–[19]. Unlike [20], we do not need a two-layer control architecture and tracking controllers; in contrast to [21], 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; [22] assumes a static linear map, while we consider system dynamics; and unlike [23], we optimize performance metrics that depend on vector quantities rather than scalar aggregates.

This work has three main contributions. First, we propose a distributed architecture and control algorithm for the optimal steady-state regulation problem, combining gradient descent with a consensus step to simultaneously solve an optimization and seek an agreement between the agents. Second, we establish convergence to a fixed point for the controller-system state, and provide guidelines for selecting the controller stepsize to ensure convergence of the controlled system. Third, we derive an explicit bound on the control error, showing that under restricted strong convexity, the controller state converges linearly to a neighborhood of the optimum. In line with the prior work [14], the neighborhood size depends on the controller stepsize, balancing gradient descent and consensus.

The paper is structured as follows. Section II defines the problem focus of thiw work. Section III introduces the proposed controller. Section IV presents the main theoretical results, followed by numerical validation in Section V. Finally, Section VI concludes the paper.

Notation. For a symmetric matrix W, we denote its eigenvalues by $\lambda_1(W) \geq \lambda_2(W) \geq \cdots \geq \lambda_N(W)$. Given a symmetric and doubly stochastic matrix W, we let the eigenvalues be sorted in a nonincreasing order: $1 = \lambda_1(W) \geq \lambda_2(W) \geq \cdots \geq \lambda_N(W) > -1$. Moreover, we let $\beta := \max\{|\lambda_2(W)|, |\lambda_N(W)|\}$.

II. Problem Setting

A. System to control

Consider a dynamical system composed of N subsystems, indexed by $\mathcal{V}_x = \{1, \ldots, N\}$; 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:

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

where $\mathcal{N}_i$ denotes the set of subsystems interacting 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 (1), $q_i \in \mathbb{R}^{r_i}$ with $E_i \in \mathbb{R}^{n_i \times r_i}$,

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.

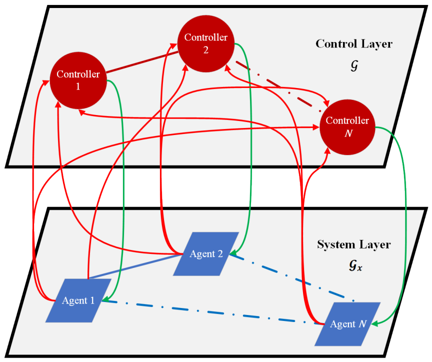

Fig. 1. Distributed system architecture considered in this work (cf. (1)). Each local controller actuates the corresponding subsystem (green lines), by using global feedback information (red lines), see (8).

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 a global output signal:

$$y^k = \sum_{i=1}^{N} C_i x_i^k + D_i u_i^k, \tag{2}$$

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

*Remark 1: **(Output model)*** Equation (2) assumes the availability of a common, or centralized, output signal. Although alternative output models could be considered (e.g., where each i measures $y_i^k = C_i x_i^k + D_i u_i^k$), the model (2) is common to many practical applications. For example, consider a swarm of drones where each drone communicates wirelessly with a common base station. □

In vector form, (1) reads as:

$$\begin{aligned} x^{k+1} &= Ax^k + Bu^k + Eq, \\ y^k &= Cx^k + Du^k, \end{aligned} \tag{3}$$

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, $q = (q_1, \ldots, q_N) \in \mathbb{R}^r$ 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 $G(z) = C(zI - A)^{-1}B + D$, $H(z) = C(zI - A)^{-1}E$ denote the transfer functions from u to y and from q to y, respectively. This definition 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 $u_k = \bar{u}, q_k = \bar{q}$ for all $k \in \mathbb{Z}_{\geq 0}$:

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

*Assumption 1: **(Stability and control properties of system)*** The model (1) is asymptotically stable, controllable, and observable. □

Thanks to Assumption 1, in what follows, we will fix a matrix $Q \succ 0$, and we let $P \succ 0$ be such that $A^\mathsf{T} PA - P = -Q$. Controllability and observability are standard assumptions to guarantee that control problems are well-defined. Moreover, we assume that the system has been pre-stabilized as in Assumption 1. This can be achieved using well-established static state feedback techniques [24].

B. Distributed structure of the controller

We consider a system controlled by distributed controllers, each co-located with a local subsystem and actuating its corresponding control variable (see Fig.1). The combination of a subsystem and its controller is referred to as an agent. Controllers collaborate through an undirected *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 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. □

Under Assumption 2, there exists a symmetric and doubly stochastic matrix $W = [w_{ij}] \in \mathbb{R}^{N \times N}$ (which will be called *mixing matrix*) such that $(i, j) \notin \mathcal{E}_u$ implies $w_{ji} = 0$, and that satisfies $\beta < 1$.

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

$$\underset{u \in \mathbb{R}^m}{\text{minimize}} \quad \sum_{i=1}^{N} \Phi_i(u, Gu + Hq), \tag{4}$$

where $\Phi_i : \mathbb{R}^m \times \mathbb{R}^p \to \mathbb{R}$. The optimization problem (4) 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 $Gu+Hq$). Moreover, the cost in (4) has a separable structure, allowing for cases where $\Phi_i(\cdot, \cdot)$ is known locally only by agent i. We also remark that optimization problem (4) is *parametrized* by q; as such, its solutions cannot be computed using standard optimization solvers, since the disturbance q is unknown and unmeasurable.

*Remark 2: **(Practical relevance of** (4)**)*** Allowing the local costs $\Phi_i(\cdot, \cdot)$ in (4) to depend on the global system input u and the global steady-state output $Gu + Hq$ enables our framework to model systems where agents have individual performance metrics (namely, $\Phi_i(\cdot, \cdot)$), but collectively optimize a shared objective, as in (4). This applies to scenarios like energy systems, where each subsystem evaluates optimality differently.

Moreover, notice that a special case of (4) is:

$$\underset{u \in \mathbb{R}^m}{\text{minimize}} \quad \sum_{i=1}^{N} \tilde{\Phi}_i(u_i, Gu + Hq), \tag{5}$$

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 $Gu + Hq$), 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. □

In the remainder, 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_{Φ_i}. □

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

$$\|\Pi^\top(\nabla\Phi(u,y) - \nabla\Phi(u',y'))\| \le L_\Phi \left\| \begin{bmatrix} u \\ y \end{bmatrix} - \begin{bmatrix} u' \\ y' \end{bmatrix} \right\|, \quad (6)$$

where $\Pi^\top := \begin{bmatrix} I_m & G^\top \end{bmatrix}$, 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 (4) by

$$\mathcal{A}^* := \{(u^*, x^*) \ : \ (u^*, x^*) \text{ is a first-order optimizer of (4)}\},$$

and we assume that this set is nonempty and closed.

III. Desing of the F-DGD Method

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

$$u^{k+1} = u^k - \eta \Pi^\top \nabla\Phi(u^k, y^k), \quad (7)$$

where $\eta > 0$ denotes a scalar stepsize, being a design parameter. The controller (7) implements a gradient-type iteration to solve the optimization (4), modified by replacing the true gradient $\Pi^\top\nabla\Phi(u^k, Gu^k + Hq)$ with a measurement-based version $\Pi^\top\nabla\Phi(u^k, y^k)$, which avoids the need to measure q. Unfortunately, (7) is inapplicable to our setting, since

(i) (7) does not respect the distributed nature of the controller considered here (cf. Section II-B), and
(ii) (7) requires centralized knowledge of the gradients $\{\nabla\Phi_i\}_{i\in\mathcal{V}_u}$, which is impractical in our case since each Φ_i is assumed to be known only locally.

With this motivation, 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:

$$u^{k+1}_{(i)} = \sum_{j=1}^{N} w_{ij} u^k_{(j)} - \eta \Pi^\top \nabla\Phi_i(u^k_{(i)}, y^k), \quad (8)$$

where w_{ij} are the entries of matrix W. According to this control law, each agent i updates its local state $u_{(i)}$ by performing two steps: 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 it applies $-\Pi^\top\nabla\Phi_i(u^k_{(i)}, y^k)$ to seek 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 $\nabla\Phi_i$. Note that each agent needs to know the steady-state map (i.e., G) and measure the global output feedback signal y^k. We call (8) *Feedback Distributed Gradient Descent (F-DGD) algorithm.*

Remark 3: ***(Relationship with distributed optimization algorithm)*** The algorithm (8) is inspired by the Distributed Gradient Descent (DGD) method [14]. While other approaches, such as EXTRA and Gradient Tracking [25], are valid alternative solutions, their investigation is left as the topic of future works. □

IV. Convergence Analysis and Error Bounds

In this section, we study the convergence of (8) when applied to control the system (3). In the remainder, we employ the following notations of stacked vectors: $u^k_{(1:N)} := (u^k_{(1)}, u^k_{(2)}, \ldots, u^k_{(N)}) \in \mathbb{R}^{mN}$ 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 (3) controlled by (8) reads as

$$\begin{aligned} x^{k+1} &= Ax^k + BSu^k_{(1:N)} + Eq, && (9a) \\ y^k &= Cx^k + DSu^k_{(1:N)}, \\ u^{k+1}_{(i)} &= \sum_{j\in\mathcal{N}_i} w_{ij} u^k_{(j)} - \eta\Pi^\top\nabla\Phi_i(u^k_{(i)}, y^k), i \in \mathcal{V}_u, && (9b) \end{aligned}$$

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

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

A. Asymptotic convergence

The following result shows that, under a suitable choice of the stepsize η, the state of (9) converges asymptotically.

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

$$\eta_1 = \frac{1 - 2\mu + \lambda_N(W)}{L_\Phi}, \qquad \eta_2 = \frac{\mu}{\lambda_1(P)L_h^2}, \quad (10)$$

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

with 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 (9) converge. □

Proof: Please see [26] for the complete proof. ■

Theorem 4.1 shows that under a sufficiently small choice of the stepsize η, the two sequences x^k and $u^k_{(i)}$, describing the system's and controller's states, converge asymptotically to a fixed point, respectively. Notice that this convergence claim is not straightforward, since the proposed controller (8) incorporates two simultaneous steps, a consensus one and a gradient step. As such, this update may oscillate or fail to

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

converge when η is chosen inadequately. The upper bounds η_1, η_2, η_3 depend on the various parameters of the system (3) and optimization problem (4); also, observe that the imposed bounds on μ guarantees that $\bar{\eta} > 0$. Finally, we notice that a sufficiently small choice for η also guarantees that there a choice of μ that satisfies $0 < \mu \leq 1 - \frac{(1-\lambda_N(W))+\eta L_\Phi}{2}$ (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 (8) as:

$$u^{k+1}_{(1:N)} = (W \otimes I) u^k_{(1:N)} - \eta\, \gamma(u^k_{(1:N)}, y^k), \qquad (11)$$

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 $k \in \mathbb{Z}_{\geq 0}$, (11) satisfies:

$$\|\gamma(u^k_{(1:N)}, y^k)\| \leq \sigma, \qquad (12)$$

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: Please see [26] for the complete proof. ■

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 (12) can still be proven, 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 [14], to establish a linear 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 ν_f if

$$(\nabla f(z) - \nabla f(z^*))^\mathsf{T}(z - z^*) \geq \nu_f \|z - z^*\|^2, \qquad (13)$$

for all $z \in \operatorname{dom} f$, where z^* is 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 ν_f and ∇f is Lipschitz continuous with constant L_f. Then,

$$\begin{aligned}(z - z^*)^\mathsf{T}&(\nabla f(z) - \nabla f(z^*)) \\ &\geq c_1 \|\nabla f(z) - \nabla f(z^*)\|^2 + c_2 \|z - z^*\|^2,\end{aligned} \qquad (14)$$

where z^* is as in (13). Moreover, for any $\theta \in [0, 1]$, we have $c_1 = \frac{\theta}{L_f}$ and $c_2 = (1 - \theta)\nu_f$.

Remark 4: Notice that, if f is strongly convex with modulus ν_f, then it is also restricted strong convexity with the same modulus [27]. In this case, (14) holds with $c_1 = \frac{1}{\nu_f + L_f}$ and $c_2 = \frac{\nu_f L_f}{\nu_f + L_f}$. □

The following is the second main result of this paper.

Theorem 4.4: ***(Control error bounds)*** Suppose the assumptions of Proposition (4.2) hold, $\eta \leq c_1$, and $(u, y) \mapsto \Phi(u, y)$ is restricted strongly convex with modulus ν_Φ. Then, (9) satisfies:

$$\|u^k_{(i)} - u^{*k}\| \leq c_3^k \|u^0_{(i)} - u^{*0}\| + \frac{c_4}{\sqrt{1 - c_3^2}} + \frac{\eta\sigma}{1 - \beta}, \qquad (15)$$

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

$(u^{*k}, x^{*k}) := \operatorname{Proj}_{\mathcal{A}^*}(\bar{u}^k, x^k)$, σ is as in (12), $\delta > 0$ is an arbitrary constant, and c_1 and c_2 are as in Lemma 4.3 with $\nu_f = \nu_\Phi / N$. Moreover, if $\Phi(u, y)$ is strongly convex, then c_1 and c_2 are as in Remark 4. □

Proof: We begin by proving (15). 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

$$\|u^k_{(i)} - u^{*k}\| \leq \|u^k_{(i)} - \bar{u}^k\| + \|\bar{u}^k - u^{*k}\|. \qquad (16)$$

The proof is organized into two main steps.

1) Bound for $\|u^k_{(i)} - \bar{u}^k\|$. By expanding (11) in time:

$$u^k_{(1:N)} = -\eta \sum_{s=0}^{k-1} (W^{k-1-s} \otimes I) \gamma(u^s_{(1:N)}, y^s). \qquad (17)$$

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

$$\begin{aligned}
\|u^k_{(i)} - \bar{u}^k\| &\leq \|u^k_{(1:N)} - \bar{u}^k_{(1:N)}\| \\
&= \|u^k_{(1:N)} - \frac{1}{N}((1_N 1_N^\mathsf{T}) \otimes I) u^k_{(1:N)}\| \\
&= \| - \eta \sum_{s=0}^{k-1} (W^{k-1-s} \otimes I) \gamma(u^s_{(1:N)}, y^s) \\
&\quad + \eta \sum_{s=0}^{k-1} \frac{1}{N}((1_N 1_N^\mathsf{T} W^{k-1-s}) \otimes I) \gamma(u^s_{(1:N)}, y^s)\| \\
&\overset{(a)}{=} \| - \eta \sum_{s=0}^{k-1} (W^{k-1-s} \otimes I) \gamma(u^s_{(1:N)}, y^s) \\
&\quad + \eta \sum_{s=0}^{k-1} \frac{1}{N}((1_N 1_N^\mathsf{T}) \otimes I) \gamma(u^s_{(1:N)}, y^s)\| \\
&= \eta \| \sum_{s=0}^{k-1} \left(\left(W^{k-1-s} - \frac{1}{N} 1_N 1_N^\mathsf{T} \right) \otimes I \right) \gamma(u^s_{(1:N)}, y^s)\| \\
&\leq \eta \sum_{s=0}^{k-1} \|W^{k-1-s} - \frac{1}{N} 1_N 1_N^\mathsf{T}\| \|\gamma(u^s_{(1:N)}, y^s)\| \\
&= \eta \sum_{s=0}^{k-1} \beta^{k-1-s} \|\gamma(u^s_{(1:N)}, y^s)\|,
\end{aligned} \qquad (18)$$

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

$$\|u^k_{(i)} - \bar{u}^k\| \le \eta \sum_{s=0}^{k-1} \beta^{k-1-s} \|\gamma(u^s_{(1:N)}, y^s)\| \le \sum_{s=0}^{k-1} \beta^{k-1-s}\sigma \le \frac{\eta\sigma}{1-\beta}. \tag{19}$$

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

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

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 (8) over i and noticing $W = [w_{ij}]$ is doubly stochastic, we obtain

$$\begin{aligned}\bar{u}^{k+1} &= \frac{1}{N}\sum_{i=1}^N u^{k+1}_{(i)} \\ &= \frac{1}{N}\sum_{i,j=1}^N w_{ij}u^k_{(j)} - \frac{\eta}{N}\sum_{i=1}^N \Pi^\mathsf{T}\nabla\Phi_i(u^k_{(i)}, y^k) \\ &= \bar{u}^k - \eta g(u^k_{(1:N)}, y^k). \end{aligned}\tag{20}$$

Before proceeding notice that the following bound holds:

$$\|\Pi^\mathsf{T}(\nabla\Phi_i(u^k_{(i)}, y^k) - \nabla\Phi_i(\bar{u}^k, y^k))\| \le L_\Phi \|u^k_{(i)} - \bar{u}^k\| \overset{(a)}{\le} \frac{\eta\sigma L_\Phi}{1-\beta}$$

by Assumptions 3, 2, and (6), and where (a) follows from (19). Moreover, we also have

$$\begin{aligned}&\|g(u^k_{(1:N)}, y^k) - \bar{g}(u^k_{(1:N)}, y^k)\| \\ &= \|\frac{1}{N}\sum_{i=1}^N \Pi^\mathsf{T}(\nabla\Phi_i(u^k_{(i)}, y^k) - \nabla\Phi_i(\bar{u}^k, y^k))\| \\ &\le \frac{1}{N} L_\Phi \sum_{i=1}^N \|u^k_{(i)} - \bar{u}^k\| \le \frac{\eta\sigma L_\Phi}{1-\beta}. \end{aligned}\tag{21}$$

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

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

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

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

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

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

where (a) follows from (21). 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 $c_1c_2 = \frac{\nu_\Phi L_\Phi}{(\nu_\Phi + L_\Phi)^2} < 1$. Therefore, we have $c_1 < 1/c_2$. When $\eta < c_1$, $(1+\eta\delta)(1-\eta c_2) > 0$.

Using $\|\bar{e}^k\|^2 \le (c_3^k)^2\|\bar{e}^0\|^2 + \frac{1-(c_3^k)^2}{1-c_3^2}c_4^2 \le (c_3^k)^2\|\bar{e}^0\|^2 + \frac{c_4^2}{1-c_3^2}$, we get

$$\|\bar{e}^k\| \le c_3^k\|\bar{e}^0\| + \frac{c_4}{\sqrt{1-c_3^2}}. \tag{22}$$

The claim thus follows by combining (19) and (22) ■

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 consensus (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 5: ***(Refinement of bound** (15)**)*** 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

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

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 matrices B, C, and E are randomly drawn from the normal distribution and A is chosen as a Schur

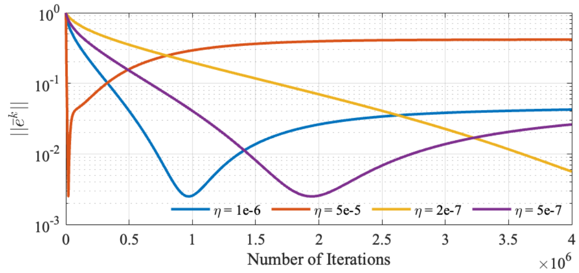

Fig. 2. Error $\bar{e}^k = \frac{1}{N}\sum_{i=1}^N \|u_{(i)}^k - u^{*k}\|$ of the proposed decentralized algorithm with different stepsizes, where $u^{*k} = \mathrm{Proj}_{\mathcal{A}^*}(\frac{1}{N}\sum_{i=1}^N u_{(i)}^k)$.

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 (9) to:

$$\underset{u\in\mathbb{R}^m}{\text{minimize}} \sum_{i=1}^{N} \frac{1}{2}\left(\|u\|_{R_i}^2 + \|Gu + Hq - y^{ref}\|_Q^2\right), \quad (23)$$

where $Q = I_p$ and $R_i = \alpha_i I_m$ are the multiplied identity matrices of the corresponding dimension. α_i is a constant, randomly chosen from $[0.001, 0.1]$ interval for each agent. We set the desired 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 constant disturbance q is generated from the standard uniform distribution.

Fig. 2 illustrates the error $\bar{e}^k$ for four 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 η causes the algorithm to converge slowly but more accurately, as predicted by Theorem 4.4.

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 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," *SIAM Journal on Optimization*, 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. Cortés, and E. Dall'Anese, "Saddle-flow dynamics for distributed feedback-based optimization," vol. 3, no. 4, pp. 948–953, 2019.

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

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

[25] 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.

[26] A. Mehrnoosh and G. Bianchin, "Distributed optimization of linear multi-agent systems via feedback-dgd," *arXiv preprint arXiv:2403.18386*, 2024.

[27] 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.