2024 IEEE 63rd Conference on Decision and Control (CDC)
December 16-19, 2024. MiCo, Milan, Italy

Extremum Seeking and Adaptive Dynamic Programming for Distributed Feedback Optimization

Tong Liu[1], Miroslav Krstić[2] and Zhong-Ping Jiang[1]

2024 IEEE 63rd Conference on Decision and Control (CDC) | 979-8-3503-1633-9/24/$31.00 ©2024 IEEE | DOI: 10.1109/CDC56724.2024.10886832

***Abstract*— This paper studies the distributed feedback optimization problem for linear multi-agent systems without precise knowledge of local costs and agent dynamics. The proposed solution is based on a hierarchical approach that uses upper-level coordinators to adjust reference signals toward the global optimum and lower-level controllers to regulate agents' outputs toward the reference signals. In the absence of precise information on local gradients and agent dynamics, an extremum-seeking mechanism is used to enforce a gradient descent optimization strategy, and an adaptive dynamic programming approach is taken to synthesize an internal-model-based optimal tracking controller. The whole procedure relies only on measurements of local costs and input-state data along agents' trajectories. Moreover, under appropriate conditions, the closed-loop signals are bounded and the output of the agents exponentially converges to a small neighborhood of the desired extremum. A numerical example is conducted to validate the efficacy of the proposed method.**

I. INTRODUCTION

Distributed Feedback Optimization (DFO) is a generalization of distributed optimization and feedback-based optimization, and seeks to achieve the optimal output agreement of multi-agent systems using real-time measurements of the agents' response [25].

Considerable studies on DFO have focused on different types of agent dynamics, such as linear systems [18], Euler-Lagrange systems [23], and nonlinear systems with certain structures [17]. It is worth noting that most of the studies rely on accurate gradient information of the cost function. However, in practical applications, obtaining precise gradient information can be challenging. For example, mobile robots can only measure the strength of a signal source without access to the exact spatial profile in an unknown signal field [4]. Extremum seeking (ES) techniques have been proposed to approximate the gradient using only measurements of the unknown cost in feedback-based optimization [13], and the applications include Nash equilibrium seeking, finite-horizon linear quadratic control, and delay compensation, to name a few [6], [16], [21], [22]. For DFO, an estimation-based ES method was proposed for unstable agent dynamics, but the output agreement objective was not considered [8], and a perturbation-based ES method was developed for multi-agent systems described as single integrators [15]. However, an ES approach to handle the unknown cost functions in the DFO design for general linear systems is still non-existent.

In addition, the exact knowledge of the agent dynamics is usually assumed for most of the DFO design. For example, all parameters of the agent dynamics should be known a priori to obtain the controllers in [1], [18]. To relax this restriction, some preliminary results have been obtained for multi-agent systems taking the parametric strict-feedback form [24], with partially linear models [28], and with unknown control direction [27]. These papers still require partial knowledge of system dynamics. On the other hand, adaptive dynamic programming (ADP) is a systematic methodology that can remove the need for accurate knowledge of system dynamics and guarantee the optimality of the controllers [11]. To the best of our knowledge, ADP has not been applied to the DFO design for general linear systems considered in this paper.

This paper investigates the DFO design for general linear multiagent systems with unknown cost functions and agent dynamics. The main contributions are summarized as follows. Firstly, a distributed reference generator is proposed in conjunction with a perturbation-based ES mechanism, using only measurements of the local cost. Secondly, a novel ADP-based optimal tracking controller is synthesized without requiring knowledge of model parameters for each agent. Finally, under these data-driven approaches, an optimal output agreement objective can be achieved with guaranteed boundedness of the closed-loop signals.

Notations. Throughout the paper, $\sigma(\cdot)$ denotes the spectrum of a matrix. $\mathcal{N} = \{1, \dots, N\}$. O_p denotes the zero matrix with size p. $X = \text{blockdiag}[X_1, X_2, \dots, X_n]$ denotes the block diagonal concatenation of the matrices $X_1, X_2, \dots, X_n$. $\otimes$ stands for the Kronecker product. I_p denotes the identity matrix with size p and p is omitted when the size is clear from the context. $[X_1; X_2]$ stands for the matrix formed by concatenating X_1 and X_2 along the rows. $(\mathbf{a}_1, \cdots, \mathbf{a}_n)$ denotes $[\mathbf{a}_1^\top, \cdots, \mathbf{a}_n^\top]^\top$ for vectors $\mathbf{a}_1, \cdots, \mathbf{a}_n$. $X \succ 0$ $(X \succeq 0)$ denotes that X is real symmetric and positive (semi-) definite. $|\cdot|$ denotes the 2-norm of a vector or matrix.

II. PROBLEM FORMULATION

This section presents the DFO problem for general linear multi-agent systems in the absence of prior knowledge on cost functions and model prameters.

Consider a multi-agent system with N heterogeneous

This work has been supported in part by the National Science Foundation under Grants CNS-2227153 and EPCN-2210320.

[1]T. Liu and Z. P. Jiang are with the Control and Networks Lab, Department of Electrical and Computer Engineering, Tandon School of Engineering, New York University, Brooklyn, NY 11201 USA (e-mails: tl3049@nyu.edu; zjiang@nyu.edu).

[2] M. Krstić is with the Department of Mechanical and Aerospace Engineering, University of California, San Diego, CA 92093-0411, USA (e-mail: krstic@ucsd.edu).

agents and agent $i \in \mathcal{N}$ is described by

$$\begin{cases} \dot{d}_i = S_i d_i \\ \dot{x}_i = A_i x_i + B_i u_i + E_i d_i \\ y_i = C_i x_i + D_i u_i + F_i d_i \end{cases} \quad (1)$$

where $x_i \in \mathbb{R}^{n_i}$, $u_i \in \mathbb{R}^{m_i}$, $d_i \in \mathbb{R}^{n_{di}}$, and $y_i \in \mathbb{R}^p$ are the state, input, external disturbance, and output, respectively. A_i, B_i, C_i, D_i, E_i, F_i, and S_i are *unknown* matrices. We impose the following assumptions on (1).

Assumption 1: The exogenous disturbance d_i is unmeasurable, the minimal polynomial of S_i is available, and all the eigenvalues of S_i are simple roots of the minimal polynomial with zero real parts.

Assumption 2: The pair (A_i, B_i) is stabilizable and for each $\lambda \in \sigma(S_i) \cup \{0\}$,

$$\text{rank}\left(\begin{bmatrix} A_i - \lambda I & B_i \\ C_i & D_i \end{bmatrix}\right) = n_i + p. \quad (2)$$

Remark 1: Assumptions 1 and 2 are standard in solving robust output regulation problems in the presence of parametric uncertainties and external disturbances [3], [5], [9], except that (2) holds when $\lambda = 0$, indicating the system capability of tracking a constant reference input [9, Remark 1.12]. This capability is further utilized for tracking control synthesis.

The communication topology among the agents is represented by a directed graph $\mathcal{G}$, which consists of N nodes, an edge set $\mathcal{E} \subseteq \mathcal{N} \times \mathcal{N}$, and a weighted adjacency matrix $\mathcal{A} = [a_{ij}]_{N \times N}$. In addition, $(i, j) \in \mathcal{E}$ is referred to as the edge from i to j and $a_{ij} > 0$ if (i, j) exists, otherwise $a_{ij} = 0$. The graph is undirected if $\mathcal{A} = \mathcal{A}^\top$. A path exists from i_1 to i_k if there exists a sequence of edges $(i_1, i_2), \ldots, (i_{k-1}, i_k)$. The graph is strongly connected if there exists a path between any two nodes. The adjacency matrix $\mathcal{A}$ is doubly stochastic if $\mathcal{A}\mathbf{1}_N = \mathbf{1}_N$ and $\mathbf{1}_N^\top \mathcal{A} = \mathbf{1}_N^\top$, where $\mathbf{1}_N = [1, \ldots, 1]^\top \in \mathbb{R}^N$. The following assumption is made on the graph.

Assumption 3: $\mathcal{G}$ is undirected, strongly connected, and $\mathcal{A}$ is doubly stochastic.

A local cost function $f_i : \mathbb{R}^p \to \mathbb{R}$ is *unknown*, but each agent i can access the *measurement* of $f_i(\cdot)$ on its output. The objective is to regulate each agent's output y_i toward a common value y^* that minimizes a global cost function $f = \sum_{i=1}^N f_i$. We impose the following assumption on the cost functions.

Assumption 4: For each $i \in \mathcal{N}$, f_i is twice continuously differentiable and there exists a positive constant μ such that

$$(b_1 - b_2)^\top (\nabla f(b_1) - \nabla f(b_2)) \geq \mu |b_1 - b_2|^2 \quad (3)$$

for all $b_1, b_2 \in \mathbb{R}^p$.

By (3), y^* is unique. Since the gradient $\nabla f(\cdot)$ is not available, exploration noise is used to obtain its approximation. Due to the existence of the exploration noise, achieving exact convergence of y_i to y^* is challenging, and a less ambitious objective is pursued here in the sense of semi-global practical stabilization. Namely, for the multi-agent system (1) with initial conditions in a prescribed compact set and for any given number $\varepsilon > 0$, assuming Assumptions 1 to 4 hold, design a distributed optimizer, that gives r_i, and a decentralized tracking controller u_i such that the closed-loop signals are bounded and

$$\limsup_{t \to \infty} |y_i(t) - y^*| < \varepsilon \quad (4)$$

for each $i \in \mathcal{N}$.

III. Data-driven Approaches

This section proposes hierarchical data-driven approaches to address the DFO problem. First, a distributed reference generator is pursued in conjunction with an ES mechanism for optimum searching. Then, an ADP-based optimal tracking controller is developed to regulate the outputs toward the designed reference signals.

A. Distributed Reference Generator

We employ the following algorithm to search for the minimizer y^*: for $i \in \mathcal{N}$,

$$\begin{aligned} \dot{r}_i &= \sum_{j \in \mathcal{N}_i} a_{ij}(r_j - r_i + q_j - q_i) - g_i(y_i) \\ \dot{q}_i &= \sum_{j \in \mathcal{N}_i} a_{ij}(r_i - r_j) \end{aligned} \quad (5)$$

where $\mathcal{N}_i$ denotes the neighbor of agent i, r_i is the reference signal, q_i is an auxiliary variable, and $g_i(y_i)$ is an estimate of the gradient. Equation (5) is revised from [30], and in the original setting, $g_i(y_i) = \nabla f_i(r_i)$ and $r_i(t)$ converges to y^* as t tends to infinity. Moreover, the convergence result still holds when $g_i(y_i) = \alpha \nabla f_i(r_i)$ where α is a small enough positive constant. This result stems from a time-scale separation property, and we are inspired to use the following ES mechanism: for $j = \{1, \ldots, p\}$,

$$g_{ij}(y_i) = a\delta w f_i(y_i) \sin(w_{ij} t) \quad (6)$$

where a, δ, and w are parameters to be designed, and $w_{ij} = w\bar{w}_{ij} \in \mathbb{R}$ with $\bar{w}_{ij} \in \mathbb{R}$ and $\bar{w}_{ij} \neq \bar{w}_{ik}$ if $j \neq k$. Let $g_i(y_i) = (g_{i1}(y_i), \ldots, g_{ip}(y_i))$. In addition, for $i \in \mathcal{N}$, a perturbed reference y_i^r is defined as

$$y_i^r = r_i + a \sin \mathbf{w}_i t \quad (7)$$

where $\mathbf{w}_i = (w_{i1}, \ldots, w_{ip})$, and $\sin \mathbf{w}_i t = (\sin w_{i1} t, \ldots, \sin w_{ip} t)$. Equations (6) and (7) are developed based on [13], [26], in which sinusoidal signals are exploited to explore the gradient information. This mechanism is effective assuming $y_i(t) = y_i^r(t)$ for $t \geq 0$. To approximate this condition, we subsequently introduce a tracking controller to steer y_i towards y_i^r.

B. Optimal Tracking Control

This section develops an optimal tracking controller for each agent without prior knowledge of the parameters in (1). Treating y_i^r as a constant, we design a reference tracking controller such that y_i converges to y_i^r asymptotically [19].

We employ an internal-model-based control design [3], [9] as follows: for $i \in \mathcal{N}$, let $\lambda^{\kappa_i} + b_1 \lambda^{(\kappa_i - 1)} + \cdots + b_{\kappa_i - 1}\lambda$ be the minimal polynomial of $A_{i0} = \text{blockdiag}[O_p, S_i]$. Then, a controllable pair (Φ_i, σ_i) can be defined as

$$\Phi_i = \left[\begin{array}{c|c} 0 & I_{\kappa_i - 1} \\ \hline 0 & [-b_{(\kappa_i-1)}, \ldots, -b_1] \end{array}\right]$$

and $\sigma_i = (0, \ldots, 0, 1) \in \mathbb{R}^{\kappa_i}$. Then, $G_{i1} = \text{blockdiag}[\Phi_i, \ldots, \Phi_i]$ and $G_{i2} = \text{blockdiag}[\sigma_i, \ldots, \sigma_i]$, which contain p copies of Φ_i and σ_i, respectively. The dynamic state feedback controller is described as

$$\begin{aligned} \dot{z}_i &= G_{i1} z_i + G_{i2} e_i \\ u_i &= K_i x_{ic} \end{aligned} \tag{8}$$

where $e_i = y_i - y_i^r$ is the tracking error, $x_{ic} = (x_i, z_i)$, and $K_i = [K_{i1}, K_{i2}]$ with $K_{i1} \in \mathbb{R}^{m_i \times n_i}$ is chosen such that $A_{ic} = A_{io} + B_{io} K_i$ is Hurwitz with

$$A_{io} = \begin{bmatrix} A_i & 0 \\ G_{i2} C_i & G_{i1} \end{bmatrix}, \quad B_{io} = \begin{bmatrix} B_i \\ G_{i2} D_i \end{bmatrix}.$$

By Assumption 2, the pair (A_{io}, B_{io}) is stabilizable [9, Lem. 1.26], which ensures the existence of K_i. Then, K_i can be determined by solving the following linear quadratic regulator problem:

$$\begin{aligned} &\min_{u_i} \int_0^\infty (x_{ic}^\top Q_i x_{ic} + u_i^\top R_i u_i) dt \\ &\text{s.t. } \dot{x}_{ic} = A_{io} x_{ic} + B_{io} u_i \end{aligned} \tag{9}$$

where $Q_i, R_i \succ 0$. One solution P_i^* of the following algebraic Riccati equation

$$A_{io}^\top P_i^* + P_i^* A_{io} + Q_i - P_i^* B_{io} R_i^{-1} B_{io}^\top P_i^* = 0 \tag{10}$$

defines a stabilizing control law $u_i = -K_i^* x_{ic}$ with $K_i^* = R_i^{-1} B_{io}^\top P_i^*$, which is also the optimal solution of (9). We have the following lemma on P_i^*.

Lemma 1 ([14]): For the differential Riccati equation

$$\dot{P}_i = A_{io}^\top P_i + P_i A_{io} + Q_i - P_i B_{io} R_i^{-1} B_{io}^\top P_i \tag{11}$$

with $P_i(t_0) \succeq 0$, $\lim_{t \to \infty} P_i(t) = P_i^*$.

Lemma 1 provides a way of solving P_i^* using (11), but this process relies on exact knowledge of A_i, B_i, C_i, and D_i. This restriction can be relaxed by the ADP techniques [2]. Since the minimal polynomial of S_i is available, a matrix $\hat{S}_i$ exists such that

$$\dot{v}_i = \hat{S}_i v_i, \quad d_i = L_i v_i \tag{12}$$

where v_i is a measurable signal and L_i is an unknown matrix (see [7, Remark 1]). Substituting (8) and (12) into (1) gives

$$\dot{x}_{ic} = A_{io} x_{ic} + B_{io} u_i + M_i v_i + \hat{y}_i^r \tag{13}$$

where $M_i = [E_i L_i; G_{i2} F_i L_i]$ and $\hat{y}_i^r = [0; -G_{i2}] y_i^r$. For any $P_{id} \succeq 0$, along the solution of (13),

$$\begin{aligned} \frac{d}{dt}(x_{ic}^\top P_{id} x_{ic}) &= 2(A_{io} x_{ic} + B_{io} u_i + M_i v_i + \hat{y}_i^r)^\top P_{id} x_{ic} \\ &= x_{ic}^\top Z_i x_{ic} + 2 u_i^\top R_i K_{id} x_{ic} \\ &\quad + 2 v_i^\top M_i^\top P_{id} x_{ic} + 2 \hat{y}_i^{r\top} P_{id} x_{ic} \end{aligned}$$

where $Z_i = A_{io}^\top P_{id} + P_{id} A_{io}$ and $K_{id} = R_i^{-1} B_{io}^\top P_{id}$, and the integration of the equation from T_k to T_{k+1} is

$$\begin{aligned} x_{ic}^\top P_{id} x_{ic}\big|_{T_k}^{T_{k+1}} &= \int_{T_k}^{T_{k+1}} x_{ic}^\top \otimes x_{ic}^\top \, dt \, \text{vec}(Z_i) \\ &\quad + 2 \int_{T_k}^{T_{k+1}} x_{ic}^\top \otimes (R_i u_i)^\top \, dt \, \text{vec}(K_{id}) \\ &\quad + 2 \int_{T_k}^{T_{k+1}} x_{ic}^\top \otimes v_i^\top \, dt \, \text{vec}(M_i^\top P_{id}) \\ &\quad + 2 \int_{T_k}^{T_{k+1}} x_{ic}^\top \otimes \hat{y}_i^{r\top} \, dt \, \text{vec}(P_{id}) \end{aligned} \tag{14}$$

where $\text{vec}(\cdot)$ denotes vectorization of a matrix. For a sequence of times $T_0 < T_1 < \cdots < T_s$, by (14),

$$\Theta_i \begin{bmatrix} \bar{Z}_i \\ \text{vec}(K_{id}) \\ \text{vec}(M_i^\top P_{id}) \end{bmatrix} = D_{x_{ic} x_{ic}} \bar{P}_{id} - 2 I_{x_{ic} \hat{y}_i^r} \text{vec}(P_{id}) \tag{15}$$

where $\Theta_i = [I_{\bar{x}_{ic}}, 2 I_{x_{ic} R_i u_i}, 2 I_{x_{ic} v_i}]$ and related definitions can be found in [20]. The following assumption is imposed to ensure that the solution of (15) is unique.

Assumption 5: Θ_i is of full column rank.

To guarantee Assumption 5, one can inject an exploration noise into u_i or y_i^r when collecting the online data [2], [11]. Under Assumption 5, for any given P_{id}, Z_i and K_{id} can be solved from (15) and used to compute $h_{id}(P_{id}) = Z_i + Q_i - K_{id}^\top R_i K_{id}$. Thus, we can define and solve a differential equation $\dot{P}_{id} = h_{id}(P_{id})$ to get P_i^*.

Proposition 1: Under Assumption 5, for the differential equation $\dot{P}_{id} = h_{id}(P_{id})$ with $P_{id}(t_0) \succeq 0$, $\lim_{t \to \infty} P_{id}(t) = P_i^*$ and $\lim_{t \to \infty} K_{id}(t) = K_i^*$.

Remark 2: Since solving $\dot{P}_{id} = h_{id}(P_{id})$ with $P_{id}(t_0) \succeq 0$ does not explicitly depend on the unknown matrices, but on the measurements of $x_{ic}(t)$, $u_i(t)$, $v_i(t)$, and $y_i^r(t)$ for some time, this process is non-model-based. A policy iteration algorithm can also be developed under appropriate conditions (see [20, Sec. IV-A]).

Remark 3: To summarize, the entire procedure is as follows: first, each agent i collects the online data to compute K_i of (8) by the ADP algorithm, then, each agent i uses (8) and (7) governed by (5) and (6) to execute the ES process.

IV. Closed-Loop Analysis

This section analyzes the stability properties of the closed-loop system using the distributed reference generators and the proposed tracking controllers. We first apply coordinate transformations to formulate a two-time-scale model. Then, we derive the stability result of the closed-loop system.

A. Coordinate Transformation

We present coordinate transformations to analyze the reference signals and the tracking error dynamics. A matrix $U \in \mathbb{R}^{N \times (N-1)}$ exists such that $T = [\mathbf{1}_N / \sqrt{N}, U]$ is orthonormal. Then, let $T_m = T \otimes I_p = [T_a, T_d]$ with $T_a \in \mathbb{R}^{Np \times p}$. Denote $r = (r_1, \ldots, r_N)$ and $q = (q_1, \ldots, q_N)$. We perform the coordination transformation $r = T_a\xi_a + T_d\xi_d$ and $q = T_a\phi_a + T_d\phi_d$ to obtain

$$\dot{\xi}_a = -T_a^\top g(y) \tag{16a}$$

$$\begin{bmatrix}\dot{\xi}_d \\ \dot{\phi}_d\end{bmatrix} = A_d \begin{bmatrix}\xi_d \\ \phi_d\end{bmatrix} - \begin{bmatrix}T_d^\top \\ 0\end{bmatrix} g(y) \tag{16b}$$

$$\dot{\phi}_a = 0 \tag{16c}$$

where $y = (y_1, \ldots, y_N)$, $g(y) = (g_1(y_1), \ldots, g_N(y_N))$, and $A_d = \begin{bmatrix}-U_d & -U_d \\ U_d & 0\end{bmatrix}$ with $U_d = (U^\top(I - \mathcal{A})U) \otimes I_p$. Moreover, the following lemma is derived.

Lemma 2: Under Assumption 3, A_d is Hurwitz.

Then, we analyze the behavior of the tracking error e_i. For $i \in \mathcal{N}$, assume K_i has been chosen such that $A_{ic} = A_{io} + B_{io}K_i$ is Hurwitz. By [9, Lem. 1.27], a matrix X_{ic} exists such that

$$\begin{aligned} X_{ic}A_{i0} - A_{ic}X_{ic} &= E_{ic} \\ C_{ic}X_c + \hat{F}_i &= 0 \end{aligned} \tag{17}$$

where $C_{ic} = [\hat{C}_i, \hat{D}_i]$ with $\hat{C}_i = C_i + D_iK_{i1}$ and $\hat{D}_i = D_iK_{i2}$, $\hat{F}_i = [-I_p, F_i]$, and $E_{ic} = [\hat{E}_i; \hat{F}_i]$ with $\hat{E}_i = [0, E_i]$. Let $\tilde{x}_{ic} = x_{ic} - X_{ic}(y_i^r, d_i)$. Then, using this transformation and (17) yields $\dot{\tilde{x}}_{ic} = A_{ic}\tilde{x}_{ic} - H_{ir}\dot{y}_i^r$, $e_i = C_{ic}\tilde{x}_{ic}$ where H_{ir} consists of first p columns of X_{ic}. Stacking all the above systems together gives

$$\dot{\tilde{x}}_c = A_c\tilde{x}_c - H_r\dot{y}^r \tag{18a}$$

$$e = C\tilde{x}_c. \tag{18b}$$

where $\tilde{x}_c = (\tilde{x}_{1c}, \ldots, \tilde{x}_{Nc})$, $y^r = (y_1^r, \ldots, y_N^r)$, $e = (e_1, \ldots, e_N)$, $A_c = \text{blockdiag}[A_{1c}, \ldots, A_{Nc}]$, $H_r = \text{blockdiag}[H_{1r}, \ldots, H_{Nr}]$, and $C = \text{blockdiag}[C_{1c}, \ldots, C_{Nc}]$. By combining (16a), (16b), and (18a), and letting $\varphi = (\xi_d, \phi_d, \tilde{x}_c)$, the closed-loop system reads

$$\dot{\xi}_a = -T_a^\top g(y) \tag{19a}$$

$$\dot{\varphi} = A_s\varphi + B_s g(y) + aw\vartheta \tag{19b}$$

where $A_s = \begin{bmatrix}A_d & 0 \\ H_d & A_c\end{bmatrix}$ with $H_d = [H_rT_dU_d, H_rT_dU_d]$, $B_s = [B_d; H_r]$ with $B_d = [-T_d^\top; 0]$, and $\vartheta = [0; 0; (\bar{w}_{11}\cos w_{11}t, \ldots, \bar{w}_{1p}\cos w_{1p}t, \ldots, \bar{w}_{Np}\cos w_{Np}t)]$. When w is small, (19a) and (19b) present two time scales [12, Sec. 11.2], where the "boundary-layer" system $\dot{\varphi} = A_s\varphi$ is exponentially stable at the origin. Hence, φ quickly "dies out", which gives rise to a reduced system:

$$\frac{d\dot{\xi}_a^r}{d\tau} = -a\delta\frac{1}{\sqrt{N}}\sum_{i=1}^N f_i\Big(\frac{1}{\sqrt{N}}\xi_a^r + a\sin\bar{\mathbf{w}}_i\tau\Big)\sin\bar{\mathbf{w}}_i\tau \tag{20}$$

where ξ_a^r is the reduced state, $\tau = wt$, and $\bar{\mathbf{w}}_i = (\bar{w}_{i1}, \ldots, \bar{w}_{ip})$ for $i \in \mathcal{N}$.

Equations (19a) and (20) can be regarded as perturbed gradient flows. In fact, let $\xi_a^* = \sqrt{N}y^*$ and $\tilde{\xi}_a = \xi_a - \xi_a^*$, and inspired by [12, Sec. 10.4] and [26, Appendix A], we can perform the change of variables:

$$\tilde{\xi}_a = \psi - a\delta\Psi(\tau, \psi) \tag{21}$$

where $\Psi(\tau, \psi) = \int_0^\tau \frac{1}{\sqrt{N}}\rho(s, \psi)\, ds$ with

$$\rho(s, \psi) = \Big(\sum_{i=1}^N \Big[f_i\Big(\frac{1}{\sqrt{N}}(\psi + \xi_a^*)\Big) + a\nabla f_i\Big(\frac{1}{\sqrt{N}}(\psi + \xi_a^*)\Big)^\top \sin\bar{\mathbf{w}}_i s\Big]\sin\bar{\mathbf{w}}_i s\Big) - \frac{a}{2}\nabla f\Big(\frac{1}{\sqrt{N}}(\psi + \xi_a^*)\Big).$$

Let $\psi \in \Delta_0$ with Δ_0 denoting a compact set. Since Ψ and $\partial\Psi/\partial\psi$ are periodic functions over τ, they are bounded for $(\tau, \psi) \in \mathbb{R} \times \Delta_0$. Then, a constant c_0 exists such that when $a\delta < c_0$, $I - a\delta(\partial\Psi/\partial\psi)$ is strictly diagonally dominant, implying that ψ is uniquely determined and continuously differentiable. Therefore, by (21),

$$\frac{d\tilde{\xi}_a}{d\tau} = \frac{d\psi}{d\tau} - a\delta\Big(\frac{\partial\Psi}{\partial\tau} + \frac{\partial\Psi}{\partial\psi}\frac{d\psi}{d\tau}\Big). \tag{22}$$

Let $O(\cdot)$ denote the order of magnitude [26, Sec. 2]. By the mean value theorem, $f_i(\frac{1}{\sqrt{N}}(\psi + \xi_a^*)) - f_i(\frac{1}{\sqrt{N}}(\psi - a\delta\Psi + \xi_a^*)) = O(a\delta)$, $\nabla f_i(\frac{1}{\sqrt{N}}(\psi + \xi_a^*)) - \nabla f_i(\frac{1}{\sqrt{N}}(\psi - a\delta\Psi + \xi_a^*)) = O(a\delta)$, and $(I - a\delta(\partial\Psi/\partial\psi))^{-1} = I + O(a\delta)$. Also, there exists $c_1 > 0$ such that $\frac{1}{\sqrt{N}}|\sum_{i=1}^N[f_i(y_i) - f_i(\frac{1}{\sqrt{N}}\xi_a + a\sin\bar{\mathbf{w}}_i\tau)]| \le c_1|\varphi|$. Using these relations and (22), we obtain

$$\frac{d\psi}{d\tau} = -\frac{a^2\delta}{2\sqrt{N}}\nabla f\Big(\frac{1}{\sqrt{N}}(\psi + \xi_a^*)\Big) + a\delta\eta(\tau, \psi, \varphi) \tag{23}$$

where $\eta = \eta_1(\tau, \psi) + \eta_2(\tau, \psi, \varphi)$ with $\eta_1 = O(a\delta + a^2\delta + a^2)$ and $|\eta_2| \le c_1|\varphi|$. Indeed, (23) denotes a perturbed gradient flow since $\frac{d\psi}{d\tau} = -\frac{a^2\delta}{2\sqrt{N}}\nabla f(\frac{1}{\sqrt{N}}(\psi + \xi_a^*))$ denotes a gradient flow that is exponentially stable at the origin.

B. Stability Analysis

We first present the stability result of the system under the coordinate transformation. Let $\zeta = (\psi, \varphi)$ and notice that $\dim\zeta = (2N - 1)p + \sum_{i=1}^N(n_i + p\kappa_i)$.

Proposition 2: For $i \in \mathcal{N}$, assume Assumptions 1 to 4 hold and A_{ic} is Hurwitz. Given any real numbers $\Delta_2 > \Delta_1 > 0$, there exist positive constants δ^*, a^* such that for every $\delta \in (0, \delta^*)$ and $a \in (0, a^*)$, a positive constant w^* exists such that when $w \in (0, w^*)$, there exist positive constants β_0, β_1 such that, for all $t_0 \ge 0$ and $|\zeta(t_0)| \le \Delta_2$, the solution of (23) and (19b) exists and satisfies:

$$|\zeta(t)| \le \beta_0|\zeta(t_0)|e^{-\beta_1(t - t_0)} + \Delta_1, \ \forall\, t \ge t_0. \tag{24}$$

Proof: The fact that A_{ic} is Hurwitz along with Lemma 2 implies that A_s is Hurwitz. It follows that for the boundary-layer system $\dot{\varphi} = A_s\varphi$, $|\varphi(t)| \le |\varphi(t_0)|e^{-\lambda_1(t - t_0))}$, $\forall t \ge t_0$ with $\varphi(t_0) \in \mathbb{R}^{\dim\zeta - p}$ for some $\lambda_1 > 0$. Similarly, for the reduced system (23) with $\varphi = 0$, let $V_r(\psi) = \psi^\top\psi$ and it follows that $\dot{V}_r(\psi)/w \le -\frac{a^2\delta\mu}{N}|\psi|^2 + 2a\delta|\psi||\eta_1|$, where we used (3). Furthermore, let $\Delta_0 = \{\psi||\psi| \le \Delta_2 + \Delta_1\}$, then a positive constant c_2 exists such that $|\eta_1| \le c_2(a\delta + a^2\delta + a^2)$ for $(\tau, \psi) \in \mathbb{R} \times \Delta_0$. By choosing $\delta^* > 0, a^* > 0$ such that $\delta^* + a^*\delta^* + a^* \le \frac{(1-\varepsilon_1)\mu\Delta_1}{6Nc_2}$ with $\varepsilon_1 \in (0, 1)$ and $a^*\delta^* < c_0$, we obtain that when $|\psi| \ge \Delta_1/3$, $\dot{V}_r \le -\frac{\varepsilon_1\mu a^2\delta w}{N}V_r$. Therefore, the solution of the reduced system satisfies $|\psi(t)| \le \max\{e^{-\theta_1 a^2\delta w(t - t_0)}|\psi(t_0)|, \Delta_1/3\}$ with $\theta_1 = \frac{\varepsilon_1\mu}{2N}$, $\forall t \ge t_0, |\psi(t_0)| \le \Delta_2$.

Moreover, the solutions of (19b) and (23) are close to the solutions of the boundary-layer system and the reduced system as $w \to 0$ on compact time intervals [12, Th. 11.1]. By [29, Th. 1], there exists a positive constant w^* with $w^* \leq \frac{\lambda_1}{\theta_1 a^2 \delta}$ such that when $w \in (0, w^*)$, for all $t_0 \geq 0$, $|\varphi(t_0)| \leq \Delta_2$, and $|\psi(t_0)| \leq \Delta_2$, the solution of (19b) and (23) exists and satisfies: $\forall\, t \geq t_0$, $|\varphi(t)| \leq |\varphi(t_0)| e^{-\lambda_1(t-t_0))} + \Delta_1/3$, $|\psi(t)| \leq e^{-\theta_1 a^2 \delta w(t-t_0)}|\psi(t_0)| + 2\Delta_1/3$. Therefore, using these estimates and the upper bound of w^* gives (24) by letting $\beta_0 = 2$ and $\beta_1 = \theta_1 a^2 \delta w$. ■

We next present the stability result of the original system. Let $\tilde{r} = r - \mathbf{1}_N \otimes y^*$, $\tilde{q} = q - T_a \phi_a$, and $\tilde{x}_{cl} = (\tilde{x}_{cl}^1, \ldots, \tilde{x}_{cl}^N)$ with $\tilde{x}_{cl}^i = x_{ic} - X_{ic}(y^*, d_i)$ for $i \in \mathcal{N}$. Note that $\phi_a(t) = \phi_a(t_0)$ for all $t \geq t_0$ by (16c).

Theorem 1: For $i \in \mathcal{N}$, under Assumptions 1 to 4, assume each agent i modeled by (1) uses the control law (8) where K_i is chosen such that A_{ic} is Hurwitz, and the perturbed reference signal (7) governed by (5) and (6). Given any positive real numbers Δ_1^*, Δ_2^*, and ε, there exist positive constants δ_1^*, a_1^* such that for every $\delta \in (0, \delta_1^*)$ and $a \in (0, a_1^*)$, a positive constant w_1^* exists such that when $w \in (0, w_1^*)$, there exist positive constants α_0, α_1 such that, for all $t_0 \geq 0$ and $|(x_i(t_0), z_i(t_0), d_i(t_0), r_i(t_0), q_i(t_0))| \leq \Delta_2^*$ with $i \in \mathcal{N}$, the solution of (1) and (5) exists and satisfies:

$$|\chi(t)| \leq \alpha_0 |\chi(t_0)| e^{-\alpha_1(t-t_0)} + \Delta_1^*, \forall\, t \geq t_0, \tag{25}$$

where $\chi = (\tilde{r}, \tilde{q}, \tilde{x}_{cl})$, and (4) is satisfied.

Proof: By the change of variables: $r = T_a \xi_a + T_d \xi_d$, $q = T_a \phi_a + T_d \phi_d$, $\tilde{x}_{ic} = x_{ic} - X_{ic}(y_i^r, d_i)$, (21), and (7), we obtain $\psi = T_a^\top \tilde{r} + a\delta\Psi$, $\xi_d = T_d^\top \tilde{r}$, $\phi_d = T_d^\top \tilde{q}$, and $\tilde{x}_c = \tilde{x}_{cl} - H_r(\tilde{r} + a\sin\bar{\mathbf{w}}\tau)$. It follows that $|\psi| \leq |\tilde{r}| + a\delta|\Psi|$, $|\xi_d| \leq |\tilde{r}|$, $|\phi_d| \leq |\tilde{q}|$, and $|\tilde{x}_c| \leq |\tilde{x}_{cl}| + |H_r|(|\tilde{r}| + a\sqrt{Np})$. Therefore, for the prescribed Δ_2^*, there exist positive constants c_0^*, Δ_2 such that when $a\delta \leq c_0^*$, $|\zeta(t_0)| \leq \Delta_2$. For any $0 < \Delta_1 < \Delta_2$, by Proposition 2, we can generate the tuple (a^*, δ^*, w^*) as an upper bound of (a, δ, w) such that the solution of (19b) and (23) exists and satisfies (24). Moreover, using the results of Proposition 2, the boundedness of Ψ, and the change of variables, we can obtain (25) and (4) by choosing $\alpha_1 = \beta_1$, appropriate α_0, Δ_1, and small enough a_1^*, δ_1^*, w_1^*. ■

Remark 4: We have solved a semi-global practical stabilization problem since Δ_2^* can be made arbitrarily large and Δ_1^* can be made arbitrarily small [10, Sec. 12.1]. By (7) and (25), the closed-loop signals are bounded and the output y exponentially converges to a neighborhood of the desired extremum $\mathbf{1}_N \otimes y^*$. Given Δ_1^*, Δ_2^*, and ε, the control objective can be achieved by first reducing a, δ sufficiently, and then reducing w sufficiently for fixed a, δ. On the other hand, a smaller ε or Δ_1^*, or a larger Δ_2^* requires smaller a, δ, w, which reduces the speed of convergence as $\alpha_1 = \theta_1 a^2 \delta w$. This trade-off aligns with the result in [26].

V. Simulation Results

We consider a rendezvous problem where four UGVs cooperate to gather at an optimal position. The communication topology of the UGVs is described by an undirected grid graph and each non-zero component of the adjacency matrix is 0.5. Let $y_i = p_i \in \mathbb{R}^2$ denote the output (position) of agent i, and the following problem is considered:

$$\min_{p_1, \cdots, p_N} \sum_{i=1}^{4} |p_i - p_i(0)|^2 \quad \text{s.t.} \quad p_i = p_j \text{ for } i, j \in \mathcal{N}$$

where $p_1(0) = (0,0)$, $p_2(0) = (0,2)$, $p_3(0) = (2,2)$, and $p_4(0) = (2,0)$ are the initial positions. The optimal position can be computed at $y^* = (p_x^*, p_y^*) = (1,1)$, but it is unknown to all the agents. The dynamics of the UGV is described by

$$\dot{x}_i = \begin{bmatrix} 0 & I_2 \\ 0 & -\epsilon_i I_2 \end{bmatrix} x_i + \begin{bmatrix} 0 \\ \epsilon_i I_2 \end{bmatrix} u_i + E_i d_i \tag{26a}$$

$$y_i = \begin{bmatrix} I_2 & 0 \end{bmatrix} x_i \tag{26b}$$

where $x_i \in \mathbb{R}^4$, $d_i \in \mathbb{R}^2$ is generated by $\dot{d}_i = S_i d_i$ with $S_i = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}$, and $E_i = [I_2; O_2]$, for $i \in \mathcal{N}$. In addition, $\epsilon_i = 1.25$ if $i \in \{1,2\}$ and $\epsilon_i = 1.33$ if $i \in \{3,4\}$. Model parameters in (26a) and (26b) are unknown to the learning algorithms. It can be checked that Assumptions 1 to 4 hold for $i \in \mathcal{N}$.

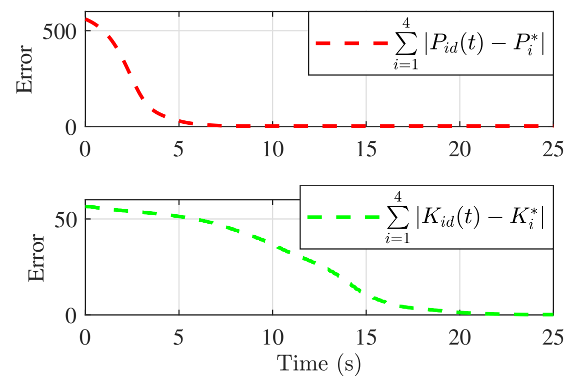

Fig. 1. Convergence of $P_{id}(t)$ and $K_{id}(t)$.

In the first stage, we apply the ADP method in Section III-B to learn an optimal tracking controller for each agent. For $i \in \mathcal{N}$, let $Q_i = 10I$, $R_i = I$, and $y_i^r = 0$, and we use a summation of sinusoidal signals as the control input to collect the data, and $\hat{S}_i = S_i^\top$ to produce the fictitious signal v_i. Let $d_1(0) = (1, 0.5)$, $d_2(0) = (1,1)$, $d_3(0) = (-1,-1)$, and $d_4(0) = (-1, 1)$. The other initial conditions are zeros except the initial positions. Consequently, Assumption 5 is satisfied by collecting the data for 2.8s with a sampling time of 0.01s. We compute P_i^* and K_i^* for comparison, and the convergence of P_{id} to P_i^* and K_{id} to K_i^* is achieved for $i \in \mathcal{N}$. The convergence result is shown in Fig. 1. Notice that it takes less than 0.2s to solve the differential equation $\dot{P}_{id} = h_{id}(P_{id})$ with $P_{id}(0) = 0$ to get the approximate optimal controller for $i \in \mathcal{N}$.

In the second stage, we use the distributed reference generators and the obtained tracking controllers to achieve the optimal output agreement objective. We set $a = 0.1$,

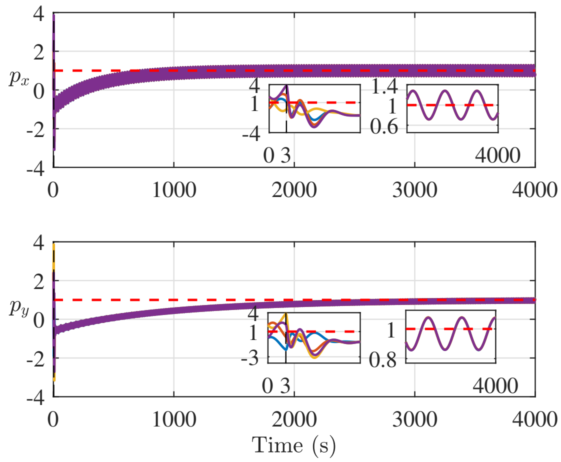

Fig. 2. Output trajectories of the UGV systems.

$\delta = w = 0.3$, and $\bar{w}_{i1} = 1, \bar{w}_{i2} = 1.5$ for $i \in \mathcal{N}$. The output trajectories of all the agents are shown in Fig. 2. When $t \leq 3$s, each agent collects the online data and computes the tracking control law (8). After that, each agent uses (8) and (7) governed by (5) and (6) to search the extremum, and the output of each agent converges to a neighborhood of y^* with $\varepsilon \leq 0.32$.

VI. Conclusions

This paper addresses the distributed feedback optimization problem for linear multi-agent systems without prior knowledge of cost functions and model parameters. We introduce an extremum-seeking mechanism for distributed reference signal design and an ADP-based method for internal-model-based optimal tracking control. Moreover, the proposed approach has achieved the semi-global practical stabilization, which guarantees the exponential convergence of the agents' output to a small neighborhood of the desired extremum. A numerical example of UGV systems illustrates the effectiveness of the proposed method. Our future work will consider the extension to nonlinear multi-agent systems.

References

[1] L. An and G.-H. Yang, "Distributed optimal coordination for heterogeneous linear multiagent systems," *IEEE Transactions on Automatic Control*, vol. 67, no. 12, pp. 6850–6857, 2021.

[2] T. Bian and Z. P. Jiang, "Value iteration and adaptive dynamic programming for data-driven adaptive optimal control design," *Automatica*, vol. 71, pp. 348–360, 2016.

[3] E. Davison, "The robust control of a servomechanism problem for linear time-invariant multivariable systems," *IEEE Transactions on Automatic Control*, vol. 21, no. 1, pp. 25–34, 1976.

[4] H. Eqab, Y. B. Salamah, I. Ahmad, and M. Morsy, "Development of source seeking algorithm for mobile robots," *Intelligent Service Robotics*, vol. 16, no. 3, pp. 393–401, 2023.

[5] B. A. Francis and W. M. Wonham, "The internal model principle for linear multivariable regulators," *Applied Mathematics and Optimization*, vol. 2, no. 2, pp. 170–194, 1975.

[6] P. Frihauf, M. Krstić, and T. Başar, "Nash equilibrium seeking in noncooperative games," *IEEE Transactions on Automatic Control*, vol. 57, no. 5, pp. 1192–1207, 2011.

[7] W. Gao and Z. P. Jiang, "Adaptive dynamic programming and adaptive optimal output regulation of linear systems," *IEEE Transactions on Automatic Control*, vol. 61, no. 12, pp. 4164–4169, 2016.

[8] M. Guay, I. Vandermeulen, S. Dougherty, and P. J. McLellan, "Distributed extremum-seeking control over networks of dynamically coupled unstable dynamic agents," *Automatica*, vol. 93, pp. 498–509, 2018.

[9] J. Huang, *Nonlinear Output Regulation: Theory and Applications*. Philadelphia, PA, USA: SIAM, 2004.

[10] A. Isidori, *Nonlinear Control Systems*. London, U.K.: Springer-Verlag, 1999, vol. II.

[11] Z. P. Jiang, T. Bian, and W. Gao, "Learning-based control: A tutorial and some recent results," *Foundations and Trends® in Systems and Control*, vol. 8, no. 3, pp. 176–284, 2020.

[12] H. K. Khalil, *Nonlinear Systems*. Englewood Cliffs, NJ, USA: Prentice-Hall, 2002.

[13] M. Krstić and H.-H. Wang, "Stability of extremum seeking feedback for general nonlinear dynamic systems," *Automatica*, vol. 36, no. 4, pp. 595–601, 2000.

[14] V. Kučera, "A review of the matrix Riccati equation," *Kybernetika*, vol. 9, no. 1, pp. 42–61, 1973.

[15] Z. Li, K. You, and S. Song, "Cooperative source seeking via networked multi-vehicle systems," *Automatica*, vol. 115, p. 108853, 2020.

[16] S.-J. Liu, M. Krstić, and T. Başar, "Batch-to-batch finite-horizon LQ control for unknown discrete-time linear systems via stochastic extremum seeking," *IEEE Transactions on Automatic Control*, vol. 62, no. 8, pp. 4116–4123, 2016.

[17] T. Liu, Z. Qin, Y. Hong, and Z. P. Jiang, "Distributed optimization of nonlinear multiagent systems: a small-gain approach," *IEEE Transactions on Automatic Control*, vol. 67, no. 2, pp. 676–691, 2021.

[18] T. Liu, Z. Qin, and Z. P. Jiang, "A new look at distributed optimal output agreement of multi-agent systems," *Automatica*, vol. 136, p. 110053, 2022.

[19] T. Liu, M. Bin, I. Notarnicola, T. Parisini, and Z. P. Jiang, "Discrete-time distributed optimization for linear uncertain multi-agent systems," in *2023 62nd IEEE Conference on Decision and Control (CDC)*. IEEE, 2023, pp. 7439–7444.

[20] T. Liu, L. Cui, B. Pang, and Z. P. Jiang, "A unified framework for data-driven optimal control of connected vehicles in mixed traffic," *IEEE Transactions on Intelligent Vehicles*, vol. 8, no. 8, pp. 4131–4145, 2023.

[21] T. R. Oliveira, M. Krstić, and D. Tsubakino, "Extremum seeking for static maps with delays," *IEEE Transactions on Automatic Control*, vol. 62, no. 4, pp. 1911–1926, 2016.

[22] J. I. Poveda, M. Krstić, and T. Başar, "Fixed-time Nash equilibrium seeking in time-varying networks," *IEEE Transactions on Automatic Control*, vol. 68, no. 4, pp. 1954–1969, 2022.

[23] Z. Qin, L. Jiang, T. Liu, and Z. P. Jiang, "Distributed optimization for uncertain Euler–Lagrange systems with local and relative measurements," *Automatica*, vol. 139, p. 110113, 2022.

[24] Z. Qin, T. Liu, and Z. P. Jiang, "Adaptive backstepping for distributed optimization," *Automatica*, vol. 141, p. 110304, 2022.

[25] Z. Qin, T. Liu, T. Liu, Z. P. Jiang, and T. Chai, "Distributed feedback optimization of nonlinear uncertain systems subject to inequality constraints," *IEEE Transactions on Automatic Control*, 2023, doi: 10.1109/TAC.2023.3343346.

[26] Y. Tan, D. Nešić, and I. Mareels, "On non-local stability properties of extremum seeking control," *Automatica*, vol. 42, no. 6, pp. 889–903, 2006.

[27] Y. Tang, "Multi-agent optimal consensus with unknown control directions," *IEEE Control Systems Letters*, vol. 5, no. 4, pp. 1201–1206, 2020.

[28] Y. Tang and X. Wang, "Optimal output consensus for nonlinear multiagent systems with both static and dynamic uncertainties," *IEEE Transactions on Automatic Control*, vol. 66, no. 4, pp. 1733–1740, 2020.

[29] A. R. Teel, L. Moreau, and D. Nesic, "A unified framework for input-to-state stability in systems with two time scales," *IEEE Transactions on Automatic Control*, vol. 48, no. 9, pp. 1526–1544, 2003.

[30] J. Wang and N. Elia, "Control approach to distributed optimization," in *Proceedings of the 48th Annual Allerton Conference on Communication, Control, and Computing*. IEEE, 2010, pp. 557–561.