On the Sample Complexity of Decentralized Linear Quadratic Regulator With Partially Nested Information Structure

Lintao Ye, *Member, IEEE*, Hao Zhu, *Senior Member, IEEE*, and Vijay Gupta, *Fellow, IEEE*

***Abstract*—In this article, we study the problem of control policy design for decentralized state-feedback linear quadratic control with a partially nested information structure, when the system model is unknown. We propose a model-based learning solution, which consists of two steps. First, we estimate the unknown system model from a single system trajectory of finite length, using least squares estimation. Next, based on the estimated system model, we design a decentralized control policy that satisfies the desired information structure. We show that the suboptimality gap between our control policy and the optimal decentralized control policy (designed using accurate knowledge of the system model) scales linearly with the estimation error of the system model. Using this result, we provide an end-to-end sample complexity result for learning decentralized controllers for a linear quadratic control problem with a partially nested information structure.**

***Index Terms*—Decentralized control, large-scale systems, optimal control, reinforcement learning, system identification, statistical learning.**

I. Introduction

In large-scale control systems, the control policy is often required to be decentralized, where different controllers may only use partial state information, when designing their local control policies. For example, a given controller may only receive a subset of the global state measurements (e.g., [1]), and there may be a delay in receiving the measurements (e.g., [2]). In general, finding a globally optimal control policy under information constraints is NP-hard, even if the system model is known at the controllers [3], [4], [5]. This has led to a large literature on identifying tractable subclasses of the problem. For instance, if the information structure describing the decentralized control problem is partially nested [6], the optimal solution to the state-feedback linear quadratic control problem can be solved efficiently using dynamic programming [7]. Other conditions, such as quadratic invariance [8], [9], have also been identified as tractable subclasses of the problem.

However, the classical work in this field assumes the knowledge of the system model at the controllers. In this work, we are interested in the situation when the system model is not known a priori [10]. In such a case, the existing algorithms do not apply. Moreover, it is not clear whether subclasses, such as problems with partially nested information patterns or where quadratic invariance is satisfied are any more tractable than the general decentralized control problem in this case.

In this article, we consider a decentralized infinite-horizon state-feedback linear quadratic regulator (LQR) control problem with a partially nested information structure [1], [7] and assume that the controllers do not know the system model. We use a model-based learning approach, where we first identify the system model, and then use it to design a decentralized control policy that satisfies the prescribed information constraints.

A. Related Work

Solving optimal control problems without prior system model knowledge has receive much attention recently. One of the most studied problems is the centralized LQR problem. For this problem, two broad classes of methods have been studied, i.e., model based learning [11], [12], [13], and model-free learning [14], [15], [16], [17]. In the model-based learning approach, a system model is first estimated from observed system trajectories using some system identification method. A control policy can then be obtained based on the estimated system model. In the model-free learning approach, the objective function in the LQR problem is first viewed as a function of the control policies. Based on zeroth-order optimization methods (e.g., [18], [19]), the optimal solution can then be obtained using gradient descent, where the gradient of the objective function is estimated from the data samples from system trajectories. Moreover, the model-based

Manuscript received 27 May 2022; accepted 2 October 2022. Date of publication 20 October 2022; date of current version 28 July 2023. The work of Lintao Ye was supported in part by the National Natural Science Foundation of China under Grant 61972170, Grant 62222205 and Grant 62203179; and in part by the National Natural Science Foundation of Hubei under Grant 2021CFB343. The work of Hao Zhu was supported in part by the NSF Grant 1802319 and Grant 2130706. The work of Vijay Gupta was supported in part by the ARO under Grant F.10052139.02.004 and in part by the AFOSR under Grant FA9550-21-1-0231. Recommended by Associate Editor A. A. Malikopoulos.

Lintao Ye is with the Key Laboratory of Image Processing and Intelligent Control, and the School of Artificial Intelligence and Automation, Ministry of Education, Huazhong University of Science and Technology, Wuhan 430074, China (e-mail: yelintao93@hust.edu.cn).

Hao Zhu is with the Department of Electrical and Computer Engineering, University of Texas at Austin, Austin, TX 78712 USA (e-mail: haozhu@utexas.edu).

Vijay Gupta is with the Elmore Family School of Electrical and Computer Engineering, Purdue University, West Lafayette, IN 47907 USA (e-mail: gupta869@purdue.edu).

Color versions of one or more figures in this article are available at https://doi.org/10.1109/TAC.2022.3215940.

Digital Object Identifier 10.1109/TAC.2022.3215940

learning approach has also been studied for the centralized linear quadratic Gaussian control problem [20]. In general, compared to model-free learning, model-based learning tends to require less data samples in order to achieve a policy of equivalent performance [21].

Most of the previous works on model-based learning for centralized LQR build on recent advances in nonasymptotic analyzes for system identification of linear dynamical systems with full state observations (e.g., [22], [23], [24], [25], [26]). Such nonasymptotic analyzes (i.e., sample complexity results) relate the estimation error of the system matrices to the number of samples used for system identification. In particular, it was shown in [23] that when using a single system trajectory, the least squares approach for system identification achieves the optimal sample complexity up to logarithmic factors.

There are few results on solving decentralized linear quadratic control problems with information constraints, when the system model is unknown. In [27], the authors studied a decentralized output-feedback linear quadratic control problem, under the assumption that the quadratic invariance condition is satisfied. The authors proposed a model-free approach and provided a sample complexity analysis. They focused on a finite-horizon setting, since gradient-based optimization methods may not converge to the optimal controller for infinite-horizon decentralized linear quadratic control problems with information constraints, even when the system model is known [28], [29]. In [30], the authors proposed a consensus-based model-free learning algorithm for multiagent decentralized LQR over an infinite horizon, where each agent (i.e., controller) has access to a subset of the global state without delay. They showed that their algorithm converges to a stationary point of the objective function in the LQR problem. In [31], the authors studied model-based learning for LQR with subspace constraints on the closed-loop responses. However, those constraints may not lead to controllers that satisfy the information constraints considered in this article (e.g., [32]).

There is a line of research on online adaptive control for centralized LQR with unknown system models, using either model-based learning [11], [26], [33], or model-free learning [34], [35]. The goal is to adaptively design a control policy online when new data samples from the system trajectory become available, and bound the corresponding regret.

B. Contributions

Here, we summarize our contributions and technical challenges in the article.

• In Section III, we provide a *sample complexity* result for estimating the system model from a single system trajectory using a least squares approach. Despite the existence of a sparsity pattern in the system model considered in our problem, we adapt the analyzes in [26] and [36] for least squares estimation of general linear system models (without any sparsity pattern) to our setting, and show that such a system identification method for general system models suffices for our ensuing analyzes.

• In Section IV, based on the estimated system model, we design a novel decentralized control policy that satisfies the given information structure. Our control policy is inspired by [7], which developed the optimal controller for the decentralized LQR problem with a partially nested information structure and known system model. The optimal controller therein depends on some internal states, each of which evolves according to an auxiliary linear system (characterized by the actual model of the original system with a disturbance term from the original system) and correlates with other internal states. Accordingly, this complicated form of the internal states makes it challenging to extend the design in [7] to the case when the system model is unknown. To tackle this, we capitalize on the observation that the optimal controller proposed in [7] can be viewed as a disturbance-feedback control policy that maps the history of past disturbances (affecting the original system) to the current control input. Thanks to this viewpoint, we put forth a control policy that uses the aforementioned estimated system model and maps the *estimates* of past disturbances to the current control input via some *estimated* internal states. More importantly, we show that the proposed control policy can be implemented in a decentralized manner that satisfies the prescribed information structure.

• In Section V-B, we characterize the performance guarantee (i.e., suboptimality) of the control policy proposed in Section IV. When we compare the performance of our control policy to that of the optimal decentralized control policy in [7], both the estimates of the past disturbances and the estimated internal states contribute to the suboptimality of our control policy, which creates the major technical challenge in our analyzes. We overcome this challenge by carefully investigating the structure of the proposed control policy, and we show that the suboptimality gap between our control policy and the optimal decentralized control policy (designed based on accurate knowledge of the system model) provided in [7] can be decomposed into two terms, both of which scale linearly with the estimation error of the system model.

• In Section V-C, we combine the above results together and provide an end-to-end sample complexity result for learning decentralized LQR with a partially nested information structure. Surprisingly, despite the existence of the information constraints and the fact that the optimal controller is a linear dynamic controller, our sample complexity result matches with that of learning centralized LQR without any information constraints [13].

An extended version of this article that includes all the omitted proofs and details can be found on arXiv as [37].

II. Preliminaries and Problem Formulation

A. Notation and Terminology

The sets of integers and real numbers are denoted as $\mathbb{Z}$ and $\mathbb{R}$, respectively. The set of integers (resp., real numbers) that are greater than or equal to $a \in \mathbb{R}$ is denoted as $\mathbb{Z}_{\geq a}$ (resp., $\mathbb{R}_{\geq a}$). The space of m-dimensional real vectors is denoted by $\mathbb{R}^m$, and the space of $m \times n$ real matrices is denoted by $\mathbb{R}^{m\times n}$. For a matrix $P \in \mathbb{R}^{n\times n}$, let $P^\top$, $\mathrm{Tr}(P)$, $\rho(P)$, and $\{\sigma_i(P) : i \in \{1, \ldots, n\}\}$ be its transpose, trace, spectral radius, and set of singular values, respectively. Without loss of generality, let the singular values of P be ordered as $\sigma_1(P) \geq \cdots \geq \sigma_n(P)$. Let $\|\cdot\|$ denote the ℓ_2 norm, i.e., $\|P\| = \sigma_1(P)$ for a matrix $P \in \mathbb{R}^{n\times n}$, and $\|x\| = \sqrt{x^\top x}$ for a vector $x \in \mathbb{R}^n$. Let $\|P\|_F = \sqrt{\mathrm{Tr}(PP^\top)}$ denote the Frobenius norm of $P \in \mathbb{R}^{n\times m}$.

A positive semidefinite matrix P is denoted by $P \succeq 0$, and $P \succeq Q$ if and only if $P - Q \succeq 0$. Let $\mathbb{S}_+^n$ (resp., $\mathbb{S}_{++}^n$) denote the set of $n \times n$ positive semidefinite (resp., positive definite) matrices. Let I denote an identity matrix whose dimension can be inferred from the context. Given any integer $n \geq 1$, we define $[n] = \{1, \ldots, n\}$. The cardinality of a finite set $\mathcal{A}$ is denoted by $|\mathcal{A}|$. Let $\mathcal{N}(\mu, \Sigma)$ denote a Gaussian distribution with mean $\mu \in \mathbb{R}^m$ and covariance $\Sigma \in \mathbb{S}_+^m$.

B. Decentralized LQR With Sparsity and Delay Constraints

In this section, we sketch the method developed in [1] and [7], which presents the optimal solution to a decentralized LQR problem with a *partially nested* information structure [6], when the system model is known a priori. First, let us consider a networked system that consists of $p \in \mathbb{Z}_{\geq 1}$ interconnected linear-time-invariant (LTI) subsystems, and let $\mathcal{V} = [p]$ be the set that contains all the p subsystems. Letting the state, input, and disturbance of the subsystem corresponding to node $i \in [p]$ be $x_i(t) \in \mathbb{R}^{n_i}$, $u_i(t) \in \mathbb{R}^{m_i}$, and $w_i(t)$, respectively, the subsystem corresponding to node i is given by

$$x_i(t+1) = \left(\sum_{j \in \mathcal{N}_i} A_{ij} x_j(t) + B_{ij} u_j(t) \right) + w_i(t) \quad \forall i \in \mathcal{V} \tag{1}$$

where $\mathcal{N}_i \subseteq [p]$ is the set of subsystems whose states and inputs directly affect the state of subsystem i, $A_{ij} \in \mathbb{R}^{n_i \times n_j}$, $B_{ij} \in \mathbb{R}^{n_i \times m_j}$, and $w_i(t) \in \mathbb{R}^{n_i}$ is a white Gaussian noise process with $w_i(t) \sim \mathcal{N}(0, \sigma_w^2 I)$ for all $t \in \mathbb{Z}_{\geq 0}$, where $\sigma_w \in \mathbb{R}_{>0}$.[1] For simplicity, we assume throughout this article that $n_i \geq m_i$ for all $i \in \mathcal{V}$. We can also write (1) as

$$x_i(t+1) = A_i x_{\mathcal{N}_i}(t) + B_i u_{\mathcal{N}_i}(t) + w_i(t) \quad \forall i \in \mathcal{V} \tag{2}$$

where $A_i \triangleq [A_{ij_1} \cdots A_{ij_{|\mathcal{N}_i|}}]$, $B_i \triangleq [B_{ij_1} \cdots B_{ij_{|\mathcal{N}_i|}}]$, $x_{\mathcal{N}_i}(t) \triangleq [x_{j_1}(t) \cdots x_{j_{|\mathcal{N}_i|}}(t)]^\top$, and $u_{\mathcal{N}_i}(t) \triangleq [u_{j_1}(t) \cdots u_{j_{|\mathcal{N}_i|}}(t)]^\top$, with $\mathcal{N}_i = \{j_1, \ldots, j_{|\mathcal{N}_i|}\}$. Further letting $n = \sum_{i \in \mathcal{V}} n_i$ and $m = \sum_{i \in \mathcal{V}} m_i$, and defining $x(t) = [x_1(t)^\top \cdots x_p(t)^\top]^\top$, $u(t) = [u_1(t)^\top \cdots u_p(t)^\top]^\top$, and $w(t) = [w_1(t)^\top \cdots w_p(t)^\top]^\top$, we can compactly write (1) into the following matrix form:

$$x(t+1) = Ax(t) + Bu(t) + w(t) \tag{3}$$

where the (i,j)th block of $A \in \mathbb{R}^{n \times n}$ (resp., $B \in \mathbb{R}^{n \times m}$), i.e., A_{ij} (resp., B_{ij}) satisfies $A_{ij} = 0$ (resp., $B_{ij} = 0$) if $j \notin \mathcal{N}_i$. We assume that $w_i(t_1)$ and $w_j(t_2)$ are independent for all $i, j \in \mathcal{V}$ with $i \neq j$ and for all $t_1, t_2 \in \mathbb{Z}_{\geq 0}$. In other words, $w(t)$ is a white Gaussian noise process with $w(t) \sim \mathcal{N}(0, \sigma_w^2 I)$ for all $t \in \mathbb{Z}_{\geq 0}$. For simplicity, we assume that $x(0) = 0$ throughout this article.[2]

Next, we use a directed graph $\mathcal{G}(\mathcal{V}, \mathcal{A})$ with $\mathcal{V} = [p]$ to characterize the (time-delayed) information flow among the subsystems in $[p]$ due to communication constraints on the subsystems. Each node in $\mathcal{G}(\mathcal{V}, \mathcal{A})$ represents a subsystem in $[p]$, and we assume that $\mathcal{G}(\mathcal{V}, \mathcal{A})$ does not have self loops. We associate any edge $(i,j) \in \mathcal{A}$ with a delay of either 0 or 1, further denoted as $i \xrightarrow{0} j$ or $i \xrightarrow{1} j$, respectively.[3] Then, we define the delay matrix corresponding to $\mathcal{G}(\mathcal{V}, \mathcal{A})$ as $D \in \mathbb{R}^{p \times p}$ such that:

i) if $i \neq j$ and there is a directed path from j to i in $\mathcal{G}(\mathcal{V}, \mathcal{A})$, then D_{ij} is equal to the sum of delays along the directed path from node j to node i with the smallest accumulative delay;
ii) if $i \neq j$ and there is no directed path from j to i in $\mathcal{G}(\mathcal{V}, \mathcal{A})$, then $D_{ij} = +\infty$;
iii) $D_{ii} = 0$ for all $i \in \mathcal{V}$.

Here, we consider the scenario where the information (e.g., state information) corresponding to subsystem $j \in \mathcal{V}$ can propagate to subsystem $i \in \mathcal{V}$ with a delay of D_{ij} (in time), if and only if there exists a directed path from j to i with an accumulative delay of D_{ij}. Note that as argued in [7], we assume that there is no directed cycle with zero accumulative delay; otherwise, one can first collapse all the nodes in such a directed cycle into a single node, and equivalently consider the resulting directed graph in the framework described earlier.

To proceed, we consider designing the control input $u(t)$ for the LTI system in (3). We focus on *state-feedback* control, i.e., we can view $u(t)$ as a policy that maps the states of the LTI system to a control input. Moreover, we require that $u(t)$ satisfy the information structure according to the directed graph $\mathcal{G}(\mathcal{V}, \mathcal{A})$ and the delay matrix $D \in \mathbb{R}^{p \times p}$, described earlier. Specifically, considering any $i \in \mathcal{V}$ and any $t \in \mathbb{Z}_{\geq 0}$, and noting that the controller corresponding to subsystem $i \in \mathcal{V}$ provides the control input $u_i(t) \in \mathbb{R}^{m_i}$, the state information that is available to the controller at $i \in \mathcal{V}$ is given by

$$\mathcal{I}_i(t) = \{x_j(k) : j \in \mathcal{V}_i, 0 \leq k \leq t - D_{ij}\} \tag{4}$$

where $\mathcal{V}_i \triangleq \{j \in \mathcal{V} : D_{ij} \neq +\infty\}$. In the sequel, we also call $\mathcal{I}_i(t)$ the *information set* of controller $i \in \mathcal{V}$ at time $t \in \mathbb{Z}_{\geq 0}$. Note that $\mathcal{I}_i(t)$ contains the states corresponding to the subsystems in $\mathcal{V}$ that have enough time to reach subsystem $i \in \mathcal{V}$ at time $t \in \mathbb{Z}_{\geq 0}$, due to the sparsity and delay constraints described earlier. Now, based on the information set $\mathcal{I}_i(t)$, we further define $\mathcal{S}(\mathcal{I}_i(t))$ to be the set that consists of all the policies that map the states in $\mathcal{I}_i(t)$ to a control input at node i. The goal is then to solve the following constrained optimization problem:

$$\begin{aligned} \min_{u(0), u(1), \ldots} \lim_{T \to \infty} \mathbb{E} \left[\frac{1}{T} \sum_{t=0}^{T-1} (x(t)^\top Q x(t) + u(t)^\top R u(t)) \right] \\ \text{s.t. } x(t+1) = Ax(t) + Bu(t) + w(t) \\ u_i(t) \in \mathcal{S}(\mathcal{I}_i(t)) \quad \forall i \in \mathcal{V} \quad \forall t \in \mathbb{Z}_{\geq 0} \end{aligned} \tag{5}$$

where $Q \in \mathbb{S}_+^n$ and $R \in \mathbb{S}_{++}^m$ are the cost matrices, and the expectation is taken with respect to $w(t)$ for all $t \in \mathbb{Z}_{\geq 0}$. Throughout this article, we always assume that the following assumption on the information propagation pattern among the subsystems in $\mathcal{V}$ holds (e.g., [7], [38]).

Assumption 1: For any $i \in \mathcal{V}$, it holds that $\mathcal{N}_i = \{j \in \mathcal{V} : D_{ij} \leq 1\}$, where $\mathcal{N}_i$ is given in (1).

Assumption 1 says that the state of subsystem $i \in \mathcal{V}$ is affected by the state and input of subsystem $j \in \mathcal{V}$, if and only if there is

[1] The analysis can be extended to the case when $w_i(t)$ is assumed to be a zero-mean white Gaussian noise process with covariance $W \in \mathbb{S}_{++}^{n_i}$. In that case, our analysis will depend on $\max_{i \in \mathcal{V}} \sigma_1(W_i)$ and $\min_{i \in \mathcal{V}} \sigma_n(W_i)$.

[2] The analysis can be extended to the case when $x(0)$ is given by a zero-mean Gaussian distribution, as one may view $x(0)$ as $w(-1)$.

[3] The framework described in this article can also be used to handle $\mathcal{G}(\mathcal{V}, \mathcal{A})$ with larger delays; see [7] for a detailed discussion.

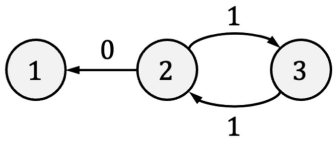


Fig. 1. Directed graph of Example 1. Node $i \in \mathcal{V}$ represents a subsystem with state $x_i(t)$ and edge $(i,j) \in \mathcal{A}$ is labeled with the information propagation delay from i to j.

a communication link with a delay of at most 1 from subsystem j to i in $\mathcal{G}(\mathcal{V},\mathcal{A})$. As shown in [7], Assumption 1 ensures that the information structure associated with the system given in (1) is partially nested [6]. Assumption 1 is frequently used in decentralized control problems (e.g., [1], [7] and the references therein), and one can see that the assumption is satisfied in networked systems where information propagates at least as fast as dynamics. To illustrate our arguments above, we introduce Example 1.

Example 1: Consider a directed graph $\mathcal{G}(\mathcal{V},\mathcal{A})$ given in Fig. 1, where $\mathcal{V}=\{1,2,3\}$ and each directed edge is associated with a delay of 0 or 1. The corresponding LTI system is then given by

$$\begin{bmatrix} x_1(t+1) \\ x_2(t+1) \\ x_3(t+1) \end{bmatrix} = \begin{bmatrix} A_{11} & A_{12} & A_{13} \\ 0 & A_{22} & A_{23} \\ 0 & A_{32} & A_{33} \end{bmatrix} \begin{bmatrix} x_1(t) \\ x_2(t) \\ x_3(t) \end{bmatrix} + \begin{bmatrix} B_{11} & B_{12} & B_{13} \\ 0 & B_{22} & B_{23} \\ 0 & B_{32} & B_{33} \end{bmatrix} \begin{bmatrix} u_1(t) \\ u_2(t) \\ u_3(t) \end{bmatrix} + \begin{bmatrix} w_1(t) \\ w_2(t) \\ w_3(t) \end{bmatrix}. \tag{6}$$

Now, in order to present the solution to (5) given in, e.g., [7], we need to construct an information graph $\mathcal{P}(\mathcal{U},\mathcal{H})$. Considering any directed graph $\mathcal{G}(\mathcal{V},\mathcal{A})$ with $\mathcal{V}=[p]$, and the delay matrix $D \in \mathbb{R}^{p\times p}$ as we described earlier, let us first define $s_j(k)$ to be the set of nodes in $\mathcal{G}(\mathcal{V},\mathcal{A})$ that are reachable from node j within k time steps, i.e., $s_j(k)=\{i\in\mathcal{V} : D_{ij}\le k\}$. The information graph $\mathcal{P}(\mathcal{U},\mathcal{H})$ is then constructed as

$$\begin{aligned} \mathcal{U} &= \{s_j(k) : k\ge 0, j\in\mathcal{V}\} \\ \mathcal{H} &= \{(s_j(k), s_j(k+1)) : k\ge 0, j\in\mathcal{V}\}. \end{aligned} \tag{7}$$

Thus, we see from (7) that each node $s\in\mathcal{U}$ corresponds to a set of nodes from $\mathcal{V}=[p]$ in the original directed graph $\mathcal{G}(\mathcal{V},\mathcal{A})$. Using a similar notation to that for the graph $\mathcal{G}(\mathcal{V},\mathcal{A})$, if there is an edge from s to r in $\mathcal{P}(\mathcal{U},\mathcal{H})$, we denote the edge as $s\to r$. Additionally, considering any $s_i(0)\in\mathcal{U}$, we write $w_i\to s_i(0)$ to indicate the fact that the noise $w_i(t)$ is injected to node $i\in\mathcal{V}$ at time $t\in\mathbb{Z}_{\ge 0}$.[4] From the above construction of the information graph $\mathcal{P}(\mathcal{U},\mathcal{H})$, one can show that the following properties hold.

Lemma 1: [7, Prop. 1] Given a directed graph $\mathcal{G}(\mathcal{V},\mathcal{A})$ with $\mathcal{V}=[p]$, the information graph $\mathcal{P}(\mathcal{U},\mathcal{H})$ constructed in (7) satisfies the following:

i) For every $r\in\mathcal{U}$, there is a unique $s\in\mathcal{U}$ such that $(r,s)\in\mathcal{H}$, i.e., $r\to s$;
ii) every path in $\mathcal{P}(\mathcal{U},\mathcal{H})$ ends at a node with a self loop;

[4]Note that we have assumed that there is no directed cycle with zero accumulative delay in $\mathcal{P}(\mathcal{U},\mathcal{H})$. Hence, one can show that for any $s_i(0)\in\mathcal{U}$, w_i is the only noise term such that $w_i\to s_i(0)$.

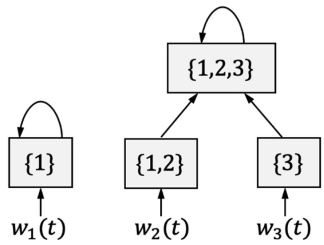


Fig. 2. Information graph of Example 1. Each node in the information graph is a subset of the nodes in the directed graph given in Fig. 1.

iii) $n\le|\mathcal{U}|\le p^2-p+1$.

Remark 1: One can see from the construction of $\mathcal{P}(\mathcal{U},\mathcal{H})$ and Lemma 1 that $\mathcal{P}(\mathcal{U},\mathcal{H})$ is a forest, i.e., a set of disconnected directed trees, where each directed tree in the forest is oriented to a node with a self loop in $\mathcal{P}(\mathcal{U},\mathcal{H})$. Specifically, $s_i(0)$ for all $i\in\mathcal{V}$ are the leaf nodes in $\mathcal{P}(\mathcal{U},\mathcal{H})$, and the nodes with self loop are root nodes in $\mathcal{P}(\mathcal{U},\mathcal{H})$.

The information graph $\mathcal{P}(\mathcal{U},\mathcal{H})$ corresponding to Example 1 is given in Fig. 2. Note that the information graph $\mathcal{P}(\mathcal{U},\mathcal{H})$ in Fig. 2 contains two disconnected directed trees, one of which is an isolated node $\{1\}\in\mathcal{U}$ with a self loop. Also notice that $s_1(0)=\{1\}$, $s_2(0)=\{1,2\}$, and $s_3(0)=\{3\}$.

Throughout this article, we assume that the elements in $\mathcal{V}=[p]$ are ordered in an increasing manner, and that the elements in s are also ordered in an increasing manner for all $s\in\mathcal{U}$. Now, for any $s,r\in\mathcal{U}$, we use A_{sr} (or $A_{s,r}$) to denote the submatrix of A that corresponds to the nodes in s and r. For example, $A_{\{1\},\{1,2\}}=[A_{11}\ A_{12}]$. In the sequel, we will also use similar notations to denote submatrices of B, Q, R and the identity matrix I. We will make the following standard assumptions (see, e.g., [7]).

Assumption 2: For any $s\in\mathcal{U}$ that has a self loop, the pair (A_{ss},B_{ss}) is stabilizable and the pair (A_{ss},C_{ss}) is detectable, where $Q_{ss}=C_{ss}^\top C_{ss}$.

Leveraging the partial nestedness of (5), the authors in [7] obtained the optimal solution to (5).

Lemma 2: [7, Corollary 4] Consider the problem given in (5), and let $\mathcal{P}(\mathcal{U},\mathcal{H})$ be the associated information graph. Suppose Assumption 2 holds. For all $r\in\mathcal{U}$, define matrices P_r and K_r recursively as

$$K_r = -(R_{rr}+B_{sr}^\top P_s B_{sr}^\top)^{-1}B_{sr}^\top P_s A_{sr} \tag{8}$$

$$P_r = Q_{rr}+K_r^\top R_{rr}K_r + (A_{sr}+B_{sr}K_r)^\top P_s(A_{sr}+B_{sr}K_r) \tag{9}$$

where for each $r\in\mathcal{U}$, $s\in\mathcal{U}$ is the unique node such that $r\to s$. In particular, for any $s\in\mathcal{U}$ that has a self loop, the matrix P_s is the unique positive semidefinite solution to the Riccati equation given by (9), and the matrix $A_{ss}+B_{ss}K_s$ is stable. The optimal solution to (5) is then given by

$$\zeta_s(t+1)=\sum_{r\to s}(A_{sr}+B_{sr}K_r)\zeta_r(t)+\sum_{w_i\to s}I_{s,\{i\}}w_i(t) \tag{10}$$

and

$$u_i^\star(t)=\sum_{r\ni i}I_{\{i\},r}K_r\zeta_r(t) \tag{11}$$

for all $t \in \mathbb{Z}_{\geq 0}$, where $\zeta_s(t)$ is an internal state initialized with $\zeta_s(0) = \sum_{w_i \to s} I_{s,\{i\}} x_i(0) = 0$ for all $s \in \mathcal{U}$. The corresponding optimal cost of (5), denoted as $J_\star$, is given by

$$J_\star = \sigma_w^2 \sum_{\substack{i \in \mathcal{V} \\ w_i \to s}} \operatorname{Tr}\left(I_{\{i\},s} P_s I_{s,\{i\}}\right). \tag{12}$$

Let us use Example 1 to illustrate the results in Lemma 2. First, considering node $\{1\} \in \mathcal{U}$ in the information graph $\mathcal{P}(\mathcal{U}, \mathcal{H})$ given in Fig. 2, we have from (10) that

$$\begin{aligned}\zeta_1(t+1) &= (A_{11} + B_{11}K_1)\zeta_1(t) + \sum_{w_i \to \{1\}} I_{\{1\},\{i\}} w_i(t) \\ &= (A_{11} + B_{11}K_1)\zeta_1(t) + w_1(t).\end{aligned}$$

Next, considering node $2 \in \mathcal{V}$ in the directed graph $\mathcal{G}(\mathcal{V}, \mathcal{A})$ given in Fig. 1, we see from (11) and Fig. 2 that

$$\begin{aligned}u_2^\star(t) = \sum_{r \ni 2} I_{\{2\},r} K_r \zeta_r(t) &= I_{\{2\},\{1,2\}} K_{\{1,2\}} \zeta_{\{1,2\}}(t) \\ &+ I_{\{2\},\{1,2,3\}} K_{\{1,2,3\}} \zeta_{\{1,2,3\}}(t)\end{aligned}$$

where K_r is given by (8).

Remark 2: Obtaining the optimal policy $u_i^\star(t)$, for any $i \in \mathcal{V}$, given by Lemma 2 requires global knowledge of the system matrices A and B, the cost matrices Q and R, and the directed graph $\mathcal{G}(\mathcal{V}, \mathcal{A})$ with the associated delay matrix D. Moreover, $u^\star(t)$ given in Lemma 2 is not a static state-feedback controller, but a linear dynamic controller based on the internal states $\zeta_r(\cdot)$ for all $r \in \mathcal{U}$. For any controller $i \in \mathcal{V}$ and for any $t \in \mathbb{Z}_{\geq 0}$, the authors in [7] proposed an algorithm to determine $\zeta_r(t)$ for all $r \in \mathcal{U}$ such that $i \in r$, and thus $u_i^\star(t)$, using only the memory maintained by the algorithm, the state information contained in the information set $\mathcal{I}_i(t)$ defined in (4), and the global information described earlier.

C. Problem Formulation and Summary of Results

We now formally introduce the problem that we will study in this article. We consider the scenario where the system matrices A and B are unknown. However, we assume that the directed graph $\mathcal{G}(\mathcal{V}, \mathcal{A})$ and the associated delay matrix D are known. Similarly to, e.g., [13], [20], we consider the scenario where we can first conduct experiments in order to estimate the unknown system matrices A and B. Specifically, starting from the initial state $x(0) = 0$, we evolve the system given in (3) for $N \in \mathbb{Z}_{\geq 1}$ time steps using a given control input sequence $\{u(0), u(1), \ldots, u(N-1)\}$, and collect the resulting state sequence $\{x(1), x(2), \ldots, x(N)\}$. Based on $\{u(0), \ldots, u(N-1)\}$ and $\{x(0), \ldots, x(N)\}$, we use a least squares approach to obtain estimates of the system matrices A and B, denoted as $\hat{A}$ and $\hat{B}$, respectively. Using the obtained $\hat{A}$ and $\hat{B}$, the goal is still to solve (5). Since the true system matrices A and B are unknown, it may no longer be possible to solve (5) optimally, using the methods introduced in Section II-B. Thus, we aim to provide a solution to (5) using $\hat{A}$ and $\hat{B}$, and characterize its performance (i.e., suboptimality) guarantees.

In the rest of this article, we first analyze the estimation error of $\hat{A}$ and $\hat{B}$ obtained from the procedure described earlier. In particular, we show in Section III that the estimation errors $\|\hat{A} - A\|$ and $\|\hat{B} - B\|$ scale as $\tilde{\mathcal{O}}(1/\sqrt{N})$ with high probability.[5] Next, in Section IV, we design a control policy $\hat{u}(\cdot)$, based on $\hat{A}$ and $\hat{B}$, which satisfies the information constraints given in (5). Supposing $\|\hat{A} - A\| \leq \varepsilon$ and $\|\hat{B} - B\| \leq \varepsilon$, where $\varepsilon \in \mathbb{R}_{>0}$, and denoting the cost of (5) corresponding to $\hat{u}(\cdot)$ as $\hat{J}$, we show in Section V-B that

$$\hat{J} - J_\star \leq C\varepsilon$$

as long as $\varepsilon \leq C_0$, where $J_\star$ is the optimal cost of (5) given by (12), and C_0 and C are constants that explicitly depend on the problem parameters of (5). Finally, combining the abovementioned results together, we show in Section V-C that with high probability and for large enough N, the following end-to-end sample complexity of learning decentralized LQR with the partially nested information structure holds

$$\hat{J} - J_\star = \tilde{\mathcal{O}}\left(\frac{1}{\sqrt{N}}\right).$$

Algorithm 1: Least Squares Estimation of A and B.

Input: parameter $\lambda > 0$ and time horizon length N
1: Initialize $x(0) = 0$
2: **For** $t = 0, \ldots, N-1$ **do**
3: Play $u(t) \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, \sigma_u^2 I)$
4: Obtain $\hat{\Theta}(N)$ using (14)
5: Extract $\hat{A}$ and $\hat{B}$ from $\hat{\Theta}(N)$

III. System Identification Using Least Squares

As we described in Section II-C, we use a least squares approach to estimate the system matrices $A \in \mathbb{R}^{n \times n}$ and $B \in \mathbb{R}^{n \times m}$, based on a single system trajectory consisting of the control input sequence $\{u(0), \ldots, u(N-1)\}$ and the system state sequence $\{x(0), \ldots, x(N)\}$, where $x(0) = 0$ and $N \in \mathbb{Z}_{\geq 1}$. Here, we draw the inputs $u(0), \ldots, u(N-1)$ independently from a Gaussian distribution $\mathcal{N}(0, \sigma_u^2 I)$, where $\sigma_u \in \mathbb{R}_{>0}$. In other words, we let $u(t) \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, \sigma_u^2 I)$ for all $t \in \{0, \ldots, N-1\}$. Moreover, we assume that the input $u(t)$ and the disturbance $w(t)$ are independent for all $t \in \{0, \ldots, N-1\}$. Note that we consider the scenario where the estimation of A and B is performed in a centralized manner using a least squares approach (detailed in Algorithm 1). However, we remark that Algorithm 1 can be carried out without violating the information constraints given by (4), since $u(t) \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, \sigma_u^2 I)$ is not a function of the states in the information set defined in (4) for any $t \in \{0, \ldots, N-1\}$. In the following, we present the least squares approach to estimate A and B, and characterize the corresponding estimation error.

A. Lest Squares Estimation of System Matrices

Let us denote

$$\Theta = [A \quad B] \text{ and } z(t) = [x(t)^\top \quad u(t)^\top]^\top \tag{13}$$

[5] Throughout this article, we let $\tilde{\mathcal{O}}(\cdot)$ hide logarithmic factors in N.

where $\Theta \in \mathbb{R}^{n\times(n+m)}$ and $z(t) \in \mathbb{R}^{n+m}$. Given the sequences $\{z(0),\dots,z(N-1)\}$ and $\{x(1),\dots,x(N)\}$, we use regularized least squares to obtain an estimate of Θ as

$$\hat{\Theta}(N) = \arg\min_{Y\in\mathbb{R}^{n\times(n+m)}} \left\{ \lambda\|Y\|_F^2 + \sum_{t=0}^{N-1} \|x(t+1) - Yz(t)\|^2 \right\} \tag{14}$$

where $\lambda \in \mathbb{R}_{>0}$ is the regularization parameter. We summarize the above least squares approach in Algorithm 1.

B. Least Squares Estimation Error

For the analysis in the sequel, we will make the following assumption, which is also made in related literature (see, e.g., [20], [39], [40]).

Assumption 3: The system matrix $A \in \mathbb{R}^{n\times n}$ is stable, and $\|A^k\| \le \kappa_0\gamma_0^k$ for all $k \in \mathbb{Z}_{\ge 0}$, where $\kappa_0 \ge 1$ and $\rho(A) < \gamma_0 < 1$.

Note that for any stable matrix A, we have from the Gelfand formula (e.g., [41]) that there exist $\kappa_0 \in \mathbb{R}_{\ge 1}$ and $\gamma_0 \in \mathbb{R}$ with $\rho(A) < \gamma_0 < 1$ such that $\|A^k\| \le \kappa_0\gamma_0^k$ for all $k \in \mathbb{Z}_{\ge 0}$. In order to characterize the estimation error of $\hat{\Theta}(N)$ given by (14), we combine ideas from [25], [26], [36], and show that the following result holds.

Proposition 1: Suppose Assumption 3 holds, and $\|A\| \le \vartheta$ and $\|B\| \le \vartheta$, where $\vartheta \in \mathbb{R}_{>0}$. Consider any $\delta > 0$. Let the input parameters to Algorithm 1 satisfy $N \ge 200(n+m)\log\frac{48}{\delta}$ and $\lambda \ge \underline{\sigma}^2/40$, where $\underline{\sigma} = \min\{\sigma_w, \sigma_u\}$, and

$$z_b = \frac{5\kappa_0}{1-\gamma_0}\overline{\sigma}\sqrt{(\|B\|^2 m + m + n)\log\frac{4N}{\delta}}$$

where κ_0 and γ_0 are given in Assumption 3, and $\overline{\sigma} = \max\{\sigma_w, \sigma_u\}$. Then, with probability at least $1-\delta$, it holds that $\|\hat{A} - A\| \le \varepsilon_0$ and $\|\hat{B} - B\| \le \varepsilon_0$, where $\hat{A}$ and $\hat{B}$ are returned by Algorithm 1, and

$$\varepsilon_0 = 4\sqrt{\frac{160}{N\underline{\sigma}^2}\left(2n\sigma_w^2(n+m)\log\frac{N + z_b^2/\lambda}{\delta} + \lambda n\vartheta^2\right)}.$$

Proof: We provide a sketched proof here and defer the complete proof to [37]. The general idea of the proof is to define a probabilistic event $\mathcal{E}$ and show that several favorable properties hold on the event $\mathcal{E}$. By showing that the event $\mathcal{E}$ holds with probability at least $1-\delta$, the result of the proposition follows. Specifically, one can first show that under the event $\mathcal{E}$, $z(t)$ defined in (13) satisfies that $\|z(t)\| \le z_b$ for all $t \in \{0,\dots,N-1\}$. Then, using [26, Lemma 6] (which is a consequence of [25, Th. 1]), one can then show that under the event $\mathcal{E}$, $\hat{\Theta}(N)$ returned by Algorithm 2 satisfies that $\|\hat{\Theta}(N) - \Theta\| \le \varepsilon_0$, where Θ is defined in (13). Since Algorithm 1 extracts $\hat{A}$ and $\hat{B}$ from $\hat{\Theta}(N)$, i.e., $\hat{\Theta}(N) = \begin{bmatrix}\hat{A} & \hat{B}\end{bmatrix}$, one can show that $\|\hat{A} - A\| \le \|\hat{\Theta}(N) - \Theta\|$ and $\|\hat{B} - B\| \le \|\hat{\Theta}(N) - \Theta\|$. ■

Several remarks pertaining to Algorithm 1 and the result in Proposition 1 are now in order. First, note that while considering the problem of learning centralized LQR without any information constraints, the authors in [13] proposed to obtain $\hat{A}$ and $\hat{B}$ from multiple system trajectories using least squares, where each trajectory starts from $x(0) = 0$. They showed that $\|\hat{A} - A\| = \mathcal{O}(1/\sqrt{N_r})$ and $\|\hat{B} - B\| = \mathcal{O}(1/\sqrt{N_r})$, where $N_r \in \mathbb{Z}_{\ge 1}$ is the number of system trajectories. In contrast, we estimate A and B from a single system trajectory, and achieve $\|\hat{A} - A\| = \tilde{\mathcal{O}}(1/\sqrt{N})$ and $\|\hat{B} - B\| = \tilde{\mathcal{O}}(1/\sqrt{N})$.

Second, note that we use the regularized least squares in Algorithm 1 to obtain estimates $\hat{A}$ and $\hat{B}$. Although least squares without regularization can also be used to obtain estimates $\hat{A}$ and $\hat{B}$ from a single system trajectory with the same $\tilde{\mathcal{O}}(1/\sqrt{N})$ finite sample guarantee (e.g., [23]), we choose to use regularized least squares considered in, e.g., [25], [26], [36]. The reason is that introducing the regularization into least squares makes the finite sample analysis more tractable (e.g., [26], [36]), which facilitates the adaption of the analysis in [26] and [36] to our setting described in this section; more details can be found in [37]. Moreover, note that the lower bound on λ required in Proposition 1 is merely used to guarantee that the denominator of the abovementioned expression for ε_0 contains the factor $1/\sqrt{N}$; choosing an arbitrary $\lambda \in \mathbb{R}_{>0}$ leads to a factor of $1/\sqrt{N-1}$. In general, one can show that choosing any $\lambda \in \mathbb{R}_{>0}$ leads to the same $\tilde{\mathcal{O}}(1/\sqrt{N})$ finite sample guarantee.

Third, note that we do not leverage the block structure (i.e., sparsity pattern) of A and B described in Section II-B, when we obtain $\hat{A}$ and $\hat{B}$ via Algorithm 1. Thus, the sparsity pattern of $\hat{A}$ and $\hat{B}$ may potentially be inconsistent with that of A and B. Nonetheless, such a potential inconsistency does not play any role in our analysis later. The reason is that the control policy to be proposed later in Section IV does not depend on the sparsity pattern of $\hat{A}$ and $\hat{B}$. Moreover, when analyzing the suboptimality of the proposed control policy in Section V, we only leverage the fact that the estimation error corresponding to submatrices in $\hat{A}$ (resp., $\hat{B}$) will be upper bounded by $\|\hat{A} - A\|$ (resp., $\|\hat{B} - B\|$). Specifically, considering any nodes s, r in the information graph $\mathcal{P}(\mathcal{U}, \mathcal{H})$ given by (7), one can show that $\|\hat{A}_{sr} - A_{sr}\| \le \|\hat{A} - A\|$.

Finally, we remark that one may also use system identification schemes and the associated sample complexity analysis dedicated to sparse system matrices (e.g., [42]). Under some extra assumptions on A and B (e.g., [42]), one may then obtain $\hat{A}$ and $\hat{B}$ that have the same sparsity pattern as A and B, and remove the logarithmic factor in N in ε_0 defined in Proposition 1. However, the assumptions on A and B made in e.g., [42] can be restrictive and hard to check in practice.

IV. Control Policy Design

While the estimation of A and B is performed in a centralized manner as we discussed in Section III-A, we assume that each controller $i \in \mathcal{V}$ receives the estimates $\hat{A}$ and $\hat{B}$ after we conduct the system identification step described in Algorithm 1. Given the matrices $\hat{A}$, $\hat{B}$, Q, and R, and the directed graph $\mathcal{G}(\mathcal{V}, \mathcal{A})$ ($\mathcal{V} = [p]$) with the delay matrix D, in this section, we design a control policy that can be implemented in a decentralized manner, while satisfying the information constraints described in Section II-B. To this end, we leverage the structure of the optimal policy $u^\star(\cdot)$ given in Lemma 2 (when A and B are

Algorithm 2: Control Policy Design for Node $i \in \mathcal{V}$.

Input: estimates $\hat{A}$ and $\hat{B}$, cost matrices Q and R, directed graph $\mathcal{G}(\mathcal{V}, \mathcal{A})$ with $\mathcal{V} = [p]$ and delay matrix D, time horizon length T

1: Construct the information graph $\mathcal{P}(\mathcal{U}, \mathcal{H})$ from (7)
2: Obtain $\hat{K}_s$ for all $s \in \mathcal{U}$ from (8)
3: Initialize $\mathcal{M}_i \leftarrow \bar{\mathcal{M}}_i$
4: **for** $t = 0, \dots T-1$ **do**
5: **for** $s \in \mathcal{L}(\mathcal{T}_i)$ **do**
6: Find $s_j(0) \in \mathcal{U}$ s.t. $j \in \mathcal{V}$ and $s_j(0) = s$
7: Obtain $\hat{w}_j(t - D_{ij} - 1)$ from (25)
8: Obtain $\hat{\zeta}_s(t - D_{ij})$ from (24)
9: $\mathcal{M}_i \leftarrow \mathcal{M}_i \cup \{\hat{\zeta}_s(t - D_{ij})\}$
10: **for** $s \in \mathcal{R}(\mathcal{T}_i)$ **do**
11: Obtain $\hat{\zeta}_s(t - D_{\max})$ from (24)
12: $\mathcal{M}_i \leftarrow \mathcal{M}_i \cup \{\hat{\zeta}_s(t - D_{\max})\}$
13: Play $\hat{u}_i(t) = \sum_{r \ni i} I_{\{i\},r} \hat{K}_r \hat{\zeta}_r(t)$
14: $\mathcal{M}_i \leftarrow \mathcal{M}_i \setminus (\{\hat{\zeta}_s(t - 2D_{\max} - 1) : s \in \mathcal{L}(\mathcal{T}_i)\} \cup \{\hat{\zeta}_s(t - D_{\max} - 1) : s \in \mathcal{R}(\mathcal{T}_i)\})$

known). Note that the optimal policy $u^\star(\cdot)$ cannot be applied to our scenario, since only $\hat{A}$ and $\hat{B}$ are available.

First, given the directed graph $\mathcal{G}(\mathcal{V}, \mathcal{A})$ with $\mathcal{V} = [p]$ and the delay matrix D, we construct the information graph $\mathcal{P}(\mathcal{U}, \mathcal{H})$ given by (7). Recall from Remark 1 that $\mathcal{P}(\mathcal{U}, \mathcal{H})$ is a forest that contains a set of disconnected directed trees. We then let $\mathcal{L}$ denote the set of all the leaf nodes in $\mathcal{P}(\mathcal{U}, \mathcal{H})$, i.e.,

$$\mathcal{L} = \{s_i(0) \in \mathcal{U} : i \in \mathcal{V}\}. \tag{15}$$

Moreover, for any $s \in \mathcal{U}$, we denote

$$\mathcal{L}_s = \{v \in \mathcal{L} : v \rightsquigarrow s\} \tag{16}$$

where we write $v \rightsquigarrow s$ if and only if there is a unique directed path from node v to node s in $\mathcal{P}(\mathcal{U}, \mathcal{H})$. In other words, $\mathcal{L}_s$ is the set of leaf nodes in $\mathcal{P}(\mathcal{U}, \mathcal{H})$ that can reach s. Moreover, for any $v, s \in \mathcal{U}$ such that $v \rightsquigarrow s$, let l_{vs} denote the length of the unique directed path from v to s in $\mathcal{P}(\mathcal{U}, \mathcal{H})$; we let $l_{vs} = 0$ if $v = s$. For example, in the information graph (associated with Example 1) given in Fig. 2, we have $\mathcal{L} = \{\{1\}, \{1, 2\}, \{3\}\}$, $\mathcal{L}_{\{1,2,3\}} = \{\{1\}, \{1, 2\}\}$, and $l_{\{1\}\{1,2,3\}} = 1$.

Next, in order to leverage the structure of the optimal policy $u^\star(\cdot)$ given in (8)–(11), we substitute (submatrices of) $\hat{A}$ and $\hat{B}$ into the right-hand sides of (8) and (9), and obtain $\hat{K}_r$ and $\hat{P}_r$ for all $r \in \mathcal{U}$. Specifically, for all $r \in \mathcal{U}$, we obtain $\hat{K}_r$, and $\hat{P}_r$ recursively as

$$\hat{K}_r = -(R_{rr} + \hat{B}_{sr}^\top \hat{P}_s \hat{B}_{sr}^\top)^{-1} \hat{B}_{sr}^\top \hat{P}_s \hat{A}_{sr} \tag{17}$$

$$\hat{P}_r = Q_{rr} + \hat{K}_r^\top R_{rr} \hat{K}_r + (\hat{A}_{sr} + \hat{B}_{sr}\hat{K}_r)^\top \hat{P}_s (\hat{A}_{sr} + \hat{B}_{sr}\hat{K}_r) \tag{18}$$

where for each $r \in \mathcal{U}$, we let $s \in \mathcal{U}$ be the unique node such that $r \to s$, and $\hat{A}_{sr}$ (resp., $\hat{B}_{sr}$) is a submatrix of $\hat{A}$ (resp., $\hat{B}$) obtained in the same manner as A_{sr} (resp., B_{sr}) described earlier. Similarly to (10), we then use $\hat{K}_r$ for all $r \in \mathcal{U}$ together with $\hat{A}$ and $\hat{B}$ to maintain an (estimated) internal state $\hat{\zeta}_r(t)$ for all $r \in \mathcal{U}$ and for all $t \in \{0, \dots, T-1\}$, which, via a similar form to (11), will lead to our control policy, denoted as $\hat{u}_i(t)$, for all $i \in \mathcal{V}$ and for all $t \in \{0, \dots, T-1\}$. Specifically, for all $i \in \mathcal{V}$ in parallel, we propose Algorithm 2 to compute the control policy

$$\hat{u}_i(t) = \sum_{r \ni i} I_{\{i\},r} \hat{K}_r \hat{\zeta}_r(t) \quad \forall t \in \{0, \dots, T-1\}. \tag{19}$$

We now describe the notations used in Algorithm 2 and hereafter. Let us consider any $i \in \mathcal{V}$. In Algorithm 2, we let $\mathcal{T}_i$ denote the set of disconnected directed trees in $\mathcal{P}(\mathcal{U}, \mathcal{H})$ such that the root node of any tree in $\mathcal{T}_i$ contains i. Slightly abusing the notation, we also let $\mathcal{T}_i$ denote the set of nodes of all the trees in $\mathcal{T}_i$. Moreover, we denote

$$\mathcal{L}(\mathcal{T}_i) = \mathcal{T}_i \cap \mathcal{L} \tag{20}$$

where $\mathcal{L}$ is defined in (15), i.e., $\mathcal{L}(\mathcal{T}_i)$ is the set of leaf nodes of all the trees in $\mathcal{T}_i$. Letting $\mathcal{R} \subseteq \mathcal{U}$ be the set of root nodes in $\mathcal{P}(\mathcal{U}, \mathcal{H})$, we denote

$$\mathcal{R}(\mathcal{T}_i) = \mathcal{T}_i \cap \mathcal{R} \tag{21}$$

where we recall from Lemma 1 that any root node in $\mathcal{P}(\mathcal{U}, \mathcal{H})$ has a self loop. We then see from the information graph $\mathcal{P}(\mathcal{U}, \mathcal{H})$ given in Fig. 2 that

$$\mathcal{L}(\mathcal{T}_1) = \{\{1\}, \{1, 2\}, \{3\}\},\ \mathcal{L}(\mathcal{T}_2) = \mathcal{L}(\mathcal{T}_3) = \{\{1, 2\}, \{3\}\}$$
$$\mathcal{R}(\mathcal{T}_1) = \{\{1\}, \{1, 2, 3\}\},\ \mathcal{R}_2(\mathcal{T}_2) = \mathcal{R}(\mathcal{T}_3) = \{1, 2, 3\}.$$

Note that if any node $s \in \mathcal{T}_i$ is a leaf node with a self loop (i.e., s is an isolated node in $\mathcal{P}(\mathcal{U}, \mathcal{H})$), we only include s in $\mathcal{L}(\mathcal{T}_i)$ (i.e., $s \in \mathcal{L}(\mathcal{T}_i)$ but $s \notin \mathcal{R}(\mathcal{T}_i)$).

Remark 3: For any $s, r \in \mathcal{L}(\mathcal{T}_i)$, let $j_1, j_2 \in \mathcal{V}$ be such that $s_{j_1}(0) = s$ and $s_{j_2}(0) = r$. In Algorithm 2, we assume that the elements in $\mathcal{L}(\mathcal{T}_i)$ are already ordered such that if $D_{ij_1} > D_{ij_2}$, then s comes before r in $\mathcal{L}(\mathcal{T}_i)$. We then let the for loop in lines 5–9 in Algorithm 2 iterate over the elements in $\mathcal{L}(\mathcal{T}_i)$ according to the abovementioned order.

Furthermore, we denote

$$D_{\max} = \max_{\substack{i,j \in \mathcal{V} \\ j \rightsquigarrow i}} D_{ij} \tag{22}$$

where we write $j \rightsquigarrow i$ if and only if there is a directed path from node j to node i in $\mathcal{G}(\mathcal{V}, \mathcal{A})$, and recall that D_{ij} is the sum of delays along the directed path from j to i with the smallest accumulative delay. Finally, the memory $\mathcal{M}_i$ of Algorithm 2 is initialized as $\mathcal{M}_i = \bar{\mathcal{M}}_i$ with

$$\bar{\mathcal{M}}_i = \big\{\hat{\zeta}_s(k) : k \in \{-2D_{\max} - 1, \dots, -D_{ij} - 1\}, s \in \mathcal{L}(\mathcal{T}_i), j \in \mathcal{V}, s_j(0) = s\big\} \cup \big\{\hat{\zeta}_s(-D_{\max} - 1) : s \in \mathcal{R}(\mathcal{T}_i)\big\} \tag{23}$$

where we initialize $\hat{\zeta}_s(k) = 0$ for all $\hat{\zeta}_s(k) \in \bar{\mathcal{M}}_i$.

Remark 4: Considering the scenario with only sparsity constraints (e.g., [1]), i.e., all the edges in $\mathcal{G}(\mathcal{V}, \mathcal{A})$ have a zero delay, we see that $D_{ij} = 0$ for all $i, j \in \mathcal{V}$ such that $j \rightsquigarrow i$, which implies via (22) that $D_{\max} = 0$.

For any $r \ni i$, the dynamics of the internal state $\hat{\zeta}_r(t)$ is given by

$$\hat{\zeta}_r(t+1) = \sum_{v \to r} (\hat{A}_{rv} + \hat{B}_{rv}\hat{K}_v)\hat{\zeta}_v(t) + \sum_{w_j \to r} I_{r,\{j\}}\hat{w}_j(t) \tag{24}$$

where $\hat{w}_j(t)$ is an estimate of the disturbance $w_j(t)$ in (2) obtained as

$$\hat{w}_j(t) = \begin{cases} 0 \text{ if } t < -1 \\ x_j(0) \text{ if } t = -1 \\ x_j(t+1) - \hat{A}_j x_{\mathcal{N}_j}(t) - \hat{B}_j \hat{u}_{\mathcal{N}_j}(t) \text{ if } t \geq 0 \end{cases} \tag{25}$$

where we replace A_j and B_j with the estimates $\hat{A}_j$ and $\hat{B}_j$ in (2), respectively, and $\hat{u}_{\mathcal{N}_j}(t)$ is the vector that collects $\hat{u}_{j_1}(t)$ for all $j_1 \in \mathcal{N}_j$, with $\mathcal{N}_j$ given in Assumption 1. We note from (24) to (25) that $\hat{\zeta}_r(0) = \sum_{w_j \to r} I_{r,\{j\}} x_j(0)$, where $x(0) = 0$ as we assumed previously. We emphasize that (24) and (25) are the keys to our control policy design, and they also enable our analyzes in Section V, where we provide a suboptimality guarantee of our control policy. As we mentioned in Section I, the motivation of the control policy $\hat{u}(\cdot)$ given by (19), (24), and (25) is that the optimal control policy given in Lemma 2 can be viewed as a disturbance-feedback controller. Since the system matrices A and B are unknown, the control policy $\hat{u}(\cdot)$ constructed in (19), (24), and (25) maps the estimates of the past disturbances given by (25) to the current control input via the estimated internal states given by (24).

Observation 1: From the structure of the information graph $\mathcal{P}(\mathcal{U}, \mathcal{H})$ defined in (7), the following hold.

(a) If r is not a leaf node in $\mathcal{P}(\mathcal{U}, \mathcal{H})$, (24) reduces to $\hat{\zeta}_r(t+1) = \sum_{v \to r} (\hat{A}_{rv} + \hat{B}_{rv}\hat{K}_v)\hat{\zeta}_v(t)$.

(b) If r is a leaf node in $\mathcal{P}(\mathcal{U}, \mathcal{H})$ that is not isolated, (24) reduces to $\hat{\zeta}_r(t+1) = \sum_{w_j \to r} I_{r,\{j\}}\hat{w}_j(t)$.

(c) If r is an isolated node in $\mathcal{P}(\mathcal{U}, \mathcal{H})$, (24) reduces to $\hat{\zeta}_r(t+1) = (\hat{A}_{rr} + \hat{B}_{rr}\hat{K}_r)\hat{\zeta}_r(t) + \sum_{w_j \to r} I_{r,\{j\}}\hat{w}_j(t)$.

We will show that in each iteration $t \in \{0, \ldots, T-1\}$ of the for loop in lines 4–14 of Algorithm 2, the internal states $\hat{\zeta}_r(t)$ for all $r \in \mathcal{U}$ such that $i \in r$ (i.e., for all $r \ni i$) can be determined, via (24), based on the current memory $\mathcal{M}_i$ of the algorithm and the state information contained in (a subset of) the information set $\mathcal{I}_i(t)$ defined in (4). As we will see, Algorithm 2 maintains, in its current memory $\mathcal{M}_i$, the internal states (with potential time delays) for a certain subset of nodes in $\mathcal{U}$, via the recursion in (24). Given those internal states, $\hat{\zeta}_r(t)$ for all $r \ni i$ can be determined using (24). Moreover, the memory $\mathcal{M}_i$ of Algorithm 2 is recursively updated in the for loop in lines 4–14 of the algorithm.

Proposition 2: Suppose any controller $i \in \mathcal{V}$ at any time step $t \in \mathbb{Z}_{\geq 0}$ has access to the states in $\tilde{\mathcal{I}}_i(t) \subseteq \mathcal{I}_i(t)$ defined as

$$\tilde{\mathcal{I}}_i(t) = \{x_j(k) : j \in \mathcal{V}_i, k \in \{t - D_{\max} - 1, \ldots, t - D_{ij}\}\} \tag{26}$$

where $\mathcal{V}_i = \{j \in \mathcal{V} : D_{ij} \neq +\infty\}$, and $\mathcal{I}_i(t)$ is defined in (4). Then, the following properties hold for Algorithm 2.

(a) The memory $\mathcal{M}_i$ of Algorithm 2 can be recursively updated such that at the beginning of any iteration $t \in \{0, \ldots, T-1\}$ of the for loop in lines 4–14 of the algorithm

$$\begin{aligned} \mathcal{M}_i = \big\{&\hat{\zeta}_s(k) : k \in \{t - 2D_{\max} - 1, \ldots, t - D_{ij} - 1\}, \\ &s \in \mathcal{L}(\mathcal{T}_i), j \in \mathcal{V}, s_j(0) = s\big\} \\ &\cup \big\{\hat{\zeta}_s(t - D_{\max} - 1) : s \in \mathcal{R}(\mathcal{T}_i)\big\} \end{aligned} \tag{27}$$

(b) The control input $\hat{u}_i(t)$ in line 13 can be determined using (24) and the states in the memory $\mathcal{M}_i$ after line 12 (and before line 14) in any iteration $t \in \{0, \ldots, T-1\}$ of the for loop in lines 4–14 of Algorithm 2.

Since the proof of Proposition 2 is technical and requires careful investigations of the structures of the directed graph $\mathcal{G}(\mathcal{V}, \mathcal{A})$ and the information graph $\mathcal{P}(\mathcal{U}, \mathcal{H})$ described in Section II-B, we use Example 1 to illustrate the steps of Algorithm 2 and the results and proof ideas of Proposition 2. The complete proof can be found in [37].

First, we note from Fig. 1 and (22) that $D_{\max} = 1$. Consider Algorithm 2 with respective to node 2 in the directed graph $\mathcal{G}(\mathcal{V}, \mathcal{A})$ given in Fig. 1. We see that $\mathcal{V}_2 = \{j \in \mathcal{V} : D_{ij} \neq \infty\} = \{2, 3\}$, which implies via (26) that $\tilde{\mathcal{I}}_2(t) = \{x_2(t-2), x_2(t-1), x_2(t), x_3(t-2), x_3(t-1)\}$ for all $t \in \{0, \ldots, T-1\}$. One can check that the initial memory $\bar{\mathcal{M}}_2$ of Algorithm 2 given by (23) satisfies (27) for $t = 0$, which implies that the memory $\mathcal{M}_2$ satisfies (27) at the beginning of iteration $t = 0$ of the for loop in lines 4–14.

To proceed, let us consider iteration $t = 0$ of the for loop in lines 4–14 of the algorithm. Noting that $\mathcal{L}(\mathcal{T}_2) = \{\{3\}, \{1, 2\}\}$ from Remark 3, Algorithm 2 first considers $s = \{3\}$ in the for loop in lines 5–9, which implies $j = 3$ in line 7. We see from (25) that in order to obtain $\hat{w}_3(t-2)$, we need to know $x_3(t-1)$, $x_3(t-2)$, $x_2(t-2)$, $\hat{u}_3(t-2)$, and $\hat{u}_2(t-2)$, where $x_3(t-1), x_3(t-2), x_2(t-2) \in \tilde{\mathcal{I}}_2(t)$, and $\hat{u}_2(t-2)$, $\hat{u}_3(t-2)$ are given by (19). One can then check that the internal states $\hat{\zeta}_r(t')$ that are needed to determine $\hat{u}_2(t-2)$ and $\hat{u}_3(t-2)$ are available in the current memory $\mathcal{M}_2$ of Algorithm 2 or become available via further applications of (24). After $\hat{w}_3(t-2)$ is obtained, we see from (24) that $\hat{\zeta}_{\{3\}}(t-1)$ can also be obtained. Algorithm 2 then updates its current memory $\mathcal{M}_2$ in line 9 and finishes the iteration with respect to $s = \{3\}$ in the for loop in lines 5–9. Next, Algorithm 2 considers $s = \{1, 2\}$ in the for loop in lines 5–9, which implies $j = 2$ in line 7. Following similar arguments to those above and noting that the current memory $\mathcal{M}_2$ of Algorithm 2 has been updated, one can show that $\hat{\zeta}_{\{1,2\}}(t)$ can be obtained from (24), based on the current memory of the algorithm. Algorithm 2 again updates its current memory $\mathcal{M}_2$ in line 9 and finishes the iteration with respect to $s = \{1, 2\}$ in the for loop in lines 5–9.

Now, recalling that $\mathcal{R}(\mathcal{T}_2) = \{1, 2, 3\}$ from Fig. 2, we see that Algorithm 2 considers $s = \{1, 2, 3\}$ in line 10. One can also check that $\hat{\zeta}_{\{1,2,3\}}(t)$ can be obtained from (24), based on the current memory of the algorithm. Finally, based on the current memory $\mathcal{M}_2$ of Algorithm 2 after line 12, one can check that the control input $\hat{u}_2(t)$ can be determined from (19). Note that Algorithm 2 removes certain internal states from its current memory in line 14 that will no longer be used. One can check that after the removal, the current memory $\mathcal{M}_2$ of Algorithm 2 will satisfy (27) at the beginning of iteration $t+1$ of the for

loop in lines 4–14 of the algorithm, where $t = 0$. One can then repeat the above arguments for iterations $t = 1, \ldots, T-1$.

Several remarks pertaining to Algorithm 2 are now in order. First, since $|\mathcal{L}(\mathcal{T}_i)| \leq p$ and $|\mathcal{R}(\mathcal{T}_i)| \leq p$, one can show via the definition of Algorithm 2 that the number of the states in the memory $\mathcal{M}_i$ of Algorithm 2 is always upper bounded by $(2D_{\max}+2)p+2p$, where we note that $D_{\max}$ defined in (22) satisfies $D_{\max} \leq p$, and p is the number of nodes in the directed graph $\mathcal{G}(\mathcal{V}, \mathcal{A})$. Moreover, one can check that Algorithm 2 can be implemented in polynomial time.

Second, it is worth noting that the control policy $\hat{u}_i(\cdot)$ for all $i \in \mathcal{U}$ that we proposed in (19) is related to the certainty equivalence approach (e.g., [43]) that has been used for learning centralized LQR without any information constraints on the controllers (e.g., [12], [13], [36]). It is known that the optimal solution to classic centralized LQR (i.e., problem (5) without the information constraints) is given by a static state-feedback controller $u^\star(t) = Kx(t)$, where K can be obtained from the solution to the Ricatti equation corresponding to A, B, Q, and R (e.g., [44]). The corresponding certainty equivalent controller simply takes the form $\hat{u}(t) = \hat{K}x(t)$, where $\hat{K}$ is obtained from the solution to the Ricatti equation corresponding to $\hat{A}$, $\hat{B}$, Q, and R, with $\hat{A}$ and $\hat{B}$ to be the estimates of A and B, respectively. While we also leverage the structure of the optimal control policy $u^\star(\cdot)$ given in (11), we cannot simply replace K_r with $\hat{K}_r$ for all $r \in \mathcal{U}$ in (11), where $\hat{K}_r$ is given by the Ricatti equations in (17)–(18). As we argued in Remark 2, this is because $u^\star(\cdot)$ is not a static state-feedback controller, but a linear dynamic controller based on the internal states $\zeta_r(\cdot)$ for all $r \in \mathcal{U}$, where the dynamics of $\zeta_r(\cdot)$ given by (10) also depends on A and B. Thus, the control policy $\hat{u}_i(\cdot)$ that we proposed in (19) is a linear dynamic controller based on $\hat{K}_r$ and the estimated internal states $\hat{\zeta}_r(\cdot)$ for all $r \in \mathcal{U}$, where the dynamics of $\hat{\zeta}_r(\cdot)$ given by (24) depends on $\hat{A}$ and $\hat{B}$. Such a more complicated form of $\hat{u}_i(\cdot)$ also creates several challenges when we analyze the corresponding suboptimality guarantees in the next section.

V. Suboptimality Guarantees

In this section, we characterize the suboptimality guarantees of the control policy $\hat{u}(\cdot)$ proposed in Section IV. To begin with, in order to explicitly distinguish the states of the system in (3) corresponding to the control policies $u^\star(\cdot)$ and $\hat{u}(\cdot)$ given by (11) and (19), respectively, we let $\hat{x}(t)$ denote the state of the system in (3) corresponding to the control policy $\hat{u}(\cdot)$ given by (19), for $t \in \mathbb{Z}_{\geq 0}$, i.e.,

$$\hat{x}(t+1) = A\hat{x}(t) + B\hat{u}(t) + w(t) \tag{28}$$

where we note from (19) that $\hat{u}(t) = \sum_{s\in\mathcal{U}} I_{\mathcal{V},s}\hat{K}_s\hat{\zeta}_s(t)$ with $\hat{K}_s$ and $\hat{\zeta}_s(t)$ given by (17) and (24), respectively, for all $s \in \mathcal{U}$. We let $x(t)$ denote the state of the system in (3) corresponding to the optimal control policy $u^\star(t)$ given by (11), for $t \in \mathbb{Z}_{\geq 0}$, i.e.,

$$x(t+1) = Ax(t) + Bu^\star(t) + w(t) \tag{29}$$

where $u^\star(t) = \sum_{s\in\mathcal{U}} I_{\mathcal{V},s}K_s\zeta_s(t)$ with K_s and $\zeta_s(t)$ given by (8) and (10), respectively, for all $s \in \mathcal{U}$. In (28)–(29), we set $\hat{x}(0) = x(0) = 0$. Moreover, for our analysis in the sequel, we introduce another control policy $\tilde{u}(t)$ given by

$$\tilde{u}_i(t) = \sum_{s\ni i} I_{\{i\},s}\hat{K}_s\tilde{\zeta}_s(t) \quad \forall i \in \mathcal{V} \tag{30}$$

for $t \in \mathbb{Z}_{\geq 0}$, where for any $s \in \mathcal{U}$, $\hat{K}_s$ is given by (17), and $\tilde{\zeta}_s(t)$ is given by

$$\tilde{\zeta}_s(t+1) = \sum_{r\to s}(A_{sr} + B_{sr}\hat{K}_r)\tilde{\zeta}_r(t) + \sum_{w_i\to s} I_{s,\{i\}}w_i(t) \tag{31}$$

with $\tilde{\zeta}_s(0) = \sum_{w_i\to s} I_{s,\{i\}}x_i(0) = 0$. We then let $\tilde{x}(t)$ denote the state of the system in (3) corresponding to $\tilde{u}_i(\cdot)$, for $t \in \mathbb{Z}_{\geq 0}$, i.e.,

$$\tilde{x}(t+1) = A\tilde{x}(t) + B\tilde{u}(t) + w(t) \tag{32}$$

where $\tilde{u}(t) = \sum_{s\in\mathcal{U}} I_{\mathcal{V},s}\hat{K}_s\tilde{\zeta}_s(t)$ from (30), and we set $\tilde{x}(0) = x(0) = 0$. Roughly speaking, the auxiliary control policy $\tilde{u}_i(\cdot)$ and the corresponding internal state $\tilde{\zeta}_s(\cdot)$ introduced earlier allow us to decompose the suboptimality gap $\hat{J} - J_\star$ of the control policy $\hat{u}(\cdot)$ into two terms that are due to $\hat{K}_s$ and $\hat{\zeta}_s(\cdot)$, respectively, for all $s \in \mathcal{V}$. We then have the following result; the proof follows directly from [7, Lemma 14] and is, thus, omitted.

Lemma 3: For any $t \in \mathbb{Z}_{\geq 0}$, the following hold: (a) $\mathbb{E}[\tilde{\zeta}_s(t)] = 0$, for all $s \in \mathcal{U}$; (b) $\tilde{x}(t) = \sum_{s\in\mathcal{U}} I_{\mathcal{V},s}\tilde{\zeta}_s(t)$; (c) $\tilde{\zeta}_{s_1}(t)$ and $\tilde{\zeta}_{s_2}(t)$ are independent for all $s_1, s_2 \in \mathcal{U}$ with $s_1 \neq s_2$.

Using the abovementioned notations, the cost of the optimization problem in (5) corresponding to the control policy $\hat{u}(\cdot)$ (i.e., $\hat{J}$) can be written as

$$\hat{J} = \limsup_{T\to\infty} \mathbb{E}\left[\frac{1}{T}\sum_{t=0}^{T-1}\left(\hat{x}(t)^\top Q\hat{x}(t) + \hat{u}(t)^\top R\hat{u}(t)\right)\right] \tag{33}$$

where we use $\limsup$ instead of $\lim$ since the limit may not exist. Furthermore, we let $\tilde{J}$ denote the cost of the optimization problem in (5) corresponding to the control policy $\tilde{u}(\cdot)$ given in (30)

$$\tilde{J} = \lim_{T\to\infty} \mathbb{E}\left[\frac{1}{T}\sum_{t=0}^{T-1}\left(\tilde{x}(t)^\top Q\tilde{x}(t) + \tilde{u}(t)^\top R\tilde{u}(t)\right)\right]. \tag{34}$$

Now, we recall from Lemma 2 that for any $s \in \mathcal{U}$ that has a self loop, the matrix $A_{ss} + B_{ss}K_s$ is stable, where K_s is given by (8). We then have from the Gelfand formula that for any $s \in \mathcal{U}$ that has a self loop, there exist $\kappa_s \in \mathbb{R}_{\geq 1}$ and $\gamma_s \in \mathbb{R}$ with $\rho(A_{ss} + B_{ss}K_s) < \gamma_s < 1$ such that $\|(A_{ss} + B_{ss}K_s)^k\| \leq \kappa_s\gamma_s^k$ for all $k \in \mathbb{Z}_{\geq 0}$. For notational simplicity, let us denote

$$\gamma = \max\left\{\max_{s\in\mathcal{R}}\gamma_s, \gamma_0\right\}, \ \kappa = \max\left\{\max_{s\in\mathcal{R}}\kappa_s, \kappa_0\right\} \tag{35}$$

where $\mathcal{R} \subseteq \mathcal{U}$ denotes the set of root nodes in $\mathcal{U}$, and $\kappa_0 \in \mathbb{R}_{\geq 1}$ and $\gamma_0 \in \mathbb{R}$ with $\rho(A) < \gamma_0 < 1$ are given in Assumption 3. Thus, we see from Assumption 3 and the abovementioned arguments that $\|(A_{ss} + B_{ss}K_s)^k\| \leq \kappa\gamma^k$ for all $s \in \mathcal{R}$ and for all $k \in \mathbb{Z}_{\geq 0}$, and $\|A^k\| \leq \kappa\gamma^k$ for all $k \in \mathbb{Z}_{\geq 0}$, where $\kappa \in \mathbb{Z}_{\geq 1}$ and $0 < \gamma < 1$. Moreover, we denote

$$\Gamma = \max\left\{\|A\|, \|B\|, \max_{s\in\mathcal{U}}\|P_s\|, \max_{s\in\mathcal{U}}\|K_s\|\right\} \tag{36}$$

$$\tilde{\Gamma} = \Gamma + 1.$$

Assumption 4: The cost matrices R and Q in (5) satisfy that $\sigma_n(R) \geq 1$ and $\sigma_m(Q) \geq 1$.

Note that the abovementioned assumption is not more restrictive than assuming that R and Q are positive definite (e.g., [12], [13], [26]). Specifically, supposing $R \succ 0$ and $Q \succ 0$, one can assume without loss of generality that $\sigma_n(R) \geq 1$ and $\sigma_m(Q) \geq 1$. This is because one can check that scaling the objective function in (5) by a positive constant does not change K_r in the optimal solution to (5) provided in Lemma 2, for any $r \in \mathcal{U}$.

A. Perturbation Bounds on Solutions to Ricatti Equations

Supposing $\|\hat{A} - A\| \leq \varepsilon$ and $\|\hat{B} - B\| \leq \varepsilon$ with $\varepsilon \in \mathbb{R}_{>0}$, in this section, we aim to provide upper bounds on the perturbations $\|\hat{P}_r - P_r\|$ and $\|\hat{K}_r - K_r\|$ for all $r \in \mathcal{U}$, where P_r (resp., $\hat{P}_r$) is given by (9) [resp., (18)], and K_r (resp., $\hat{K}_r$) is given by (8) [resp., (17)]. First, we note from Lemma 2 that for any $r \in \mathcal{U}$ that has a self loop, (9) [resp., (18)] reduces to the Ricatti equation in P_r (resp., $\hat{P}_r$). The following results characterize the bounds on $\|\hat{P}_r - P_r\|$ and $\|\hat{K}_r - K_r\|$, for all $r \in \mathcal{U}$. The proofs can be found in [37], which use ideas from [12], and algebraic manipulations based on (8)–(9) and (17)–(18).

Lemma 4: Suppose Assumptions 2 and 4 hold, and $\|\hat{A} - A\| \leq \varepsilon$ and $\|\hat{B} - B\| \leq \varepsilon$, where $\varepsilon \in \mathbb{R}_{>0}$. Then, for any $r \in \mathcal{U}$ that has a self loop, the following hold:

$$\|\hat{P}_r - P_r\| \leq 6\frac{\kappa^2}{1-\gamma^2}\tilde{\Gamma}^5(1+\sigma_1(R^{-1}))\varepsilon \leq \frac{1}{6} \tag{37}$$

$$\|\hat{K}_r - K_r\| \leq 18\frac{\kappa^2}{1-\gamma^2}\tilde{\Gamma}^8(1+\sigma_1(R^{-1}))\varepsilon \leq 1 \tag{38}$$

and

$$\|(A_{rr} + B_{rr}\hat{K}_r)^k\| \leq \kappa(\frac{\gamma+1}{2})^k \quad \forall k \geq 0 \tag{39}$$

under the assumption that

$$\varepsilon \leq \frac{1}{768}\frac{(1-\gamma^2)^2}{\kappa^4}\tilde{\Gamma}^{-11}(1+\sigma_1(R^{-1}))^{-2} \tag{40}$$

where P_r (resp., $\hat{P}_r$) is given by (9) [resp., (18)], K_r (resp., $\hat{K}_r$) is given by (8) [resp., (17)], γ and κ are defined in (35), and $\tilde{\Gamma}$ is defined in (36).

Lemma 5: Suppose Assumptions 2 and 4 hold, and $\|\hat{A} - A\| \leq \varepsilon$ and $\|\hat{B} - B\| \leq \varepsilon$, where $\varepsilon \in \mathbb{R}_{>0}$. Then, for any $r \in \mathcal{U}$ that does not have a self loop, the following hold:

$$\|\hat{K}_r - K_r\| \leq \frac{18\kappa^2\tilde{\Gamma}^8}{1-\gamma^2}(1+\sigma_1(R^{-1}))(20\tilde{\Gamma}^9\sigma_1(R))^{l_{rs}-1}\varepsilon \leq 1 \tag{41}$$

and

$$\|\hat{P}_r - P_r\| \leq \frac{6\kappa^2\tilde{\Gamma}^5}{1-\gamma^2}(1+\sigma_1(R^{-1}))(20\tilde{\Gamma}^9\sigma_1(R))^{l_{rs}}\varepsilon \leq \frac{1}{6} \tag{42}$$

under the assumption that

$$\varepsilon \leq \frac{(1-\gamma^2)^2}{768\kappa^4}\tilde{\Gamma}^{-11}(1+\sigma_1(R^{-1}))^{-2}(20\tilde{\Gamma}^9\sigma_1(R))^{-D_{\max}} \tag{43}$$

where l_{rs} is the length of the unique directed path from node r to node s in $\mathcal{P}(\mathcal{U}, \mathcal{H})$ with $s \in \mathcal{U}$ to be the unique root node that is reachable from r, and $D_{\max}$ is defined in (22).

Consider any $r \in \mathcal{U}$ with a self loop and suppose (40) holds. One can show via (39) and [12, Lemma 12] that $\hat{K}_r$ given by (17) is also stabilizing for the pair $(\hat{A}_{rr}, \hat{B}_{rr})$, i.e., $\hat{A}_{rr} + \hat{B}_{rr}\hat{K}_r$ is stable (see our arguments for (72) in the Appendix for more details). Moreover, it is well-known (e.g., [44]) that a stabilizing solution $\hat{P}_r$ to the Ricatti equation in (18) exists if and only if $(\hat{A}_{rr}, \hat{B}_{rr})$ is stabilizable and $(\hat{A}_{rr}, C_{rr})$ (with $Q_{rr} = C_{rr}^T C_{rr}$) is detectable.[6] The abovementioned arguments together also imply that $(\hat{A}_{rr}, \hat{B}_{rr})$ is stabilizable and $(\hat{A}_{rr}, C_{rr})$ (with $Q_{rr} = C_{rr}^\top C_{rr}$) is detectable for all $r \in \mathcal{U}$, under the assumption on ε given by (40).

B. Perturbation Bounds on Costs

Suppose $\|\hat{A} - A\| \leq \varepsilon$ and $\|\hat{B} - B\| \leq \varepsilon$, where $\varepsilon \in \mathbb{R}_{>0}$. In this section, we aim to provide an upper bound on $\hat{J} - J_\star$, where $J_\star$ and $\hat{J}$ are given by (12) and (33), respectively. To this end, we first provide upper bounds on $\tilde{J} - J_\star$ and $\hat{J} - \tilde{J}$, where $\tilde{J}$ is given by (34), which will lead to the upper bound on $\hat{J} - J_\star$. We start with the following result; the proof can be found in the Appendix.

Lemma 6: Suppose Assumptions 2 and 4 hold, and $\|\hat{A} - A\| \leq \varepsilon$ and $\|\hat{B} - B\| \leq \varepsilon$, where $\varepsilon \in \mathbb{R}_{>0}$ satisfies (43). Then, for any $s \in \mathcal{U}$

$$\lim_{t\to\infty} \mathbb{E}\left[\tilde{\zeta}_s(t)\tilde{\zeta}_s(t)^\top\right] \preceq \frac{4p\sigma_w^2\tilde{\Gamma}^{4D_{\max}}\kappa^2}{1-\gamma^2} I \tag{44}$$

where $p = |\mathcal{V}|$, κ, and γ are defined in (35), $\tilde{\Gamma}$ is defined in (36), and $D_{\max}$ is defined in (22).

For our analysis in the sequel, we further define $\tilde{P}_r$ recursively, for all $r \in \mathcal{U}$, as

$$\tilde{P}_r = Q_{rr} + \hat{K}_r^\top R_{rr}\hat{K}_r + (A_{sr} + B_{sr}\hat{K}_r)^\top \tilde{P}_s(A_{sr} + B_{sr}\hat{K}_r) \tag{45}$$

where $\hat{K}_r$ is given by (17), and $s \in \mathcal{U}$ is the unique node such that $r \to s$. We then have the following result, which gives an upper bound on $\tilde{J} - J_\star$.

Proposition 3: Suppose Assumption 2 and 4 hold, and $\|\hat{A} - A\| \leq \varepsilon$ and $\|\hat{B} - B\| \leq \varepsilon$, where $\varepsilon \in \mathbb{R}_{>0}$ satisfies (43). It holds that

$$\tilde{J} = \sigma_w^2 \sum_{\substack{i\in\mathcal{V} \\ w_i \to s}} \mathrm{Tr}\left(I_{\{i\},s}\tilde{P}_s I_{s,\{i\}}\right) \tag{46}$$

where $\tilde{J}$ is defined in (34). Moreover, consider the optimal cost $J_\star$ given by (12). For any $\varphi \in \mathbb{R}_{>0}$,

$$\tilde{J} - J_\star \leq \frac{72\kappa^4\sigma_w^2 npq}{(1-\gamma^2)^2}\tilde{\Gamma}^{4D_{\max}+8}(\Gamma^3 + \sigma_1(R))(1+\sigma_1(R^{-1})) \times (20\tilde{\Gamma}^9\sigma_1(R))^{D_{\max}}\varepsilon + \varphi \tag{47}$$

where $p = |\mathcal{V}|$ and $q = |\mathcal{U}|$, κ and γ are defined in (35), $\tilde{\Gamma}$ is defined in (36), and $D_{\max}$ is defined in (22).

Proof: First, since ε satisfies (43) [and thus (40)], we have from (39) in Lemma 5 that $A_{ss} + B_{ss}\hat{K}_s$ is stable for any $s \in \mathcal{U}$ that has a self loop. Now, using similar arguments to those for [7, proofs of Theorem 2 and Corollary 4], and leveraging Lemma 3 and (30)–(31), (17), and (45), one can show that (46) holds. Since the proof of (47) is more involved and technical, we defer it

[6] A solution $\hat{P}_r$ to the Ricatti equation in (19) is said to be stabilizing if $\hat{A}_{rr} + \hat{B}_{rr}\hat{K}_r$ [with $\hat{K}_r$ given by (18)] is stable.

to [37] in the interest of space. The proof idea is to first telescope the summation on the right-hand side of (34) properly. Using the telescoped summation corresponding to $\tilde{J}$, one can leverage Lemma 3 and [14, Lemma 12] and show that $\tilde{J} - J_\star$ satisfies that

$$\tilde{J} - J_\star \leq \frac{4p\sigma_w^2 \tilde{\Gamma}^{4D_{\max}} \kappa^2}{1-\gamma^2} \mathrm{Tr}\left(\sum_{r\in\mathcal{U}} (K_r - \hat{K}_r)^\top \times (R_{rr} + B_{sr}^\top P_s B_{sr})(K_r - \hat{K}_r) \right) + \varphi$$

for all $\varphi > 0$. Note that ε is assumed to satisfy (43). Moreover, recalling $|\mathcal{U}| = q$ and $n_i \geq m_i$ for all $i \in \mathcal{V}$ as we assumed previously, and the fact that $\|\hat{K}_r - K_r\| \leq 1$ for all $r \in \mathcal{U}$, one can then use Lemmas 4–5 and show that (47) holds. ■

Next, we aim to provide an upper bound on $\hat{J} - \tilde{J}$. We first prove the following result. The proof can be found in [37], which follows from Lemma 3 and similar arguments to those for the proof of Lemma 6.

Lemma 7: Suppose Assumptions 2 and 4 hold, and $\|\hat{A} - A\| \leq \varepsilon$ and $\|\hat{B} - B\| \leq \varepsilon$, where ε satisfies (43). Then, for any $s \in \mathcal{U}$ and for any $t \in \mathbb{Z}_{\geq 0}$

$$\mathbb{E}\left[\|\tilde{\zeta}_s(t)\|^2\right] \leq \frac{4np\sigma_w^2 \tilde{\Gamma}^{4D_{\max}} \kappa^2}{1-\gamma^2} \tag{48}$$

where $\tilde{\zeta}_s(t)$ is given in (31), $p = |\mathcal{V}|$, κ and γ are defined in (35), $\tilde{\Gamma}$ is defined in (36), and $D_{\max}$ is defined in (22). Moreover, for any $t \in \mathbb{Z}_{\geq 0}$

$$\mathbb{E}\left[\|\tilde{x}(t)\|^2\right] \leq \frac{4npq^2\sigma_w^2 \tilde{\Gamma}^{4D_{\max}} \kappa^2}{1-\gamma^2} \tag{49}$$

and

$$\mathbb{E}\left[\|\tilde{u}(t)\|^2\right] \leq \frac{4npq^2\sigma_w^2 \tilde{\Gamma}^{4D_{\max}+2} \kappa^2}{1-\gamma^2} \tag{50}$$

where $\tilde{x}(t)$ and $\tilde{u}(t)$ are given by (32) and (30), respectively, and $q = |\mathcal{U}|$.

For notational simplicity in the sequel, let us denote

$$\begin{aligned} \zeta_b &= \sqrt{\frac{4np\sigma_w^2 \tilde{\Gamma}^{4D_{\max}} \kappa^2}{1-\gamma^2}} \\ \bar{\varepsilon} &= \frac{(1-\gamma)^3}{768\kappa^4 pq} (\tilde{\Gamma}+1)^{-2} \tilde{\Gamma}^{-9} (1+\sigma_1(R^{-1}))^{-2} \\ &\quad \times (20(\tilde{\Gamma}+1)^2 \tilde{\Gamma}^7 \sigma_1(R))^{-D_{\max}}. \end{aligned} \tag{51}$$

We then have the following results; the proofs of Lemma 8 and Proposition 4 are included in the Appendix.

Lemma 8: Suppose Assumptions 2–4 hold, and $\|\hat{A} - A\| \leq \bar{\varepsilon}$ and $\|\hat{B} - B\| \leq \bar{\varepsilon}$. Then, for all $t \in \mathbb{Z}_{\geq 0}$,

$$\sqrt{\mathbb{E}\left[\|\hat{u}(t) - \tilde{u}(t)\|^2\right]} \leq \frac{58\kappa^2 (\tilde{\Gamma}+1)^{2D_{\max}+3} p^2 q^2}{(1-\gamma)^2} \zeta_b \bar{\varepsilon} \tag{52}$$

and

$$\sqrt{\mathbb{E}\left[\|\hat{x}(t) - \tilde{x}(t)\|^2\right]} \leq \frac{58\kappa^3 \Gamma (\tilde{\Gamma}+1)^{2D_{\max}+3} p^2 q^2}{(1-\gamma)^3} \zeta_b \bar{\varepsilon} \tag{53}$$

where $\hat{u}(t)$ and $\hat{x}(t)$ are given by (19) and (28), respectively.

The abovementioned result also implies the following.

Corollary 1: Suppose Assumptions 2–4 hold. and $\|\hat{A} - A\| \leq \bar{\varepsilon}$ and $\|\hat{B} - B\| \leq \bar{\varepsilon}$. Then, for all $t \in \mathbb{Z}_{\geq 0}$,

$$\sqrt{\mathbb{E}\left[\|\hat{x}(t)\|^2\right]} \leq \frac{58\kappa^3 \Gamma (\tilde{\Gamma}+1)^{2D_{\max}+3} p^2 q^2}{(1-\gamma)^3} \zeta_b \bar{\varepsilon} + q\zeta_b \tag{54}$$

and

$$\sqrt{\mathbb{E}\left[\|\hat{u}(t)\|^2\right]} \leq \frac{58\kappa^2 (\tilde{\Gamma}+1)^{2D_{\max}+3} p^2 q^2}{(1-\gamma)^2} \zeta_b \bar{\varepsilon} + q\tilde{\Gamma}\zeta_b. \tag{55}$$

Proposition 4: Suppose Assumptions 2–4 hold, and $\|\hat{A} - A\| \leq \bar{\varepsilon}$ and $\|\hat{B} - B\| \leq \bar{\varepsilon}$. Then, for $\hat{J}$ and $\tilde{J}$ defined in (33) and (34), respectively,

$$\begin{aligned} \hat{J} - \tilde{J} &\leq \frac{696\kappa^6 \sigma_w^2 n p^4 q^3}{(1-\gamma)^4(1-\gamma^2)} \tilde{\Gamma}^{4D_{\max}+2} (\tilde{\Gamma}+1)^{2D_{\max}+3} \\ &\quad \times (\sigma_1(Q) + \sigma_1(R)) \bar{\varepsilon} \end{aligned} \tag{56}$$

where κ and γ are defined in (35), $p = |\mathcal{V}|$ and $q = |\mathcal{U}|$, $\tilde{\Gamma}$ is defined in (36), and $D_{\max}$ is defined in (22).

Suppose $\|\hat{A} - A\| \leq \varepsilon$ and $\|\hat{B} - B\| \leq \varepsilon$ with $\varepsilon \in \mathbb{R}_{>0}$. We see from the results in Propositions 3–4 that $\hat{J} - J_\star \leq C\varepsilon$ if $\varepsilon \leq C_0$, where C and C_0 are constants that depend on the problem parameters.

C. Sample Complexity Result

We are now in place to present the sample complexity result for learning decentralized LQR with the partially nested information structure described in Section II-B.

Theorem 1: Suppose Assumptions 2–4 hold, and Algorithm 1 is used to obtain $\hat{A}$ and $\hat{B}$. Moreover, suppose $\|A\| \leq \vartheta$ and $\|B\| \leq \vartheta$, where $\vartheta \in \mathbb{R}_{>0}$, and $D_{\max} \leq D$, where $D_{\max}$ is defined in (22) and D is a universal constant. Consider any $\delta > 0$. Let the input parameters to Algorithm 1 satisfy $N \geq \alpha/\bar{\varepsilon}$ and $\lambda \geq \underline{\sigma}^2/40$, where

$$z_b = \frac{5\kappa_0}{1-\gamma_0} \overline{\sigma} \sqrt{(\|B\|^2 m + m + n) \log \frac{4N}{\delta}}$$

and

$$\alpha = \frac{160}{\underline{\sigma}^2} \left(2n\sigma_w^2 (n+m) \log \frac{N + z_b^2/\lambda}{\delta} + \lambda n \vartheta^2 \right)$$

where κ_0 and γ_0 are given in Assumption 3, $\underline{\sigma} = \min\{\sigma_w, \sigma_u\}$, $\overline{\sigma} = \max\{\sigma_w, \sigma_u\}$, and $\bar{\varepsilon}$ is defined in (51). Then, with probability at least $1 - \delta$

$$\begin{aligned} \hat{J} - J_\star &\leq C_1 \frac{\kappa^6 \sigma_w^2 n p^4 q^3}{(1-\gamma^2)^2} \tilde{\Gamma}^{11D+5} (\tilde{\Gamma}+1)^{2D+3} \\ &\quad \times (\Gamma^3 + \sigma_1(R) + \sigma_1(Q)) \sigma_1(R)^D \sqrt{\frac{\alpha}{N}} \end{aligned} \tag{57}$$

where $\hat{J}$ and $J_\star$ are given in (33) and (12), respectively, C_1 is a universal constant, κ and γ are defined in (35), and Γ and $\tilde{\Gamma}$ are defined in (36), $p = |\mathcal{V}|$, and $q = |\mathcal{U}|$.

Proof: Note that the results in Propositions 3–4 hold, if $\|\hat{A} - A\| \leq \bar{\varepsilon}$ and $\|\hat{B} - B\| \leq \bar{\varepsilon}$ with $\bar{\varepsilon}$ given in (51). Thus, letting $N \geq \frac{\alpha}{\bar{\varepsilon}}$ and $\lambda \geq \underline{\sigma}^2/40$, one can first check that $N \geq 200(n+m)\log\frac{48}{\delta}$, and then obtain from Proposition 1 that with probability at least $1 - \delta$, $\hat{A}$ and $\hat{B}$ returned by Algorithm 1 satisfy $\|\hat{A} - A\| \leq \bar{\varepsilon}$ and $\|\hat{B} - B\| \leq \bar{\varepsilon}$. Noting that $D_{\max} \leq D$,

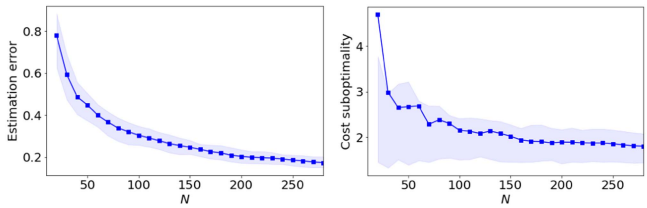
where D is a universal constant, and setting $\varphi = 1/\sqrt{N}$ in Proposition 3, one can show via Propositions 3–4 that (57) holds with probability at least $1-\delta$. ■

Thus, we have shown a $\tilde{\mathcal{O}}(1/\sqrt{N})$ end-to-end sample complexity result for learning decentralized LQR with the partially nested information structure. In other words, we relate the number of data samples used for estimating the system model to the performance of the control policy proposed in Section IV. Note that our result in Theorem 1 matches with the $\mathcal{O}(1/\sqrt{N})$ sample complexity result (up to logarithm factors in N) provided in [13] for learning centralized LQR without any information constraints. Also note that the sample complexity for learning centralized LQR has been improved to $\mathcal{O}(1/N)$ in [12]. Specifically, the authors in [12] showed that the gap between the cost $\hat{J}$ corresponding to the control policy they proposed and the optimal cost $J_\star$ is upper bounded by $\mathcal{O}(\varepsilon^2)$ if ε is sufficiently small, where $\|\hat{A}-A\| \le \varepsilon$ and $\|\hat{B}-B\| \le \varepsilon$. Due to the additional challenges introduced by the information constraints on the controllers (see our discussions at the end of Section IV), we leave investigating the possibility of improving our sample complexity result in Theorem 1 for future work.

VI. Numerical Results

In this section, we illustrate the sample complexity result provided in Theorem 1 with numerical experiments, where the numerical experiments are conducted based on Example 1. Specifically, we consider the LTI system given by (6) with the corresponding directed graph and information graph given by Figs. 1 and 2, respectively. Under the sparsity pattern of A and B specified in (6), we generate the nonzero entries in $A \in \mathbb{R}^{3\times 3}$ and $B \in \mathbb{R}^{3\times 3}$ independently by the Gaussian distribution $\mathcal{N}(0,1)$ while satisfying Assumption 3. We set the covariance of the zero-mean white Gaussian noise process $w(t)$ to be I, and set the cost matrices to be $Q = 2I$ and $R = 5I$. Moreover, we set the input sequence used in the system identification algorithm (Algorithm 1) to be $u(t) \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0,I)$ for all $t \in \{0,\dots,N-1\}$. In order to approximate the value of $\hat{J}$ defined in (33), we simulate the system using Algorithm 2 for $T = 2000$ and obtain $\hat{J} \approx \frac{1}{T}\sum_{t=0}^{T-1}(\tilde{x}(t)^\top Q\tilde{x}(t) + \tilde{u}(t)^\top R\tilde{u}(t))$. Fixing the randomly generated matrices A and B described earlier, the numerical results presented in this section are obtained by averaging over 100 independent experiments.

In Fig. 3(a), we plot the estimation error $\left\|[\hat{A}\quad \hat{B}] - [A\quad B]\right\|$ corresponding to Algorithm 1 when we range the number of the data samples used in Algorithm 1 from $N = 20$ to $N = 280$. Similarly, in Fig. 3(b), we plot the curve corresponding to the cost suboptimality $\hat{J} - J_\star$, where $J_\star$ is obtained by the closed-form expression given in (12). According to Fig. 3, we observe that the estimation error and the cost suboptimality share a similar dependency pattern on N. The similar dependency on N aligns with the results shown in Proposition 1 and Theorem 1 that both the estimation error and the cost suboptimality scale as $\tilde{\mathcal{O}}(1/\sqrt{N})$, which is a consequence of the results shown in Propositions 3–4 that the cost suboptimality scales linearly with the estimation error. The results presented in Fig. 3 then also imply that our suboptimality results provided in Propositions 3–4 can be tight for certain instances of the problem. Finally, we observe from the shaded regions in Fig. 3 that the cost suboptimality is more sensitive to the randomness introduced by the random input $u(t) \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0,I)$ for $t \in \{0,\dots,N-1\}$ and the noise $w(t) \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0,I)$ for $t \in \mathbb{Z}_{\ge 0}$, when we run the 100 experiments described above. This is potentially due to the fact that we approximated the cost suboptimality as $\frac{1}{T}\sum_{t=0}^{T-1}(\hat{x}(t)^\top Q\hat{x}(t) + \hat{u}(t)^\top R\hat{u}(t)) - J_\star$ with $T = 2000$.



(a) (b)

Fig. 3. Both the performance of Algorithm 1 and the performance of Algorithm 2 are plotted against the number of data samples used for estimating the system model, where shaded regions display quartiles. (a) Estimation error versus N (b) Cost suboptimality versus N.

VII. Conclusion

In this article, we considered the problem of control policy design for decentralized state-feedback linear quadratic control with a partially nested information structure, when the system model is unknown. We took a model-based learning approach consisting of two steps. First, we estimated the unknown system model from a single system trajectory of finite length, using least squares estimation. Next, we designed a control policy based on the estimated system model, which satisfies the desired information constraints. We showed that the suboptimality gap between our control policy and the optimal decentralized control policy (designed using accurate knowledge of the system model) scales linearly with the estimation error of the system model. Combining the above results, we provided an end-to-end sample complexity of learning decentralized controllers for state-feedback linear quadratic control with a partially nested information structure.

Appendix
Proofs for Suboptimality Guarantees

Proof of Lemma 6

First, let us consider any $s \in \mathcal{U}$ that has a self loop. Noting the construction of the information graph $\mathcal{P} = (\mathcal{U},\mathcal{H})$ given in (7), one can show that (31) can be rewritten as

$$\tilde{\zeta}_s(t+1) = (A_{ss} + B_{ss}\hat{K}_s)\tilde{\zeta}_s(t) + \sum_{v\in\mathcal{L}_s} H(v,s) \sum_{w_j \to v} I_{v,\{j\}} w_j(t - l_{vs}) \tag{58}$$

where $\mathcal{L}_s = \{v \in \mathcal{L} : v \rightsquigarrow s\}$ is the set of leaf nodes in $\mathcal{P}(\mathcal{U},\mathcal{H})$ that can reach s, l_{vs} is the length of the (unique) directed path from node v to node s in $\mathcal{P}(\mathcal{U},\mathcal{H})$ with $l_{vs} = 0$ if $v = s$, and $H(v,s) \triangleq (A_{sr_1} + B_{sr_1}\hat{K}_{r_1})\cdots(A_{r_{l_{vs}-1}v} + B_{r_{l_{vs}-1}v}\hat{K}_v)$

with $H(v,s)=I$ if $v=s$, where $\hat{K}_r$ is given by (17) for all $r\in\mathcal{U}$, and $v, r_{l_{vs}-1},\dots,r_1,s$ are the nodes along the directed path from v to s in $\mathcal{P}(\mathcal{U},\mathcal{H})$. Recalling from (31) that $\tilde{\zeta}_s(0)=\sum_{w_i\to s} I_{s,\{i\}}x_i(0)=0$, in (58), we set $w_j(t-l_{vs})=0$ if $t-l_{vs}<0$. Now, under the assumption on ε given in (43), we see from (41) in Lemma 5 that $\|\hat{K}_r\|\le\tilde{\Gamma}$, which implies that $\|A_{sr}+B_{sr}\hat{K}_r\|\le\tilde{\Gamma}^2$, for all $r\in\mathcal{U}$ with $r\neq s$. Noting that $l_{vs}\le D_{\max}$ from the construction of $\mathcal{P}(\mathcal{U},\mathcal{H})$, we have $\|H(v,s)\|\le\tilde{\Gamma}^{2D_{\max}}$, for all $v\in\mathcal{L}_s$. Considering any $t\in\mathbb{Z}_{\ge0}$ and denoting

$$\eta_s(t)=\sum_{v\in\mathcal{L}_s}H(v,s)\sum_{w_j\to v}I_{v,\{j\}}w_j(t-l_{vs}) \tag{59}$$

we have

$$\mathbb{E}\left[\eta_s(t)\eta_s(t)^\top\right]=\mathbb{E}\left[\sum_{v\in\mathcal{L}_s}\sum_{w_j\to v}H(v,s)I_{v,\{j\}}w_j(t-l_{vs})\right.$$
$$\left.\times\, w_j(t-l_{vs})^\top I_{\{j\},v}H(v,s)^\top\right]$$

where we use the fact from $w(t)\sim\mathcal{N}(0,\sigma_w^2 I)$ that $w_{j_1}(t)$ and $w_{j_2}(t)$ are independent for all $j_1,j_2\in\mathcal{V}$ with $j_1\neq j_2$, and the fact that for any $v\in\mathcal{U}$ with $s_j(0)=v$, w_j is the only noise term such that $w_j\to v$ (see Footnote 2). Moreover, we see that $\eta_s(t_1)$ and $\eta_s(t_2)$ are independent for all $t_1,t_2\in\mathbb{Z}_{\ge0}$ with $t_1\neq t_2$, and that $\eta_s(t)$ is independent of $\tilde{\zeta}_s(t)$ for all $t\in\mathbb{Z}_{\ge0}$. Now, considering any $k\in\mathbb{Z}_{\ge0}$ such that $k-l_{vs}\ge0$ for all $v\in\mathcal{L}_s$, and noting that $w(t)\sim\mathcal{N}(0,\sigma_w^2I)$ for all $t\in\mathbb{Z}_{\ge0}$, we have

$$\mathbb{E}\left[\eta_s(k)\eta_s(k)^\top\right]=\sigma_w^2\sum_{v\in\mathcal{L}_s}\sum_{w_j\to v}H(v,s)I_{v,\{j\}}I_{\{j\},v}H(v,s)^\top. \tag{60}$$

Let us denote the right-hand side of (60) as $\bar{W}_s$, and denote

$$\tilde{L}_{ss}=A_{ss}+B_{ss}\hat{K}_s.$$

Fixing any $\tau\in\mathbb{Z}_{\ge1}$ such that $\tau-l_{vs}\ge0$ for all $v\in\mathcal{L}_s$, and considering any $t\ge\tau$, one can then unroll (58) and show that

$$\mathbb{E}\left[\tilde{\zeta}_s(t)\tilde{\zeta}_s(t)^\top\right]=\tilde{L}_{ss}^{t-\tau}\mathbb{E}\left[\tilde{\zeta}_s(\tau)\tilde{\zeta}_s(\tau)^\top\right](\tilde{L}_{ss}^\top)^{t-\tau}$$
$$+\sum_{k=0}^{t-\tau-1}L_{ss}^k\bar{W}_s(\tilde{L}_{ss}^\top)^k. \tag{61}$$

Under the assumption on ε given in (43), one can obtain from Lemma 4 that $\|\tilde{L}_{ss}^k\|\le\kappa(\frac{\gamma+1}{2})^k$ for all $k\ge0$, where $0<\frac{\gamma+1}{2}<1$, which implies that $\tilde{L}_{ss}$ is stable. It follows that

$$\lim_{t\to\infty}\mathbb{E}\left[\tilde{\zeta}_s(t)\tilde{\zeta}_s(t)^\top\right]\preceq\frac{4\|\bar{W}_s\|\kappa^2}{1-\gamma^2}I.$$

Noting that $|\mathcal{L}_s|\le p$ from the definition of $\mathcal{P}(\mathcal{U},\mathcal{H})$ given in (7), and that for any $v\in\mathcal{U}$ with $s_j(0)=v$, w_j is the only noise term such that $w_j\to v$, as we argued above, we have from (60) that

$$\|\bar{W}_s\|\le\sigma_w^2 p\max_{v\in\mathcal{L}_s}\|H(v,s)\|^2\le\sigma_w^2p\tilde{\Gamma}^{4D_{\max}}$$

where the second inequality follows from the fact that $\|H(v,s)\|\le\tilde{\Gamma}^{2D_{\max}}$ as we argued above. It then follows that (44) holds.

Next, let us consider any $s\in\mathcal{U}$ that does not have a self loop. Similarly to (58), one can rewrite (31) as

$$\tilde{\zeta}_s(t+1)=\sum_{v\in\mathcal{L}_s}H(v,s)\sum_{w_j\to v}I_{v,\{j\}}w_j(t-l_{vs}).$$

Using similar arguments to those above, one can show that (44) also holds. ■

Proof of Lemma 8

Due to the space constraint, we omitted some technical details in this proof; the complete proof can be found in [37].

For notational simplicity in this proof, we denote

$$\delta_h=p(\tilde{\Gamma}+1)^{2D_{\max}-1},\ \beta=\tilde{\Gamma}^{2D_{\max}}$$

$$\Lambda_1=q\tilde{\Gamma}\left(\frac{2\kappa p(\beta+1)}{1-\gamma}+\frac{32\kappa^2p(\tilde{\Gamma}+1)}{(1-\gamma)^2}\right)\left(1+\frac{\kappa\Gamma}{1-\gamma}\right)$$

and

$$\Lambda_2=\frac{2pq\tilde{\Gamma}\kappa}{1-\gamma}\left((\beta+1)q(\tilde{\Gamma}+1)+\delta_h\right)\zeta_b$$
$$+\frac{16\kappa^2pq\tilde{\Gamma}(\tilde{\Gamma}+1)}{(1-\gamma)^2}\left(2q(\tilde{\Gamma}+1)+2\right)\zeta_b.$$

We first prove (52). Note that $\frac{\kappa}{1-\gamma}>1$, and that $\beta+1\le\tilde{\Gamma}^{2D_{\max}+1}$ since $\tilde{\Gamma}=\Gamma+1\ge2$, where the inequality follows from the fact via (9) and (36) that $\Gamma\ge\sigma_m(Q)\ge1$. Based on the abovementioned notations, one can then show that

$$1.1\Lambda_2\le\frac{58\kappa^2(\tilde{\Gamma}+1)^{2D_{\max}+3}p^2q^2}{(1-\gamma)^2}\zeta_b. \tag{62}$$

Thus, in order to show that (52) holds for all $t\in\mathbb{Z}_{\ge0}$, it suffices to show that $\mathbb{E}[\|u(t)-\tilde{u}(t)\|^2]\le(1.1\Lambda_2\bar{\varepsilon})^2$ holds for all $t\in\mathbb{Z}_{\ge0}$. To this end, we prove via an induction on $t=0,1,\dots$. For any $t\in\mathbb{Z}_{\ge0}$, we recall from (19) and (30) that $\hat{u}_i(t)=\sum_{r\ni i}I_{\{i\},r}\hat{K}_r\hat{\zeta}_r(t)$ and $\tilde{u}_i(t)=\sum_{r\ni i}I_{\{i\},r}\hat{K}_r\tilde{\zeta}_r(t)$, respectively, for all $i\in\mathcal{V}$, where $\hat{\zeta}_r(t)$ and $\tilde{\zeta}_r(t)$ are given by (24) and (31), respectively, and $\hat{K}_r$ is given by (17), for all $r\in\mathcal{U}$. As we argued before, in (24) and (31) we have $\hat{\zeta}_r(0)=\tilde{\zeta}_r(0)=\sum_{w_i\to r}I_{r,\{i\}}x_i(0)$ for all $r\in\mathcal{U}$. Hence, we have $\hat{u}(0)=\tilde{u}(0)$, which implies that (52) holds for $t=0$, completing the proof of the base step of the induction.

For the induction step, suppose $\mathbb{E}[\|\hat{u}(k)-\tilde{u}(k)\|^2]\le(1.1\Lambda_2\bar{\varepsilon})^2$ holds for all $k\in\{0,\dots,t\}$. Now, considering any $k\in\{0,\dots,t\}$, we can unroll the expressions of $\hat{x}(k)$ and $\tilde{x}(k)$ given by (28) and (32), respectively, and obtain

$$\hat{x}(k)=A^k\hat{x}(0)+\sum_{k'=0}^{k-1}A^{k-k'-1}(B\hat{u}(k')+w(k'))$$

$$\tilde{x}(k)=A^k\tilde{x}(0)+\sum_{k'=0}^{k-1}A^{k-k'-1}(B\tilde{u}(k')+w(k'))$$

where we note that $\hat{x}(0)=\tilde{x}(0)=x(0)$. It then follows that

$$\sqrt{\mathbb{E}\left[\|\hat{x}(k)-\tilde{x}(k)\|^2\right]}$$
$$\le\sum_{k'=0}^{k-1}\sqrt{\mathbb{E}\left[\|A^{k-k'-1}B(\hat{u}(k')-\tilde{u}(k'))\|^2\right]}$$

$$\leq \Gamma 1.1\Lambda_2\bar{\varepsilon}\sum_{k'=0}^{k-1}\|A^{k-k'-1}\| \leq 1.1\Gamma\Lambda_2\bar{\varepsilon}\frac{\kappa}{1-\gamma} \tag{63}$$

where the first inequality follows from [37, Lemma 14]. To obtain the first inequality in (63), we use the induction hypothesis. To obtain the second inequality in (63), we use the fact that $\|A^{k'}\| \leq \kappa\gamma^{k'}$ (with $0<\gamma<1$), for all $k' \in \mathbb{Z}_{\geq 0}$, from Assumption 3. Recalling from our arguments in Section IV [particularly, (25)], one can show that

$$\hat{w}(k) = \hat{x}(k+1) - \hat{A}\hat{x}(k) - \hat{B}\hat{u}(k)$$

where $\hat{w}(k) = \begin{bmatrix}\hat{w}_1(k)^\top & \cdots & \hat{w}_p(k)^\top\end{bmatrix}^\top$ is an estimate of $w(k)$ in (3). From (28), we see that

$$w(k) = \hat{x}(k+1) - A\hat{x}(k) - B\hat{u}(k).$$

Recall from Lemma 7 that $\mathbb{E}[\|\tilde{x}(k)\|^2] \leq q^2\zeta_b^2$ and $\mathbb{E}[\|\tilde{u}(k)\|^2] \leq q^2\tilde{\Gamma}^2\zeta_b^2$, for all $k \in \mathbb{Z}_{\geq 0}$. One can use (63) and [37, Lemma 14] and show that

$$\mathbb{E}\left[\|\hat{x}(k)\|^2\right] \leq \left(\frac{1.1\Gamma\Lambda_2\kappa}{1-\gamma}\bar{\varepsilon} + q\zeta_b\right)^2.$$

Moreover, noting the induction hypothesis, one can show that

$$\mathbb{E}\left[\|\hat{u}(k)\|^2\right] \leq (1.1\Lambda_2\bar{\varepsilon} + q\tilde{\Gamma}\zeta_b)^2.$$

Combining the abovementioned arguments and using [37, Lemma 14], one can now show that

$$\mathbb{E}\left[\|\hat{w}(k) - w(k)\|^2\right] \leq \left(q(\tilde{\Gamma}+1)\zeta_b + \frac{1.1\Gamma\Lambda_2\kappa}{1-\gamma}\bar{\varepsilon} + 1.1\Lambda_2\right)^2\bar{\varepsilon}^2.$$

Denoting

$$\delta_w = q(\tilde{\Gamma}+1)\zeta_b + \frac{1.1\Gamma\Lambda_2\kappa}{1-\gamma}\bar{\varepsilon} + 1.1\Lambda_2\bar{\varepsilon} \tag{64}$$

we have

$$\mathbb{E}\left[\|\hat{w}(k) - w(k)\|^2\right] \leq \delta_w^2\bar{\varepsilon}^2 \quad \forall k \in \{0,\ldots,t\}.$$

Moreover, note that

$$\mathbb{E}\left[\|w(k)\|^2\right] = \mathbb{E}\left[\mathrm{Tr}(w(k)w(k)^\top)\right] = \mathrm{Tr}\left(\mathbb{E}\left[w(k)w(k)^\top\right]\right) = n\sigma_w^2 \leq \zeta_b^2 \quad \forall k \in \mathbb{Z}_{\geq 0}.$$

To proceed, let us consider any $s \in \mathcal{U}$ that has a self loop. Recalling the arguments in the proof of Lemmas 6, we can rewrite (31) as

$$\tilde{\zeta}_s(t+1) = (A_{ss} + B_{ss}\hat{K}_s)\tilde{\zeta}_s(t) + \eta_s(t) \tag{65}$$

with

$$\eta_s(t) = \sum_{v\in\mathcal{L}_s} H(v,s)\sum_{w_j\to v} I_{v,\{j\}}w_j(t-l_{vs}) \tag{66}$$

where $\mathcal{L}_s = \{v \in \mathcal{L} : v \rightsquigarrow s\}$ is the set of leaf nodes in $\mathcal{P}(\mathcal{U},\mathcal{H})$ that can reach s, l_{vs} is the length of the (unique) directed path from node v to node s in $\mathcal{P}(\mathcal{U},\mathcal{H})$ with $l_{vs}=0$ if $v=s$, and

$$H(v,s) = (A_{sr_1} + B_{sr_1}\hat{K}_{r_1})\cdots(A_{r_{l_{vs}-1}v} + B_{r_{l_{vs}-1}v}\hat{K}_v)$$

with $H(v,s) = I$ if $v=s$. We also recall from the arguments in the proof of Lemma 6 that $\|H(v,s)\| \leq \beta$ for all $v \in \mathcal{L}_s$. We then see from (60) in the proof of Lemma 6 and the definition of ζ_b in (51) that

$$\mathbb{E}\left[\|\eta_s(k)\|^2\right] = \mathbb{E}\left[\mathrm{Tr}(\eta_s(k)\eta_s(k)^\top)\right] = \mathrm{Tr}\left(\mathbb{E}\left[\eta_s(k)\eta_s(k)^\top\right]\right) \leq \sigma_w^2 np\beta^2 \leq \zeta_b^2 \quad \forall k \in \mathbb{Z}_{\geq 0}.$$

Similarly, one can rewrite (24) as

$$\hat{\zeta}_s(t+1) = (\hat{A}_{ss} + \hat{B}_{ss}\hat{K}_s)\hat{\zeta}_s(t) + \hat{\eta}_s(t) \tag{67}$$

where

$$\hat{\eta}_s(t) = \sum_{v\in\mathcal{L}_s}\hat{H}(v,s)\sum w_j \to vI_{v,\{j\}}\hat{w}_j(t-l_{vs})$$

where

$$\hat{H}(v,s) = (\hat{A}_{sr_1} + \hat{B}_{sr_1}\hat{K}_{r_1})\cdots(\hat{A}_{r_{l_{vs}-1}v} + \hat{B}_{r_{l_{vs}-1}v}\hat{K}_v)$$

with $\hat{H}(v,s) = I$ if $v=s$. For any $k \in \{0,\ldots,t\}$, one can then show via the abovementioned arguments that

$$\sqrt{\mathbb{E}\left[\|\eta_s(k) - \hat{\eta}_s(k)\|^2\right]} \leq (\delta_h\zeta_b\bar{\varepsilon} + (\delta_h\bar{\varepsilon}+\beta)\delta_w\bar{\varepsilon}) \tag{68}$$

and

$$\sqrt{\mathbb{E}\left[\|\hat{\eta}_s(k)\|^2\right]} \leq p\left(\delta_h\zeta_b\bar{\varepsilon} + (\delta_h\bar{\varepsilon}+\beta)\delta_w\bar{\varepsilon}\right) + \zeta_b. \tag{69}$$

Now, let us denote $\tilde{L}_{ss} = A_{ss} + B_{ss}\hat{K}_s$ and $\hat{L}_{ss} = \hat{A}_{ss} + \hat{B}_{ss}\hat{K}_s$. Recalling that $\hat{\zeta}_s(0) = \tilde{\zeta}_s(0) = \sum_{w_i\to s} I_{s,\{i\}}x_i(0)$, where $x(0)=0$ as we assumed before, one can unroll (65) and (67), and show that

$$\hat{\zeta}_s(t+1) - \tilde{\zeta}_s(t+1) = \sum_{k=0}^{t}\left(\hat{L}_{ss}^{t-k}\hat{\eta}_s(k) - \tilde{L}_{ss}^{t-k}\tilde{\eta}_s(k)\right). \tag{70}$$

Since $\|\hat{A}-A\| \leq \bar{\varepsilon}$ and $\|\hat{B}-B\| \leq \bar{\varepsilon}$, where $\bar{\varepsilon}$ satisfies (40), as we argued above, we have from Lemma 4 that

$$\|\tilde{L}_{ss}^k\| \leq \kappa\left(\frac{\gamma+1}{2}\right)^k \quad \forall k \in \mathbb{Z}_{\geq 0} \tag{71}$$

where $\kappa \in \mathbb{R}_{\geq 1}$ and $\gamma \in \mathbb{R}$, with $0<\gamma<1$. Moreover, since $\|\hat{L}_{ss} - \tilde{L}_{ss}\| = \|\hat{A}_{ss} - A_{ss} + \hat{K}_s(\hat{B}_{ss} - B_{ss})\| \leq (\tilde{\Gamma}+1)\bar{\varepsilon}$, one can use [12, Lemma 5] and prove that

$$\|\hat{L}_{ss}^k - \tilde{L}_{ss}^k\| \leq k\kappa^2\left(\frac{\gamma+3}{4}\right)^{k-1}(\tilde{\Gamma}+1)\bar{\varepsilon} \quad \forall k \in \mathbb{Z}_{\geq 0} \tag{72}$$

based on the choice of $\bar{\varepsilon}$ in (51). Combining (68)–(72), one can then show that

$$\sqrt{\mathbb{E}\left[\|\hat{\zeta}_s(t+1) - \tilde{\zeta}_s(t+1)\|^2\right]} \leq \frac{2\kappa p}{1-\gamma}\left((\beta+1)\delta_w + \delta_h\zeta_b\right)\bar{\varepsilon} + \frac{16\kappa^2(\tilde{\Gamma}+1)p}{(1-\gamma)^2}\left(2\delta_w + 2\zeta_b\right)\bar{\varepsilon}. \tag{73}$$

Now, substituting (64) into the right-hand side of (73), one can show that

$$\sqrt{\mathbb{E}\left[\|\hat{\zeta}_s(t+1) - \tilde{\zeta}_s(t+1)\|^2\right]} \leq \frac{1}{q\tilde{\Gamma}}\left(\Lambda_1(1.1\Lambda_2\bar{\varepsilon}) + \Lambda_2\right)\bar{\varepsilon} \tag{74}$$

where we note that $\Lambda_1 > 0$ and $\Lambda_2 > 0$ by their definitions.

Next, considering any $s \in \mathcal{U}$ that does not have a self loop, we have from the arguments in the proof of Lemma 6 that (31) can be rewritten as $\tilde{\zeta}_s(t+1) = \eta_s(t)$, where $\eta_s(t)$ is defined in (66). Using similar arguments to those above, one can then show that (74) also holds.

Further recalling (19) and (30), we know that $\hat{u}(t+1) = \sum_{s\in\mathcal{U}} I_{\mathcal{V},s}\hat{K}_s\hat{\zeta}_s(t+1)$ and $\tilde{u}(t+1) = \sum_{s\in\mathcal{U}} I_{\mathcal{V},s}\hat{K}_s\tilde{\zeta}_s(t+1)$. Using (74), one can show that

$$\sqrt{\mathbb{E}\left[\|\hat{u}(t+1) - \tilde{u}(t+1)\|^2\right]} \leq \left(\Lambda_1(1.1\Lambda_2\bar{\varepsilon}) + \Lambda_2\right)\bar{\varepsilon}.$$

Moreover, one can prove that $\bar{\varepsilon}$ given in (51) satisfies that $0 < \bar{\varepsilon} \leq \frac{1}{11\Lambda_1}$, which further implies that

$$\Lambda_1(1.1\Lambda_2\bar{\varepsilon})\bar{\varepsilon} + \Lambda_2\bar{\varepsilon} \leq 1.1\Lambda_2\bar{\varepsilon}$$

completing the induction step.

Next, using similar arguments to those for (63), we have $\sqrt{\mathbb{E}[\|\hat{x}(t) - \tilde{x}(t)\|^2]} \leq \frac{1.1\Gamma\Lambda_2\kappa}{1-\gamma}\bar{\varepsilon}$ for all $t \in \mathbb{Z}_{\geq 0}$. It then follows from (62) that (53) holds for all $t \in \mathbb{Z}_{\geq 0}$. ■

Proof of Proposition 4

For notational simplicity in this proof, we denote

$$\Lambda = \frac{58\kappa^2(\tilde{\Gamma}+1)^{2D_{\max}+3}p^2q^2}{(1-\gamma)^2}. \tag{75}$$

For all $t \in \mathbb{Z}_{\geq 0}$, we then see from Lemma 8 that

$$\mathbb{E}\left[\|\hat{u}(t) - \tilde{u}(t)\|^2\right] \leq (\Lambda\zeta_b\bar{\varepsilon})^2$$

and

$$\mathbb{E}\left[\|\hat{x}(t) - \tilde{x}(t)\|^2\right] \leq \left(\frac{\kappa\Gamma}{1-\gamma}\Lambda\zeta_b\bar{\varepsilon}\right)^2$$

where $\hat{u}(k)$ (resp., $\tilde{u}(k)$) is given by (19) [resp., (30)], $\hat{x}(k)$ (resp., $\tilde{x}(k)$) is given by (28) [resp., (32)], and ζ_b is defined in (51). Similarly, we see from Corollary 1 that

$$\mathbb{E}\left[\|\hat{x}(t)\|^2\right] \leq \left(\frac{\kappa\Gamma}{1-\gamma}\Lambda\zeta_b\bar{\varepsilon} + q\zeta_b\right)^2$$

and

$$\mathbb{E}\left[\|\hat{u}(t)\|^2\right] \leq (\Lambda\zeta_b\bar{\varepsilon} + q\tilde{\Gamma}\zeta_b)^2$$

for all $t \in \mathbb{Z}_{\geq 0}$. To proceed, we have the following

$$\hat{J} - \tilde{J} = \limsup_{T\to\infty} \mathbb{E}\left[\frac{1}{T}\sum_{t=0}^{T-1}\left(\hat{x}(t)^\top Q\hat{x}(t) - \tilde{x}(t)^\top Q\tilde{x}(t) + \hat{u}(t)^\top R\hat{u}(t) - \tilde{u}(t)^\top R\tilde{u}(t)\right)\right]. \tag{76}$$

Now, considering any term in the summation on the right-hand side of (76), and dropping the dependency on t for notational simplicity, we have the following:

$$\begin{aligned}
&\mathbb{E}\left[\hat{x}^\top Q\hat{x} - \tilde{x}^\top Q\tilde{x}\right] \\
&\leq \mathbb{E}\left[\|Q\hat{x}\|\|\hat{x} - \tilde{x}\|\right] + \mathbb{E}\left[\|\hat{x} - \tilde{x}\|\|Q\tilde{x}\|\right] \\
&\leq \sqrt{\mathbb{E}\left[\|Q\hat{x}\|^2\right]\mathbb{E}\left[\|\hat{x} - \tilde{x}\|^2\right]} + \sqrt{\mathbb{E}\left[\|\hat{x} - \tilde{x}\|^2\right]\mathbb{E}\left[\|Q\tilde{x}\|^2\right]} \\
&\leq \sigma_1(Q)\left(\frac{\kappa\Gamma\Lambda\zeta_b}{1-\gamma}\bar{\varepsilon} + q\zeta_b\right)\frac{\kappa\Gamma\Lambda\zeta_b}{1-\gamma}\bar{\varepsilon} + \sigma_1(Q)\frac{\kappa\Gamma\Lambda\zeta_b}{1-\gamma}\bar{\varepsilon}q\zeta_b \\
&= \sigma_1(Q)\left(\frac{\kappa\Gamma\Lambda\zeta_b}{1-\gamma}\bar{\varepsilon} + 2q\zeta_b\right)\frac{\kappa\Gamma\Lambda\zeta_b}{1-\gamma}\bar{\varepsilon}
\end{aligned} \tag{77}$$

where the first two inequalities follow from the Cauchy–Schwartz inequality, and the third inequality follows from the upper bounds on $\mathbb{E}[\|\hat{x}\|^2]$, $\mathbb{E}[\|\hat{x} - \tilde{x}\|^2]$, and $\mathbb{E}[\|\tilde{x}\|^2]$ given above and in Lemma 7. Similarly, one can prove that

$$\mathbb{E}\left[\hat{u}^\top R\hat{u} - \tilde{u}^\top R\tilde{u}\right] \leq \sigma_1(R)(\Lambda\zeta_b\bar{\varepsilon} + 2q\tilde{\Gamma}\zeta_b)\Lambda\zeta_b\bar{\varepsilon}. \tag{78}$$

Combining (77) and (78) together, we obtain from (76) that

$$\begin{aligned}
\hat{J} - \tilde{J} &\leq \sigma_1(Q)\left(\frac{\kappa\Gamma\Lambda\zeta_b}{1-\gamma}\bar{\varepsilon} + 2q\zeta_b\right)\frac{\kappa\Gamma\Lambda\zeta_b}{1-\gamma}\bar{\varepsilon} \\
&\quad + \sigma_1(R)(\Lambda\zeta_b\bar{\varepsilon} + 2q\tilde{\Gamma}\zeta_b)\Lambda\zeta_b\bar{\varepsilon} \\
&\leq \left(\frac{\kappa\tilde{\Gamma}\zeta_b}{1-\gamma}\right)^2(\Lambda^2\bar{\varepsilon} + 2q\Lambda)(\sigma_1(Q) + \sigma_1(R))\bar{\varepsilon} \\
&\leq \left(\frac{\kappa\tilde{\Gamma}\zeta_b}{1-\gamma}\right)^2 3\Lambda pq(\sigma_1(Q) + \sigma_1(R))\bar{\varepsilon}
\end{aligned} \tag{79}$$

where the second inequality follows from the fact that $\frac{\kappa\tilde{\Gamma}}{1-\gamma} \geq 1$. To obtain (79), one can show that $\Lambda^2\bar{\varepsilon} \leq \Lambda pq$. Finally substituting the expressions for ζ_b and Λ given in (51) and (75), respectively, we obtain from (79) that (56) holds. ■

REFERENCES

[1] P. Shah and P. A. Parrilo, "$\mathcal{H}_2$-optimal decentralized control over posets: A state-space solution for state-feedback," *IEEE Trans. Autom. Control*, vol. 58, no. 12, pp. 3084–3096, Dec. 2013.

[2] A. Lamperski and J. C. Doyle, "Dynamic programming solutions for decentralized state-feedback LQG problems with communication delays," in *Proc. Amer. Control Conf.*, 2012, pp. 6322–6327.

[3] H. S. Witsenhausen, "A counterexample in stochastic optimum control," *SIAM J. Control*, vol. 6, no. 1, pp. 131–147, 1968.

[4] C. H. Papadimitriou and J. Tsitsiklis, "Intractable problems in control theory," *SIAM J. Control Optim.*, vol. 24, no. 4, pp. 639–654, 1986.

[5] V. D. Blondel and J. N. Tsitsiklis, "A survey of computational complexity results in systems and control," *Automatica*, vol. 36, no. 9, pp. 1249–1274, 2000.

[6] Y.-C. Ho et al., "Team decision theory and information structures in optimal control problems—Part I," *IEEE Trans. Autom. control*, vol. AC-17, no. 1, pp. 15–22, Feb. 1972.

[7] A. Lamperski and L. Lessard, "Optimal decentralized state-feedback control with sparsity and delays," *Automatica*, vol. 58, pp. 143–151, 2015.

[8] M. Rotkowitz and S. Lall, "A characterization of convex problems in decentralized control," *IEEE Trans. Autom. Control*, vol. 50, no. 12, pp. 1984–1996, Dec. 2005.

[9] M. C. Rotkowitz and N. C. Martins, "On the nearest quadratically invariant information constraint," *IEEE Trans. Autom. Control*, vol. 57, no. 5, pp. 1314–1319, May 2012.

[10] Z.-S. Hou and Z. Wang, "From model-based control to data-driven control: Survey, classification and perspective," *Inf. Sci.*, vol. 235, pp. 3–35, 2013.

[11] Y. Abbasi-Yadkori and C. Szepesvári, "Regret bounds for the adaptive control of linear quadratic systems," in *Proc. Conf. Learn. Theory*, 2011, pp. 1–26.

[12] H. Mania, S. Tu, and B. Recht, "Certainty equivalence is efficient for linear quadratic control," in *Proc. Int. Conf. Neural Inf. Process. Syst.*, 2019, pp. 10154–10164.

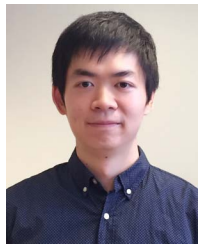
[13] S. Dean, H. Mania, N. Matni, B. Recht, and S. Tu, "On the sample complexity of the linear quadratic regulator," *Foundations Comput. Math.*, vol. 20, no. 4, pp. 633–679, 2020.

[14] M. Fazel, R. Ge, S. Kakade, and M. Mesbahi, "Global convergence of policy gradient methods for the linear quadratic regulator," in *Proc. Int. Conf. Mach. Learn.*, 2018, pp. 1467–1476.

[15] K. Zhang, B. Hu, and T. Basar, "Policy optimization for $\mathcal{H}_2$ linear control with $\mathcal{H}_\infty$ robustness guarantee: Implicit regularization and global convergence," in *Proc. Learn. Dyn. Control Conf.*, 2020, pp. 179–190.

[16] D. Malik, A. Pananjady, K. Bhatia, K. Khamaru, P. L. Bartlett, and M. J. Wainwright, "Derivative-free methods for policy optimization: Guarantees for linear quadratic systems," *J. Mach. Learn. Res.*, vol. 21, no. 21, pp. 1–51, 2020.

[17] B. Gravell, P. M. Esfahani, and T. H. Summers, "Learning optimal controllers for linear systems with multiplicative noise via policy gradient," *IEEE Trans. Autom. Control*, vol. 66, no. 11, pp. 5283–5298, Nov. 2021.
[18] S. Ghadimi and G. Lan, "Stochastic first-and zeroth-order methods for nonconvex stochastic programming," *SIAM J. Optim.*, vol. 23, no. 4, pp. 2341–2368, 2013.
[19] Y. Nesterov and V. Spokoiny, "Random gradient-free minimization of convex functions," *Foundations Comput. Math.*, vol. 17, no. 2, pp. 527–566, 2017.
[20] Y. Zheng, L. Furieri, M. Kamgarpour, and N. Li, "Sample complexity of linear quadratic gaussian (LQG) control for output feedback systems," in *Proc. Learn. Dyn. Control Conf.*, 2021, pp. 559–570.
[21] S. Tu and B. Recht, "The gap between model-based and model-free methods on the linear quadratic regulator: An asymptotic viewpoint," in *Proc. Conf. Learn. Theory*, 2019, pp. 3036–3083.
[22] M. K. S. Faradonbeh, A. Tewari, and G. Michailidis, "Finite time identification in unstable linear systems," *Automatica*, vol. 96, pp. 342–353, 2018.
[23] M. Simchowitz, H. Mania, S. Tu, M. I. Jordan, and B. Recht, "Learning without mixing: Towards a sharp analysis of linear system identification," in *Proc. Conf. On Learn. Theory*, 2018, pp. 439–473.
[24] T. Sarkar and A. Rakhlin, "Near optimal finite time identification of arbitrary linear dynamical systems," in *Proc. Int. Conf. Mach. Learn.*, 2019, pp. 5610–5618.
[25] Y. Abbasi-Yadkori, D. Pál, and C. Szepesvári, "Improved algorithms for linear stochastic bandits," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 24, pp. 2312–2320, 2011.
[26] A. Cohen, T. Koren, and Y. Mansour, "Learning linear-quadratic regulators efficiently with only $\sqrt{T}$ regret," in *Proc. Int. Conf. Mach. Learn.*, 2019, pp. 1300–1309.
[27] L. Furieri, Y. Zheng, and M. Kamgarpour, "Learning the globally optimal distributed LQ regulator," in *Proc. Learn. Dyn. Control Conf.*, 2020, pp. 287–297.
[28] H. Feng and J. Lavaei, "On the exponential number of connected components for the feasible set of optimal decentralized control problems," in *Proc. Amer. Control Conf.*, 2019, pp. 1430–1437.
[29] J. Bu, A. Mesbahi, M. Fazel, and M. Mesbahi, "LQR through the lens of first order methods: Discrete-time case," 2019, *arXiv:1907.08921*.
[30] Y. Li, Y. Tang, R. Zhang, and N. Li, "Distributed reinforcement learning for decentralized linear quadratic control: A derivative-free policy optimization approach," *IEEE Trans. Autom. Control*, to be published, doi: 10.1109/TAC.2021.3128592.
[31] S. Fattahi, N. Matni, and S. Sojoudi, "Efficient learning of distributed linear-quadratic control policies," *SIAM J. Control Optim.*, vol. 58, no. 5, pp. 2927–2951, 2020.
[32] Y. Zheng, L. Furieri, A. Papachristodoulou, N. Li, and M. Kamgarpour, "On the equivalence of Youla, system-level, and input–output parameterizations," *IEEE Trans. Autom. Control*, vol. 66, no. 1, pp. 413–420, Jan. 2021.
[33] S. Dean, H. Mania, N. Matni, B. Recht, and S. Tu, "Regret bounds for robust adaptive control of the linear quadratic regulator," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 31, pp. 4192–4201, 2018.
[34] Y. Abbasi-Yadkori, P. Bartlett, K. Bhatia, N. Lazic, C. Szepesvari, and G. Weisz, "Politex: Regret bounds for policy iteration using expert prediction," in *Proc. Int. Conf. Mach. Learn.*, 2019, pp. 3692–3702.
[35] A. B. Cassel and T. Koren, "Online policy gradient for model free learning of linear quadratic regulators with $\sqrt{T}$ regret," in *Proc. Int. Conf. Mach. Learn.*, 2021, pp. 1304–1313.
[36] A. Cassel, A. Cohen, and T. Koren, "Logarithmic regret for learning linear quadratic regulators efficiently," in *Proc. Int. Conf. Mach. Learn.*, 2020, pp. 1328–1337.
[37] L. Ye, H. Zhu, and V. Gupta, "On the sample complexity of decentralized linear quadratic regulator with partially nested information structure," 2021, *arXiv:2110.07112*.
[38] J. Yu, D. Ho, and A. Wierman, "Online stabilization of unknown networked systems with communication constraints," 2022, *arXiv:2203.02630*.
[39] S. Lale, K. Azizzadenesheli, B. Hassibi, and A. Anandkumar, "Logarithmic regret bound in partially observable linear dynamical systems," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 33, pp. 20876–20888, 2020.
[40] M. Simchowitz, K. Singh, and E. Hazan, "Improper learning for non-stochastic control," in *Proc. Conf. Learn. Theory*, 2020, pp. 3320–3436.
[41] R. A. Horn and C. R. Johnson, *Matrix Analysis*. Cambridge, U.K.: Cambridge Univ. Press, 2012.
[42] S. Fattahi, N. Matni, and S. Sojoudi, "Learning sparse dynamical systems from a single sample trajectory," in *Proc. IEEE Conf. Decis. Control*, 2019, pp. 2682–2689.
[43] K. J. Åström and B. Wittenmark, *Adaptive Control*. Chelmsford, MA, USA: Courier Corporation, 2008.
[44] D. P. Bertsekas, *Dyn. Program. and Optimal Control*: vol. 2, 4th ed. Belmont, MA, USA: Athena Scientific, 2017.



Lintao Ye (Member, IEEE) received the B.E. degree in material science and engineering from the Huazhong University of Science and Technology, Wuhan, China, in 2015, the M.S. degree in mechanical engineering in 2017, and the Ph.D. degree in electrical and computer engineering in 2020, both from Purdue University, West Lafayette, IN, USA.

He is currently a Lecturer with the School of Artificial Intelligence and Automation, Huazhong University of Science and Technology, Wuhan, China. He was a Postdoctoral Researcher with the University of Notre Dame, Notre Dame, IN, USA. His research interests are in the areas of optimization algorithms, control theory, estimation theory, and network science.

Hao Zhu (Senior Member, IEEE) received the B.S. degree in electrical engineering from Tsinghua University, Beijing, China, in 2006, and the M.Sc. and Ph.D. degrees in electrical engineering from the University of Minnesota in 2009 and 2012, respectively.

She is currently an Associate Professor of Electrical and Computer Engineering (ECE) with The University of Texas at Austin, Austin, TX, USA. From 2012 to 2017, she was a Postdoctoral Research Associate and then an Assistant Professor of ECE with the University of Illinois at Urbana-Champaign, Champaign, IL, USA. Her research focus is on developing innovative algorithmic solutions for problems related to learning and optimization for future energy systems. Her current interest includes physics-aware and risk-aware machine learning for power system operations, and energy management system design under the cyber-physical coupling.

Dr. Zhu is a recipient of the NSF CAREER Award and an invited attendee to the US NAE Frontier of Engr. (USFOE) Symposium, and also the faculty advisor for three Best Student Papers awarded at the North American Power Symposium. She is currently an Editor of IEEE TRANSACTIONS ON SMART GRID and IEEE TRANSACTIONS ON SIGNAL PROCESSING.

Vijay Gupta (Fellow, IEEE) received the B.Tech. degree from the Indian Institute of Technology, Delhi, New Delhi, India, and the M.S. and Ph.D. degrees from the California Institute of Technology, Pasadena, CA, USA, all in electrical engineering, in 2001, 2002, and 2007, respectively.

He is currently a Professor with the Elmore Family School of Electrical and Computer Engineering, Purdue University, West Lafayette, IN, USA. His research and teaching interests are broadly in the interface of communication, control, distributed computation, and human decision making.

Dr. Gupta was the recipient of the 2018 Antonio J Rubert Award from the IEEE Control Systems Society, the 2013 Donald P. Eckman Award from the American Automatic Control Council, and a 2009 National Science Foundation (NSF) CAREER Award.