Optimal control of network-coupled subsystems: Spectral decomposition and low-dimensional solutions

Shuang Gao *Member, IEEE,* and Aditya Mahajan *Senior Member, IEEE*

***Abstract*—In this paper we investigate the optimal control of network-coupled subsystems with coupled dynamics and costs. The dynamics coupling may be represented by the adjacency matrix, the Laplacian matrix, or any other symmetric matrix corresponding to an underlying weighted undirected graph. Cost couplings are represented by two coupling matrices which have the same eigenvectors as the coupling matrix in the dynamics. We use the spectral decomposition of these three coupling matrices to decompose the overall system into $(L+1)$ systems with decoupled dynamics and cost, where L is the number of linearly independent eigendirections associated with non-zero eigenvalue triples of the three coupling matrices. Furthermore, the optimal control input at each subsystem can be computed by solving $(L_{\text{dist}}+1)$ decoupled Riccati equations where L_{dist} $(L_{\text{dist}} \leq L)$ is the number of distinct non-zero eigenvalue triples of the three coupling matrices. A salient feature of the result is that, given the spectral decompositions of the couplings, the solution complexity does not directly depend on the number of subsystems. Therefore, the proposed solution framework provides a scalable method for synthesizing and implementing optimal control laws for large-scale network-coupled subsystems.**

***Index Terms*—Optimal control, linear systems, large-scale systems, Riccati equations, spectral decompositions.**

I. Introduction

A. Motivation

The recent proliferation of low cost sensors and actuators has given rise to many networked control systems such as the Internet of Things, smart grids, smart buildings, etc., where multiple subsystems are connected over a network. In such systems, the evolution of the state of a subsystem depends on its local state and local control and is also influenced by the states and controls of its neighbors. Such networks are often referred to as large-scale systems or complex networks, and various aspects of such systems have been investigated since the early 1970s [2], [3], including issues such as controllability [4], [5], observability [5], [6], control energy [7], distributed control [8]–[10] and decentralized control [11]–[14].

A key theme for investigating large-scale networked control systems is to identify conditions under which the optimal control laws may be synthesized and implemented with low-complexity. Such conditions include simplified control objectives (e.g., consensus [8]–[10] and synchronization [15]), simplified control inputs (e.g., pinning control [16]–[18] and ensemble control [19]), simplified coupling between subsystems (e.g., symmetric interconnections [5], [11], [12], [20], [21], exchangeable or anonymous subsystems [22]–[25], sparse connections or structure reduction [26], [27], decoupled dynamics [28], hierarchical decompositions [29] and patterned systems [30]), approximate optimality (e.g., mean-field games [31]–[33], control based on approximate aggregations [34], approximate distributed control [35], [36], and graphon control [37], [38]).

In this paper, we propose a decomposition method for large-scale network-coupled subsystems which relies on the spectral decomposition of the dynamic and cost couplings among the subsystems. Several related approaches have been considered in the literature. An earlier approach similar in spirit to ours is [34], which considered the problem of approximating a high-dimensional system with a low-dimensional system using state aggregation. Both exact and approximate solutions were proposed. Spectral decomposition of large-scale systems with symmetric interconnected subsystems have been considered in [11], [12]. Algebraic decomposition of mean-field coupled subsystems has been considered in [23]–[25]. Algebraic decomposition with cost couplings and no dynamics couplings is considered in [28], [39]. Similar problems under the graphon LQR framework are studied in [38], [40]. A key feature which distinguishes our approach from these works above is that our approach is applicable to models where the couplings in dynamics and costs among subsystems are not homogeneous and such couplings could be dense or spare, and that we establish optimal solutions rather than an approximate solution. Another line of related work is graphical games ([26], [41, Chapter 6]) where the coupling of the utility function depends on an underlying graph. In contrast to these, we consider a control problem and propose a different type of decomposition.

B. Contributions of this paper

In this paper, we investigate a control system with multiple subsystems connected over an undirected graph. Each subsystem has a local state and takes a local control action. The evolution of the state of each subsystem depends on its local state and local control as well as a weighted combination (which we call the network field) of the states and controls of its neighbors. Moreover, the weights in the network field,

S. Gao and A. Mahajan are with the Department of Electrical and Computer Engineering, McGill University, Montreal, QC, Canada. Emails: sgao@cim.mcgill.ca, aditya.mahajan@mcgill.ca.

The work of the second author was supported in part by the Innovation for Defence Excellence and Security (IDEaS) Program of the Canadian Department of National Defence through grant CFPMN2-30

Preliminary version [1] of this work was presented at the 58th IEEE Conference on Decision and Control, Nice, France, December, 2019.

which are represented by a coupling matrix, may correspond to the adjacency matrix, Laplacian matrix or any other symmetric matrix that characterizes the underlying graph. The subsystems can also be coupled in the (quadratic) costs via two coupling matrices that share the same eigenvectors with the coupling matrix in the dynamics. The objective is to choose the control of each subsystem to minimize the total cost over time. The above model is a linear quadratic regulation problem and a centralized solution can be obtained by solving $nd_x \times nd_x$-dimensional Riccati equation, where n is the number of subsystems and d_x is the dimension of the state of each subsystem. In this paper, we propose an alternative solution that has low complexity and may be implemented in a local manner with aggregated (or projected) state information and local state information. For some particular cases, the control can be implemented in a distributed manner that relies on neighbourhood information and local information.

The main contributions of this paper are the following:

- A spectral decomposition technique is devoloped to decompose the linear quadratic control problem for network-coupled dynamical subsystems into $L+1$ decoupled subproblems, where L is the number of linearly independent eigendirections associated with non-zero eigenvalue triples of the three coupling matrices.
- These $L+1$ decoupled subproblems can be solved by solving only $L_{\text{dist}}+1$ decoupled Riccati equations of dimension $d_x \times d_x$, where L_{dist} is the number of distinct non-zero eigenvalue triples of the three coupling matrices and d_x is the state dimension of each subsystem. In contrast, a direct centralized solution requires solving an $nd_x \times nd_x$-dimensional Riccati equation where n is the number of subsystems. We note that the inequalities $L_{\text{dist}} \leq L \leq n$ always hold. Thus the method proposed in this paper leads to considerable simplifications in synthesizing optimal control laws.
- To implement the optimal control input, each subsystem needs to know the $(L+1)d_x$-dimensional vector of local components of eigen and auxiliary states (which are defined later in the paper). In contrast, to implement the centralized solution, each subsystem needs to know the nd_x dimensional global state. Thus, in applications such as [42]–[45] where $L \ll n$, the method proposed in this paper leads to considerable simplification in implementing the optimal control law.
- The solution method is extended to solve stochastic linear quadratic control problems for network-coupled subsystems.
- The solution method is applied to study consensus problems to establish optimal distributed control solutions for some particular cases.

C. Notations and definitions

We use $\mathbb{N}$ and $\mathbb{R}$ to denote respectively the sets of natural and real numbers. The notation $A=[a_{ij}]$ means that a_{ij} is the (i,j)th element of the matrix A. For a vector v, v_i denotes its ith element. For a matrix A, $A^\intercal$ denotes its transpose. Given vectors $v^1, \ldots, v^n$, $\text{cols}(v^1,\ldots,v^n)$ denotes the matrix formed by horizontally stacking the vectors. For any $n \in \mathbb{N}$, $\mathbb{1}_n$ denotes the n-dimensional vector of ones, $\mathbb{1}_{n\times n}$ denotes the $n \times n$-dimensional matrix of ones, and I_n denotes the $n \times n$-dimensional identity matrix.

A pair (A,B) is *stabilizable* if there exists a matrix L such that $A+BL$ is Hurwitz (i.e., all its eigenvalues have negative real parts). A pair (C,A) is *detectable* if there exists a matrix F such that $A^\intercal + FC^\intercal$ is Hurwitz.

II. System model and problem formulation

A. System model

Consider a network consisting of n nodes connected over an undirected weighted graph $\mathcal{G}(\mathcal{N},\mathcal{E},W)$, where $\mathcal{N} = \{1,\ldots,n\}$ is the set of nodes, $\mathcal{E} \subseteq \mathcal{N}\times\mathcal{N}$ is the unordered set of edges, and $W=[w_{ij}] \in \mathbb{R}^{n\times n}$ is the weighted adjacency matrix. Let $M=[m_{ij}] \in \mathbb{R}^{n\times n}$ be a symmetric coupling matrix corresponding to the underlying graph $\mathcal{G}(\mathcal{N},\mathcal{E},W)$. For instance, M may represent the underlying adjacency matrix (i.e., $M=W$) or represent the underlying Laplacian matrix (i.e., $M=\text{diag}(W\mathbb{1}_n)-W$). For any node $i \in \mathcal{N}$, $\mathcal{N}_i \coloneqq \{j \in \mathcal{N} : (i,j) \in \mathcal{E}\}$ denotes the set of neighbors of node i. Note that the edge set $\mathcal{E}$ is allowed to include self-loops. Therefore the set $\mathcal{N}_i$ may contain node i.

The system operates in continuous time for either a finite interval $[0,T]$ or an infinite interval $[0,\infty)$. A state $x_i(t) \in \mathbb{R}^{d_x}$ and a control input $u_i(t) \in \mathbb{R}^{d_u}$ are associated with each node $i \in \mathcal{N}$. At time $t=0$, the system starts from an initial state $(x_i(0))_{i\in\mathcal{N}}$ and for $t>0$, the state of node i evolves according to

$$\dot{x}_i(t) = Ax_i(t) + Bu_i(t) + Dx_i^{\mathcal{G}}(t) + Eu_i^{\mathcal{G}}(t), \qquad (1)$$

where A, B, D and E are matrices of appropriate dimensions and

$$x_i^{\mathcal{G}}(t) = \sum_{j\in\mathcal{N}} m_{ij}x_j(t) \quad \text{and} \quad u_i^{\mathcal{G}}(t) = \sum_{j\in\mathcal{N}} m_{ij}u_j(t) \qquad (2)$$

are the locally perceived *network field* of states and control actions at node i. It is assumed that all the different subsystems have the same parameter matrices A, B, D and E.

We follow an atypical representation of the "vectorized" dynamics. Define

$$x(t) = \text{cols}(x_1(t),\ldots,x_n(t)),$$
$$u(t) = \text{cols}(u_1(t),\ldots,u_n(t)),$$

as the global state and control actions of the system, and

$$x^{\mathcal{G}}(t) = \text{cols}(x_1^{\mathcal{G}}(t),\ldots,x_n^{\mathcal{G}}(t)),$$
$$u^{\mathcal{G}}(t) = \text{cols}(u_1^{\mathcal{G}}(t),\ldots,u_n^{\mathcal{G}}(t)),$$

as the global network field of states and actions. Note that $x(t), x^{\mathcal{G}}(t) \in \mathbb{R}^{d_x\times n}$ and $u(t), u^{\mathcal{G}}(t) \in \mathbb{R}^{d_u\times n}$ are matrices and not vectors. The system dynamics may be written as

$$\dot{x}(t) = Ax(t) + Bu(t) + Dx^{\mathcal{G}}(t) + Eu^{\mathcal{G}}(t). \qquad (3)$$

Furthermore, we may write

$$x^{\mathcal{G}}(t) = x(t)M^\intercal = x(t)M \text{ and } u^{\mathcal{G}}(t) = u(t)M^\intercal = u(t)M.$$

B. System performance and control objective

At time $t \in [0, T)$, the system incurs an instantaneous cost

$$c(x(t), u(t)) = \sum_{i\in\mathcal{N}} \sum_{j\in\mathcal{N}} \big[g_{ij} x_i(t)^\intercal Q x_j(t) + h_{ij} u_i(t)^\intercal R u_j(t)\big], \tag{4}$$

and at the terminal time T, the system incurs a terminal cost

$$c_T(x(T)) = \sum_{i\in\mathcal{N}} \sum_{j\in\mathcal{N}} g_{ij} x_i(T)^\intercal Q_T x_j(T), \tag{5}$$

where Q, Q_T, and R are matrices of appropriate dimensions and g_{ij} and h_{ij} are real-valued weights.

We are interested in the following optimization problems.

Problem 1 *Choose a control trajectory* $u\colon [0,T) \to \mathbb{R}^{d_u \times n}$ *to minimize*

$$J(u) = \int_0^T c(x(t), u(t))dt + c_T(x(T)) \tag{6}$$

subject to the dynamics in (3).

Problem 2 *Choose a control trajectory* $u\colon [0,\infty) \to \mathbb{R}^{d_u \times n}$ *to minimize*

$$J(u) = \int_0^\infty c(x(t), u(t))dt. \tag{7}$$

subject to the dynamics in (3).

C. Assumptions on the cost in the model

(A0) The weight matrices $G = [g_{ij}]$ and $H = [h_{ij}]$ are respectively given by

$$G = q_0 I + q_1 M_q, \quad H = r_0 I + r_1 M_r$$

where M_q and M_r are symmetric matrices that share with M the same set of orthonormal eigenvectors $\{v^1, ..., v^n\}$ associated with all their eigenvalues.

For any two $n\times n$-dimensional symmetric matrices M_1 and M_2, one can verify that the following statements are equivalent: (i) M_1 and M_2 commute (i.e., $M_1M_2 = M_2M_1$); (ii) M_1 and M_2 are simultaneously diagnolizable (i.e., there exists an invertible matrix P such that $P^{-1}M_1P$ and $P^{-1}M_2P$ are both diagonal matrices); (iii) M_1 and M_2 are simultaneously diagnolizable by an orthogonal matrix; (iv) M_1 and M_2 share the same set of n orthonormal eigenvectors.

An important special case of Assumption (A0) is that G and H are both polynomials of M, that is,

$$G = \sum_{k=0}^{K_G} \mathbf{q}_k M^k \quad \text{and} \quad H = \sum_{k=0}^{K_H} \mathbf{r}_k M^k \tag{8}$$

where K_G and K_H denote the degrees of the polynomials and $\{\mathbf{q}_k\}_{k=0}^{K_G}$ and $\{\mathbf{r}_k\}_{k=0}^{K_H}$ are real coefficients (as those cases in [1]). Here we use bold face letters to differentiate them from the coefficients q_0, q_1, r_0 and r_1 in Assumption (A0).

When Assumption (A0) holds, let $(\lambda^\ell, \lambda_r^\ell, \lambda_q^\ell)$ denote the (real) eigenvalue triple of the coupling matrix triple (M, M_r, M_q) associated with the shared eigenvector v^ℓ, $\ell \in \{1, ..., n\}$. We call the eigenvalue triple a *non-zero eigenvalue triple* if at least one eigenvalue in the triple is non-zero. Let L denote the number of linearly independent eigendirections associated with non-zero eigenvalue triples of (M, M_q, M_r). Without loss of generality, let $v^1, ..., v^L$ denote the shared orthonormal eigenvectors associated with all the non-zero eigenvalue triples.

For ease of notation, for $\ell \in \{1, \dots, L\}$, define

$$q^\ell = q_0 + q_1\lambda_q^\ell \quad \text{and} \quad r^\ell = r_0 + r_1\lambda_r^\ell.$$

(A1) The matrices Q and Q_T are symmetric and positive semi-definite and R is symmetric and positive definite.

(A2) For $\ell \in \{1, \dots, L\}$, q^ℓ is non-negative and r^ℓ is strictly positive. Moreover $q_0 \geq 0$ and $r_0 > 0$.

Assumption (A2) ensures that for any $y \in \mathbb{R}^n$, $y^\intercal G y \geq 0$ and $y^\intercal H y > 0$. Assumptions (A1) and (A2) ensure that $G \otimes Q$ and $G \otimes Q_T$ are symmetric positive semi-definite, and $H \otimes R$ is symmetric positive definite, which are standard sufficient conditions for finite-horizon LQR problems to have a unique optimal solution (see for instance [46]).

D. Some remarks on the assumptions on the cost function

Since the subsystems (or agents) are coupled in the dynamics over an underlying graph, it may be reasonable to assume that the cost structure respects the same graph structure. The polynomials in (8) allow us to consider cost coupling structures which may involve not only the immediate neighbourhood but also multiple-hop neighbourhood connections. We present a few examples with different coefficients $\{\mathbf{q}_k\}_{k=0}^{K_G}$ and $\{\mathbf{r}_k\}_{k=0}^{K_H}$ in (8) in the following:

1) If $K_G = K_H = 0$, $\mathbf{q}_0 = 1$, $\mathbf{r}_0 = 1$ and all other coefficients are zero, then $G = H = I$. In this case, the instantaneous cost reduces to

$$c(x(t), u(t)) = \sum_{i=1}^n [x_i(t)^\intercal Q x_i(t) + u_i(t)^\intercal R u_i(t)].$$

Thus, the problem is equivalent to the social optimal control problem where the cost is the summation of the costs of all the subsystems.

2) If $K_G = 2$, $K_H = 1$, $\mathbf{q}_0 = 1, \mathbf{q}_1 = -2, \mathbf{q}_2 = 1$, $\mathbf{r}_0 = 1$ and all other coefficients are zero, then $G = (I - M)^2$ and $H = I$. If, furthermore, the matrix $M = \frac{1}{n}\mathbb{1}_n\mathbb{1}_n^\intercal$, then the instantaneous cost reduces to

$$c(x(t), u(t)) = \sum_{i=1}^n [(x_i(t) - \bar{x}(t))^\intercal Q (x_i(t) - \bar{x}(t)) + u_i(t)^\intercal R u_i(t)],$$

where $\bar{x}(t) := \frac{1}{n}\sum_{i=1}^n x_i(t)$, which is similar to the cost of the social optimal mean field control problem [47].

3) If $K_G = 2$, $K_H = 0$, $\mathbf{q}_2 = 1$, $\mathbf{r}_0 = 1$, all other coefficients are zero, and the coupling matrix is the Laplacian matrix, then $G = M^2, H = I, M = \mathcal{L} := \text{diag}(W\mathbb{1}_n - W)$. The instantaneous cost reduces to

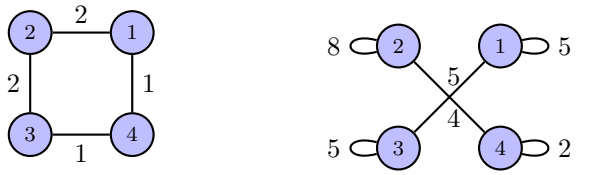
$$(x(t), u(t)) = \sum_{i=1}^n \Big[e_i(t)^\intercal Q e_i(t) + u_i(t)^\intercal R u_i(t)\Big],$$

where the local state error for subsystem i is given by $e_i(t) := \sum_{j\in\mathcal{N}_i} w_{ij}(x_i(t) - x_j(t))$. If, furthermore,

there are no couplings in the dynamics and $A = 0$, this structure then produces the optimal control problem that can be exactly associated with a distributed control problem (see Section V for more details).

4) K_G and K_H can be ∞ as long as the limit of the corresponding polynomial series is well defined. Such examples include the exponential function $G = \exp(M) = \sum_{k=0}^{\infty} \frac{1}{k!} M^k$, and the inverse function $G = (I - \gamma M)^{-1} = \sum_{k=0}^{\infty} \gamma^k M^k$ (when the spectral radius $\rho(M)$ of M satisfies $\rho(M) < \gamma^{-1}$).

E. Salient features of the model



(a) A graph $\mathcal{G}$ (b) 2-hop neighborhood of $\mathcal{G}$

Fig. 1: A graph and its 2-hop neighborhood.

We highlight salient features of the model via an example. Consider a system with 4 nodes connected via a network shown in Fig. 1(a), with

$$G = \mathbf{q}_0 I + \mathbf{q}_1 M + \mathbf{q}_2 M^2 \text{ and } H = \mathbf{r}_0 I + \mathbf{r}_1 M + \mathbf{r}_2 M^2,$$

where M and M^2 are the weighted adjacency matrix of the graph $\mathcal{G}$ and that of the 2-hop neighborhood of $\mathcal{G}$, respectively, given by

$$M = \begin{bmatrix} 0 & 2 & 0 & 1 \\ 2 & 0 & 2 & 0 \\ 0 & 2 & 0 & 1 \\ 1 & 0 & 1 & 0 \end{bmatrix} \quad \text{and} \quad M^2 = \begin{bmatrix} 5 & 0 & 5 & 0 \\ 0 & 8 & 0 & 4 \\ 5 & 0 & 5 & 0 \\ 0 & 4 & 0 & 2 \end{bmatrix}.$$

1) Salient features of the dynamics: For this example,

$$x_1^{\mathcal{G}}(t) = 2x_2(t) + x_4(t), \quad x_2^{\mathcal{G}}(t) = 2x_1(t) + 2x_3(t),$$
$$x_3^{\mathcal{G}}(t) = 2x_2(t) + x_4(t), \quad x_4^{\mathcal{G}}(t) = x_1(t) + x_3(t).$$

Thus, each subsystem is affected by its neighbors. The influence of each neighbor is not homogeneous but depends on the weight associated with the corresponding edge in the graph. Furthermore, the network field $x^{\mathcal{G}}(t)$ is not homogeneous and varies from subsystem to subsystem.

2) Salient features of the cost: If M is the weighted adjacency matrix of the graph $\mathcal{G}$, the matrix M^k, $k \in \mathbb{N}$, represents the weighted adjacency matrix of the k-hop neighborhood of $\mathcal{G}$. Thus, $G = \mathbf{q}_0 I + \mathbf{q}_1 M + \mathbf{q}_2 M^2$ means that each node has a coupling of $\mathbf{q}_0$ with its own state, a coupling of $\mathbf{q}_1$ with its 1-hop neighborhood and a coupling of $\mathbf{q}_2$ with its 2-hop neighborhood. Similar interpretation holds for H. Note that

$$G = \mathbf{q}_0 I + \mathbf{q}_1 M + \mathbf{q}_2 M^2 = \begin{bmatrix} \mathbf{q}_0+5\mathbf{q}_2 & 2\mathbf{q}_1 & 5\mathbf{q}_2 & \mathbf{q}_0+\mathbf{q}_1 \\ 2\mathbf{q}_1 & \mathbf{q}_0+8\mathbf{q}_2 & 2\mathbf{q}_1 & 4\mathbf{q}_2 \\ 5\mathbf{q}_2 & 2\mathbf{q}_1 & \mathbf{q}_0+5\mathbf{q}_2 & \mathbf{q}_1 \\ \mathbf{q}_1 & 4\mathbf{q}_2 & \mathbf{q}_1 & \mathbf{q}_0+2\mathbf{q}_2 \end{bmatrix}.$$

Thus, the agents are not interchangeable, i.e., in general, $G_{ii} \neq G_{jj}$ and $G_{ki} \neq G_{kj}$.

III. Spectral decomposition of the system

Since the coupling matrix M is real and symmetric, it admits spectral factorizations. Under Assumption (A0), the following simultaneous spectral decompositions hold for M, M_q and M_r:

$$M = \sum_{\ell=1}^{L} \lambda^\ell v^\ell v^{\ell\mathsf{T}}, \; M_q = \sum_{\ell=1}^{L} \lambda_q^\ell v^\ell v^{\ell\mathsf{T}}, \; M_r = \sum_{\ell=1}^{L} \lambda_r^\ell v^\ell v^{\ell\mathsf{T}}. \tag{9}$$

In the rest of this section, we decompose the dynamics and the cost based on the above spectral decomposition. Our decompositions may be viewed as generalizations of mean-field decompositions used in [23]–[25] to heterogenous networks.

A. Spectral decomposition of the dynamics

For $\ell \in \{1, \dots, L\}$, define eigenstates and eigencontrol actions as

$$x^\ell(t) = x(t) v^\ell v^{\ell\mathsf{T}}, \tag{10}$$
$$u^\ell(t) = u(t) v^\ell v^{\ell\mathsf{T}}, \tag{11}$$

respectively. Multiplying both sides of (3) by $v^\ell v^{\ell\mathsf{T}}$, we get

$$\dot{x}^\ell(t) = (A + \lambda^\ell D) x^\ell(t) + (B + \lambda^\ell E) u^\ell(t), \tag{12}$$

where we have used the fact that $M v^\ell v^{\ell\mathsf{T}} = \lambda^\ell v^\ell v^{\ell\mathsf{T}}$. Let $x_i^\ell(t)$ and $u_i^\ell(t)$ denote the i-th column of these matrices, i.e.,

$$x^\ell(t) = \operatorname{cols}(x_1^\ell(t), \dots, x_n^\ell(t)),$$
$$u^\ell(t) = \operatorname{cols}(u_1^\ell(t), \dots, u_n^\ell(t)).$$

Therefore, the dynamics (12) can be written as a collection of decoupled "local" dynamics: for $i \in \mathcal{N}$,

$$\dot{x}_i^\ell(t) = (A + \lambda^\ell D) x_i^\ell(t) + (B + \lambda^\ell E) u_i^\ell(t). \tag{13}$$

Using the spectral factorization (9), we may write:

$$x^{\mathcal{G}}(t) = x(t) M = \sum_{\ell=1}^{L} \lambda^\ell x^\ell(t), \tag{14}$$

$$u^{\mathcal{G}}(t) = u(t) M = \sum_{\ell=1}^{L} \lambda^\ell u^\ell(t). \tag{15}$$

Now, define auxiliary state and auxiliary control actions as

$$\breve{x}(t) = x(t) - \sum_{\ell=1}^{L} x^\ell(t) \quad \text{and} \quad \breve{u}(t) = u(t) - \sum_{\ell=1}^{L} u^\ell(t).$$

Then, by subtracting (12) from (3) and substituting (14) and (15), we get

$$\dot{\breve{x}}(t) = A \breve{x}(t) + B \breve{u}(t). \tag{16}$$

Note that $\breve{x}(t) \in \mathbb{R}^{d_x \times n}$ and $\breve{u}(t) \in \mathbb{R}^{d_u \times n}$. Let $\breve{x}_i(t)$ and $\breve{u}_i(t)$ denote the i-th column of these matrices, i.e.,

$$\breve{x}(t) = \operatorname{cols}(\breve{x}_1(t), \dots, \breve{x}_n(t)),$$
$$\breve{u}(t) = \operatorname{cols}(\breve{u}_1(t), \dots, \breve{u}_n(t)).$$

Therefore, the dynamics (16) of the auxiliary state can be written as a collection of decoupled "local" dynamics:

$$\dot{\breve{x}}_i(t) = A\breve{x}_i(t) + B\breve{u}_i(t), \quad i \in \mathcal{N}. \tag{17}$$

The above decomposition may be summarized as follows.

Proposition 1 *The local state and control at each node $i \in \mathcal{N}$ may be decomposed as*

$$x_i(t) = \breve{x}_i(t) + \sum_{\ell=1}^{L} x_i^\ell(t), \tag{18}$$

$$u_i(t) = \breve{u}_i(t) + \sum_{\ell=1}^{L} u_i^\ell(t), \tag{19}$$

where the dynamics of $\breve{x}_i(t)$ depend on only $\breve{u}_i(t)$ and are given by (17) *and the dynamics of $x_i^\ell(t)$ depends on only $u_i^\ell(t)$ and are given by* (13). □

B. Spectral decomposition of the cost

For any $n \times n$ matrix $P = [p_{ij}]$, any $d \times n$ matrices $x = \operatorname{cols}(x_1, \dots, x_n)$, and $y = \operatorname{cols}(y_1, \dots, y_n)$, we use the following short hand notation:

$$\langle x, y \rangle_P = \sum_{i \in \mathcal{N}} \sum_{j \in \mathcal{N}} p_{ij} x_i^\mathsf{T} y_j. \tag{20}$$

Proposition 2 *The instantaneous cost may be written as*

$$c(x(t), u(t)) = \langle x(t), Qx(t) \rangle_G + \langle u(t), Ru(t) \rangle_H,$$

which can be simplified as follows:

$$\begin{aligned}
&\langle x(t), Qx(t) \rangle_G \\
&\qquad = \sum_{i \in \mathcal{N}} \Big[q_0 \breve{x}_i(t)^\mathsf{T} Q \breve{x}_i(t) + \sum_{\ell=1}^{L} q^\ell x_i^\ell(t)^\mathsf{T} Q x_i^\ell(t) \Big], \\
&\langle u(t), Ru(t) \rangle_H \\
&\qquad = \sum_{i \in \mathcal{N}} \Big[r_0 \breve{u}_i(t)^\mathsf{T} R \breve{u}_i(t) + \sum_{\ell=1}^{L} r^\ell u_i^\ell(t)^\mathsf{T} R u_i^\ell(t) \Big]. \quad \square
\end{aligned}$$

See Appendix for the proof.

IV. Main results: Structure and synthesis of optimal control strategies

A. Finite horizon setup

The main result for the finite horizon setup is the following.

Theorem 1 *For $\ell \in \{1, \dots, L\}$, let $P^\ell \colon [0,T] \to \mathbb{R}^{d_x \times d_x}$ be the solution to the backward Riccati differential equation*

$$\begin{aligned}
-\dot{P}^\ell(t) &= (A + \lambda^\ell D)^\mathsf{T} P^\ell(t) + P^\ell(t)(A + \lambda^\ell D) \\
&\quad - P^\ell(t)(B + \lambda^\ell E)(r^\ell R)^{-1}(B + \lambda^\ell E)^\mathsf{T} P^\ell(t) + q^\ell Q
\end{aligned} \tag{21}$$

with the final condition $P^\ell(T) = q^\ell Q_T$. Similarly, let $\breve{P} \colon [0,T] \to \mathbb{R}^{d_x \times d_x}$ be the solution to the backward Riccati differential equation

$$-\dot{\breve{P}}(t) = A^\mathsf{T} \breve{P}(t) + \breve{P}(t) A - \breve{P}(t) B (r_0 R)^{-1} B^\mathsf{T} \breve{P}(t) + q_0 Q \tag{22}$$

with the final condition $\breve{P}(T) = q_0 Q_T$.

Then, under assumptions (A0)*,* (A1) *and* (A2)*, the optimal control strategy for Problem 1 is given by*

$$u_i(t) = -\breve{K}(t)\breve{x}_i(t) - \sum_{\ell=1}^{L} K^\ell(t) x_i^\ell(t), \quad i \in \mathcal{N}, \tag{23}$$

where

$$\begin{aligned}
\breve{K}(t) &= (r_0 R)^{-1} B^\mathsf{T} \breve{P}(t), \\
K^\ell(t) &= (r^\ell R)^{-1} (B + \lambda^\ell E)^\mathsf{T} P^\ell(t).
\end{aligned}$$ □

PROOF Consider the following collections of dynamical systems:

- Eigensystem (ℓ, i), $\ell \in \{1, \dots, L\}$, $i \in \mathcal{N}$, with state $x_i^\ell(t)$, control inputs $u_i^\ell(t)$, dynamics
$$\dot{x}_i^\ell(t) = (A + \lambda^\ell D) x_i^\ell(t) + (B + \lambda^\ell E) u_i^\ell(t),$$
and cost
$$J_i^\ell(u_i^\ell) = \int_0^T \big[q^\ell x_i^\ell(t)^\mathsf{T} Q x_i^\ell(t) + r^\ell u_i^\ell(t)^\mathsf{T} R u_i^\ell(t) \big] dt + q^\ell x_i^\ell(T)^\mathsf{T} Q x_i^\ell(T).$$
- Auxiliary system i, $i \in \mathcal{N}$, with state $\breve{x}_i(t)$, control inputs $\breve{u}_i(t)$, dynamics
$$\dot{\breve{x}}_i(t) = A\breve{x}_i(t) + B\breve{u}_i(t),$$
and cost
$$\breve{J}_i(\breve{u}_i) = \int_0^T \big[q_0 \breve{x}_i(t)^\mathsf{T} Q \breve{x}_i(t) + r_0 \breve{u}_i(t)^\mathsf{T} R \breve{u}_i(t) \big] dt + q_0 \breve{x}_i(T)^\mathsf{T} Q \breve{x}_i(T).$$

Note that all systems have decoupled dynamics and decoupled nonnegative cost. By Proposition 2, we have

$$J(u) = \sum_{i \in \mathcal{N}} \Big[\breve{J}_i(\breve{u}_i) + \sum_{\ell=1}^{L} J_i^\ell(u_i^\ell) \Big].$$

Thus, instead of solving:

(CP1) choose control trajectory $u \colon [0,T) \to \mathbb{R}^{d_u \times n}$ to minimize $J(u)$,

we can equivalently solve the following problems:

(CP2) choose control trajectory $u_i^\ell \colon [0,T) \to \mathbb{R}^{d_u}$ to minimize $J_i^\ell(u_i^\ell)$ for $i \in \mathcal{N}$, $\ell \in \{1, \dots, L\}$,

(CP3) choose control trajectory $\breve{u}_i \colon [0,T) \to \mathbb{R}^{d_u}$ to minimize $\breve{J}_i(\breve{u}_i)$ for $i \in \mathcal{N}$.

Given the solutions of Problems (CP2) and (CP3), we can use Proposition 1 and choose $u_i(t)$ according to (19).

Problems (CP2) and (CP3) are standard optimal control problems and their solution are given as follows. Let $P^\ell \colon [0,T] \to \mathbb{R}^{d_x \times d_x}$ and $\breve{P} \colon [0,T] \to \mathbb{R}^{d_x \times d_x}$ be as given by (21) and (22). Then, for all $i \in \mathcal{N}$, the optimal solution of (CP2) is given by $u_i^\ell(t) = K^\ell(t) x_i^\ell(t)$, $\ell \in \{1, \dots, L\}$, and the solution of (CP3) is given by $\breve{u}_i(t) = \breve{K}(t)\breve{x}_i(t)$. The result follows by combining the above two equations using (19). ■

Remark 1 Based on the definition of $\breve{x}_i(t)$, the control in (23) can be equivalently written as

$$u_i(t) = -\breve{K}(t) x_i(t) - \sum_{\ell=1}^{L} \big(K^\ell(t) - \breve{K}(t) \big) x_i^\ell(t),$$

where the first part represents a local state feedback and the second part represents offset terms proportional to eigen states. □

Remark 2 Although the eigenstates $\{x_i^\ell(t)\}_{\ell=1}^L$ depend on the eigenvectors $(v^1,\dots,v^L)$, the corresponding Riccati equations (21) only depend on the eigenvalue triples $\{(\lambda^\ell,\lambda_r^\ell,\lambda_q^\ell)\}_{\ell=1}^L$. So, if the coupling matrices have repeated eigenvalue triples, as is the case when there are certain symmetries in the underlying graphs, eigendirections with the same eigenvalue triples have the same Riccati equation. Therefore, we only need to solve $L_{\text{dist}}+1$, Riccati equations, where L_{dist} denotes the number of distinct non-zero eigenvalue triples of the coupling matrices (M,M_r,M_q). □

Remark 3 The Riccati equations (21)–(22) are significantly simpler to solve compared to the naive centralized Riccati equation. Each Riccati equation in (21)–(22) is of dimension $d_x\times d_x$, while the centralized Riccati equation is of dimension $nd_x\times nd_x$. So, even if one of the coupling matrices (M, M_q and M_r) is full rank (i.e., $L=n$) and all eigenvalue triples are distinct, solving the n "one-dimensional" Riccati equations (21)–(22) is significantly simpler than solving one centralized "n-dimensional" Riccati equation. For coupling matrices where $L\ll n$, such savings become more drastic. □

B. Infinite horizon setup

Let $Q^{\frac12}$ denote the symmetric positive semi-definite matrix that satisfies ${Q^{\frac12}}^\intercal Q^{\frac12}=Q$. For infinite horizon problems, we further impose the following standard assumptions.

(A3) (A,B) is stabilizable and $(q_0^{\frac12}Q^{\frac12},A)$ is detectable.

(A4) For all $\ell\in\{1,...,L\}$, $(A+\lambda^\ell D,B+\lambda^\ell E)$ is stabilizable and $(q^{\ell\,\frac12}Q^{\frac12},A+\lambda^\ell D)$ is detectable .

Theorem 2 *Suppose assumptions* (A0)–(A4) *hold. For* $\ell\in\{1,\dots,L\}$, *let* $P^\ell\in\mathbb{R}^{d_x\times d_x}$ *be the unique symmetric positive semi-definite solution to the algebraic Riccati equation*

$$\begin{aligned}0=&(A+\lambda^\ell D)^\intercal P^\ell+P^\ell(A+\lambda^\ell D)\\&-P^\ell(B+\lambda^\ell E)(r^\ell R)^{-1}(B+\lambda^\ell E)^\intercal P^\ell+q^\ell Q.\end{aligned}\tag{24}$$

Similarly, let $\breve P\in\mathbb{R}^{d_x\times d_x}$ *be the unique symmetric positive semi-definite solution to the algebraic Riccati equation*

$$0=A^\intercal\breve P+\breve PA-\breve PB(r_0R)^{-1}B^\intercal\breve P+q_0Q.\tag{25}$$

Then the optimal control strategy for Problem 2 is given by

$$u_i(t)=-\breve K\breve x_i(t)-\sum_{\ell=1}^L K^\ell x_i^\ell(t),\tag{26}$$

with $\breve K=(r_0R)^{-1}B^\intercal\breve P$ *and* $K^\ell=(r^\ell R)^{-1}(B+\lambda^\ell E)^\intercal P^\ell$.□

The proof follows along the similar lines as the proof of Theorem 1. Under the extra assumptions (A3) and (A4), one only needs to replace the finite horizon costs with the infinite horizon costs and then solve decoupled LQR problems by solving the corresponding algebraic Riccati equations. (A3) and (A4) ensure the existence of solutions to the algebraic Riccati equations (25) and (24) (see e.g. [48]).

C. Remarks on the information structure and the implementation of the optimal strategy

Since we are interested in regulating a deterministic system, we may implement the optimal control law either using open-loop (i.e. pre-computed) control inputs or using closed-loop (i.e. state feedback) control inputs. For both implementations, the eigenvalue triples $\{(\lambda^\ell,\lambda_q^\ell,\lambda_r^\ell)\}_{\ell=1}^L$ need to be known at all subsystems.

For the open-loop implementation, one can write

$$u_i(t)=-\breve K(t)\breve\Phi(t,0)\breve x_i(0)-\sum_{\ell=1}^L K^\ell(t)\Phi^\ell(t,0)x_i^\ell(0),\tag{27}$$

where the state transition matrices $\breve\Phi(t,0)$ and $\Phi^\ell(t,0)$ are respectively given by $\breve\Phi(t,0)=\exp\Big(\int_0^t\big(A-B\breve K(\tau)\big)d\tau\Big)$ and $\Phi^\ell(t,0)=\exp\Big(\int_0^t\big(A+\lambda^\ell D-(B+\lambda^\ell E)K^\ell(\tau)\big)d\tau\Big)$. Thus, to implement the control action, subsystem i needs to know $\breve x_i(0)$ and $\{x_i^\ell(0)\}_{\ell=1}^L$, which can be obtained using one of the following three information structures:

1) All subsystems know the initial condition $x(0)$ and the eigendirections $\{v^\ell\}_{\ell=1}^L$. Using these, subsystem i can compute $\{x_i^\ell(0)\}_{\ell=1}^L$ and $\breve x_i(0)$, and implement (27).
2) Subsystem $i, i\in\mathcal{N}$, knows its local initial state $x_i(0)$ and its local initial eigensystem states $\{x_i^\ell(0)\}_{\ell=1}^L$. Then subsystem i can compute $\breve x_i(0)$ and implement (27).
3) All subsystems knows the initial state $\{x(0)v^\ell\}_{\ell=1}^L$. In addition, subsystem i knows $v_i:=(v_i^1,\cdots,v_i^L)$ and its local initial state $x_i(0)$. Then subsystem i can compute $\{x_i^\ell(0)\}_{\ell=1}^L$ and $\breve x_i(0)$, and implement (27).

The closed-loop implementation, which is given by (23) or (26), can be obtained by using one of the three information structures described above with $x(0)$, $x_i(0)$ and $x_i^\ell(0)$ replaced by $x(t)$, $x_i(t)$ and $x_i^\ell(t)$, respectively.

Furthermore, for the information structures in 2) and 3), a mixed implementation which combines open-loop and close-loop implementations can also be obtained via only replacing $x_i(0)$ by $x_i(t)$ in 2) and 3). In the mixed implementation, for any subsystem $i\in\mathcal{N}$, the close-loop part corresponds to the individual state $x_i(t)$ and the open-loop part corresponds to the terms $\{x_i^\ell(0)\}_{\ell=1}^L$ or $\{x(0)v^\ell\}_{\ell=1}^L$ which involve the aggregate of initial states of all subsystems.

V. Applications to consensus

Consensus refers to a distributed coordination problem in which nodes connected over a graph update their local states based on the states of their neighbors. The simplest objective is for all nodes to converge to a "consensus" value starting from any initial state $x(0)$, i.e., $\lim_{t\to\infty}\|x_i(t)-x_j(t)\|=0,\ \forall i,j\in\mathcal{N}$. There are various consensus protocols (i.e., rules to update the state at each node as a function of the state of the nearest neighbors and its own state), which have different rates of convergence. We refer readers to [8], [9], [49] for an overview. Often these consensus protocols are hand crafted based on intuitions. In this section, we show that the standard consensus protocol naturally emerges as the optimal solution of an appropriately chosen networked control problem.

In particular, consider a (non-negatively) weighted connected undirected graph $\mathcal{G}(\mathcal{V},\mathcal{E},W)$ where W represents its adjacency matrix. Now consider the system dynamics

$$\dot{x}_i(t) = u_i(t), \quad i \in \mathcal{N} \tag{28}$$

which is a special case of (1) with $A=0, B=I, D=0$, and $E=0$. Furthermore, consider the cost function

$$c(x(t),u(t)) = \langle x(t), Qx(t)\rangle_{M^2} + \langle u(t), Ru(t)\rangle_I \tag{29}$$

where $M = \mathrm{diag}(W\mathbb{1}_n) - W$ is the graph Laplacian matrix, and Q and R are arbitrary symmetric positive definite matrices. The rank of the Laplacian matrix of a (non-negatively weighted) connected graph is $n-1$ and all non-zero eigenvalues are positive. Thus, $L = n-1$ for this setup.

Proposition 3 *The solution to Problem 2 with the dynamics in* (28) *and the cost in* (29) *is given by*

$$u_i(t) = -R^{-1}\Pi \sum_{\ell=1}^{n-1} \lambda^\ell x_i^\ell(t), \quad i \in \mathcal{N}, \tag{30}$$

where Π *denotes the symmetric positive semi-definite solution to* $\Pi R^{-1}\Pi = Q$. □

PROOF Since $B=I$, $Q>0$, $R>0$, $q^\ell = (\lambda^\ell)^2 > 0$, $r^\ell = 1$ $q_0 = 0$, and $r_0 = 1$, (A0)–(A4) are obviously satisfied. An application of Theorem 2 yields the following optimal control

$$u_i(t) = -\sum_{\ell=1}^{n-1} R^{-1}P^\ell x_i^\ell(t), \quad i \in \mathcal{N}, \tag{31}$$

where P^ℓ is the symmetric positive semi-definite solution to the algebraic Riccati equation

$$0 = -P^\ell R^{-1} P^\ell + (\lambda^\ell)^2 Q. \tag{32}$$

Note that $\mathbf{q}_0 = 0$ in this example implies the solution to the auxiliary Riccati equation in (25) is $\breve{P} = 0$. Hence $\breve{K} = 0$ in (26) and the control law (31) does not contain the auxiliary part. Let $\Pi = (\lambda^\ell)^{-1} P^\ell$. Substituting P^ℓ in (32), Π is then given by the symmetric positive semi-definite solution to $\Pi R^{-1}\Pi = Q$. Hence the optimal control is given by (30). ■

Now, recall that

$$\sum_{\ell=1}^{n-1} \lambda^\ell x_i^\ell(t) = x_i^{\mathcal{G}}(t) = \sum_{j\in\mathcal{N}} m_{ji}x_j = \sum_{j\in\mathcal{N}} m_{ij}x_j = \sum_{j\in\mathcal{N}} w_{ij}(x_i - x_j)$$

Therefore, the optimal control may be written as

$$u_i(t) = -R^{-1}\Pi \sum_{j\in\mathcal{N}} w_{ij}(x_i(t) - x_j(t)), \quad i \in \mathcal{N}. \tag{33}$$

Thus the optimal control is the same as the standard consensus protocol in [8], [49]. A similar result was established in [50, Theorem 4.6] using a more sophisticated proof argument.

VI. Generalizations to stochastic systems

A. Stochastic networked control problem

In this section we consider a model similar to Section II-A but with stochastic dynamics. As before, there are n subsystems that are connected over an undirected weighted graph $\mathcal{G}(\mathcal{N},\mathcal{E},W)$, with an associated symmetric coupling matrix M. For any $i \in \mathcal{N}$, the state $x_i(t)$, the control $u_i(t)$, and the network fields $x_i^{\mathcal{G}}(t)$ and $u_i^{\mathcal{G}}(t)$ are defined as before. The difference is that rather than being deterministic, the system dynamics are stochastic and are given by

$$dx_i(t) = \Big[Ax_i(t) + Bu_i(t) + Dx_i^{\mathcal{G}}(t) + Eu_i^{\mathcal{G}}(t)\Big]dt + F dw_i(t), \tag{34}$$

for all $i \in \mathcal{N}$, where the matrices A, B, D, E and F are as before, F is a matrix of an appropriate dimension, the initial states $(x_i(0))_{i\in\mathcal{N}}$ are deterministic, and $\{w_i(t) \in \mathbb{R}^{d_w} : i \in \mathcal{N}, t \geq 0\}$ are standard (d_w-dimensional) Brownian motions that are independent across nodes.

As before, there is an instantaneous cost $c(x(t),u(t))$ for $t \in [0,T)$, and a terminal cost $c(x(T))$, given by (4) and (5).

Let $\mathcal{F}(t)$ denote the σ-algebra generated by $\{w(\tau) : 0 \leq \tau \leq t\}$ where $w(\tau) := \mathrm{cols}\big(w_1(\tau),\ldots,w_n(\tau)\big)$.

We are interested in the following optimization problem.

Problem 3 *Choose an* $\mathcal{F}(t)$*-adapted control* $u\colon [0,T) \to \mathbb{R}^{d_u\times n}$ *to minimize*

$$J(u) = \mathbb{E}\left[\int_0^T c(x(t),u(t))dt + c_T(x(T))\right], \tag{35}$$

subject to the system dynamics in (34) and initial conditions $(x_i(0))_{i\in\mathcal{N}}$.

B. Decompositions

Recall that $w(t) := \mathrm{cols}\big(w_1(t),\ldots,w_n(t)\big) \in \mathbb{R}^{d_w\times n}$. We introduce the following noise processes in eigendirections and the auxiliary direction: for any $i \in \mathcal{N}$ and $\ell \in \{1,...,L\}$,

$$w_i^\ell(t) := w(t)v^\ell v_i^\ell \quad \text{and} \quad \breve{w}_i(t) := w_i(t) - \sum_{\ell=1}^{L} w_i^\ell(t).$$

The corresponding matrix representations are given by

$$\breve{w}(t) := \mathrm{cols}\big(\breve{w}_1(t),\ldots,\breve{w}_n(t)\big),$$
$$w^\ell(t) := \mathrm{cols}\big(w_1^\ell(t),\ldots,w_n^\ell(t)\big).$$

Clearly, $w^\ell(t) = w(t)v^\ell v^{\ell\mathsf{T}}$ and $\mathbb{E}[w_i^\ell] = \mathbb{E}[\breve{w}_i] = 0$.

Lemma 1 *The following statements hold for all* $t \in [0,T]$, $i,j \in \{1,\ldots,n\}$, $\ell,h \in \{1,\ldots,L\}$*:*

1) $\breve{w}_i(t)$ *and* $w_i^\ell(t)$ *are independent.*
2) $\breve{w}_j(t)$ *and* $w_i^\ell(t)$ *are independent if and only if* $v_j^\ell(v_j^\ell - v_i^\ell) = 0$.
3) $\breve{w}_i(t)$ *and* $\breve{w}_j(t)$ *are independent if and only if* $v_i^\ell v_j^\ell = 0$.
4) $w_i^\ell(t)$ *and* $w_j^\ell(t)$ *are independent if and only if* $v_i^\ell v_j^\ell = 0$.
5) *If* $i \neq j$ *and* $\ell \neq h$*, then* $w_i^\ell(t)$ *and* $w_j^h(t)$ *are independent.* □

PROOF Since for any fixed time $t \in [0,T]$, $\breve{w}_i(t)$, $\breve{w}_j(t)$, $w_j^\ell(t)$ and $w_k^h(t)$ are Gaussian random variables with zero mean, they

are independent if and only if the covariance matrix is zero. By explicitly computing the covariance matrices, results in Lemma 1 are verified. ∎

Since w_i^ℓ and $\breve{w}_i$ are linear combinations of independent standard Brownian motions, they themselves are Brownian motions. It is easy to verify that for $s>0,\ t\geq 0$, $\text{var}(w_i^\ell(t+s)-w_i^\ell(t)) = s(v_i^\ell)^2 I_{d_w}$, $\text{var}(\breve{w}_i(t+s)-\breve{w}_i(t)) = s\big(1-\sum_{\ell=1}^L (v_i^\ell)^2\big) I_{d_w}$. Hence the intensities of w_i^ℓ and $\breve{w}_i$ are $|v_i^\ell|$ and $\big(1-\sum_{\ell=1}^L (v_i^\ell)^2\big)^{\frac{1}{2}}$, respectively. Since $\{v^\ell\}_{\ell=1}^L$ forms an orthonormal basis of a subspace in $\mathbb{R}^n$, one can verify that $1-\sum_{\ell=1}^L (v_i^\ell)^2 \geq 0$.

Recall the definition of $\breve{x}_i(t)$, $\breve{u}_i(t)$, $x^\ell(t)$ and $u^\ell(t)$. Following arguments similar to the deterministic case, we obtain the following stochastic differential equations for the decomposed dynamics

$$dx_i^\ell(t) = \Big[(A+\lambda^\ell D)x_i^\ell(t) + (B+\lambda^\ell E)u_i^\ell(t)\Big]dt + F dw_i^\ell(t), \tag{36}$$

$$d\breve{x}_i(t) = \big[A\breve{x}_i(t) + B\breve{u}_i(t)\big]dt + F d\breve{w}_i(t), \tag{37}$$

for all $i\in\mathcal{N}$, $\ell\in\{1,\dots,L\}$. Following the proof argument of Proposition 2, we obtain

$$J(u) = \sum_{i\in\mathcal{N}}\Big[\breve{J}_i(\breve{u}_i) + \sum_{\ell=1}^L J_i^\ell(u_i^\ell)\Big], \tag{38}$$

where for all $i\in\mathcal{N}$ and $\ell\in\{1,\dots,L\}$,

$$J_i^\ell(u_i^\ell) = \mathbb{E}\Big[\int_0^T \big(q^\ell x_i^\ell(t)^\intercal Q x_i^\ell(t) + r^\ell u_i^\ell(t)^\intercal R u_i^\ell(t)\big)dt + q^\ell x_i^\ell(T)^\intercal Q_T x_i^\ell(T)\Big], \tag{39}$$

$$\breve{J}_i(\breve{u}_i) = \mathbb{E}\Big[\int_0^T \big(q_0\breve{x}_i(t)^\intercal Q\breve{x}_i(t) + r_0\breve{u}_i(t)^\intercal R\breve{u}_i(t)\big)dt + q_0\breve{x}_i(T)^\intercal Q_T\breve{x}_i(T)\Big]. \tag{40}$$

C. Optimal control solution

Theorem 3 *Under assumptions* (A0)*,* (A1) *and* (A2)*, the optimal control strategy for Problem 3 is the same as the strategy in Theorem 1 given by* (23)*. Furthermore, the optimal cost is given by*

$$V(x(0)) = \sum_{i\in\mathcal{N}}\Big(\breve{V}_i(\breve{x}_i(0)) + \sum_{\ell=1}^L V_i^\ell(x_i^\ell(0))\Big), \tag{41}$$

where for $i\in\mathcal{N}$ and $\ell\in\{1,...,L\}$,

$$\breve{V}_i(\breve{x}_i(0)) = \breve{x}_i(0)^\intercal\breve{P}(0)\breve{x}_i(0) + \big(1-\sum_{\ell=1}^L (v_i^\ell)^2\big)\int_0^T \text{Tr}\big(\breve{P}(t)FF^\intercal\big)dt, \tag{42}$$

$$V_i^\ell(x_i^\ell(0)) = x_i^\ell(0)^\intercal P^\ell(0)x_i^\ell(0) + (v_i^\ell)^2\int_0^T \text{Tr}\big(P^\ell(t)FF^\intercal\big)dt. \tag{43}$$

□

PROOF The dynamics in (34) can be decomposed into (36) and (37), and the decomposition of the cost in (35) follows (38), (39) and (40). Therefore, Problem 3 can be equivalently decomposed into the linear quadratic control problems defined by (36) and (39), and the linear quadratic control problems given by (37) and (40), where $i\in\mathcal{N}$. Note that the Brownian motions are not necessarily independent across all the decoupled problems as illustrated in Lemma 1. However, following the certainty equivalence principle for linear quadratic Gaussian problems (see e.g., [51]), we obtain the same optimal control feedback gain as the deterministic case, which does not depend on the intensity of the Brownian motion. This, together with the non-negativity of each term in (38) under assumptions (A1) and (A2), implies that solving the decomposed linear quadratic control problems independently yields the optimal feedback gain for and hence optimal solution to Problem 3. Therefore, the optimal feedback gains are the same as those in Theorem 1 for the linear quadratic control problems and the optimal control is given by (23). The optimal costs for the decomposed linear quadratic control problems are given by (42) and (43) (see for instance [52]) and hence the optimal cost for Problem 3 is given by (41). ∎

Note that the intensity of the Brownian motion does not influence the optimal feedback gain but it effects the optimal cost under optimal control.

Remark 4 The result of Theorem 3 generalizes to the infinite horizon long run average cost setup and the infinite horizon discounted cost setup in a natural manner. For each of these setups, the optimal control law will be of the same form as Theorem 3 but the control gains will be time homogeneous and determined by the solution of an algebraic Riccati equation.□

D. A special case: mean-field coupling

Suppose the graph $\mathcal{G}$ is a complete graph (with self-loops) with all edge weights equal to $\frac{1}{n}$. Let M be its adjacency matrix. Then $M=\frac{1}{n}\mathbb{1}_{n\times n}$ has rank 1 and $\lambda_1=1$ is the only non-zero eigenvalue with the normalized eigenvector $v^1=\frac{1}{\sqrt{n}}[1,\dots,1]^\intercal$. Then $x^1(t)=x(t)v^1v^{1\intercal}=x(t)M$. Thus, the eigenstate $x_i^1(t)=\frac{1}{n}\sum_{j=1}^n x_j(t), i\in\mathcal{N}$, is the same for all subsystems and we denote it by $\bar{x}(t)$. Let $M_q=M_r=M$. According to Theorem 1, the Riccati equation of eigensystem is given by

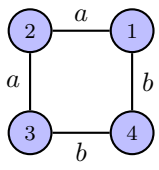
$$-\dot{\bar{P}}(t) = (A+D)^\intercal\bar{P}(t) + \bar{P}(t)(A+D) - \bar{P}(t)(B+E)(r_1R)^{-1}(B+E)^\intercal\bar{P}(t) + q_1Q, \tag{44}$$

where $\bar{P}(t):=P^1(t)$ and the final condition $\bar{P}(T)=q_1Q_T$. The Riccati equation for the auxiliary system is given by

$$-\dot{\breve{P}}(t) = A^\intercal\breve{P}(t) + \breve{P}(t)A - \breve{P}(t)B(r_0R)^{-1}B^\intercal\breve{P}(t) + q_0Q$$

with $\breve{P}(T)=q_0Q_T$. The optimal control strategy is given by $u_i(t) = -\breve{K}(t)(x_i(t)-\bar{x}(t)) - \bar{K}(t)\bar{x}(t)$, where $\breve{K}(t) = (r_0R)^{-1}B^\intercal\breve{P}(t)$ and $\bar{K}(t)=(r_1R)^{-1}(B+E)^\intercal\bar{P}(t)$.

The above result is similar in spirit to [24, Theorem 1 and Theorem 4], which were derived for discrete time systems.



$$\begin{bmatrix} 0 & a & 0 & b \\ a & 0 & a & 0 \\ 0 & a & 0 & b \\ b & 0 & b & 0 \end{bmatrix}$$

Fig. 2: Graph $\mathcal{G}$ with $n = 4$ nodes and its adjacency matrix

VII. Illustrative Examples

A. Adjacency matrix coupling

Consider a network with $n = 4$ subsystems connected over a graph $\mathcal{G}$, as shown in Fig. 2, with its adjacency matrix as the coupling matrix M. Note that $L = \text{rank}(M) = 2$. Consider the following couplings in the cost

$$G = I - 2M + M^2 \quad \text{and} \quad H = I. \tag{45}$$

For the ease of notation define $\rho = \sqrt{2(a^2+b^2)}$ and $\theta = \tan^{-1}(b/a)$. Then it is easy to verify that the non-zero eigenvalues of M are $\lambda^1 = -\rho$ and $\lambda^2 = \rho$. The corresponding eigenvectors are $v^1 = \left[-\frac{1}{2} \ \frac{\sin(\theta)}{\sqrt{2}} \ -\frac{1}{2} \ \frac{\cos(\theta)}{\sqrt{2}}\right]^\mathsf{T}$ and $v^2 = \left[\frac{1}{2} \ \frac{\sin(\theta)}{\sqrt{2}} \ \frac{1}{2} \ \frac{\cos(\theta)}{\sqrt{2}}\right]^\mathsf{T}$. Observe that $q^\ell = (1-\lambda^\ell)^2$ is non-negative and $r^\ell = 1$ is strictly positive, $\ell \in \{1, 2\}$. Thus the model satisfies assumption (A2).

To illustrate how to use the result of Theorem 1, let's pick a subsystem, say subsystem 1, and consider the calculations that need to be carried out at that subsystem. Recall that for all $i \in \mathcal{N}$, $x_i^\ell(0) = x(0)v^\ell v_i^\ell$. Thus

$$x_1^1(0) = \frac{1}{4}x_1(0) - \frac{\sin(\theta)}{2\sqrt{2}}x_2(0) + \frac{1}{4}x_3(0) - \frac{\cos(\theta)}{2\sqrt{2}}x_4(0),$$
$$x_1^2(0) = \frac{1}{4}x_1(0) + \frac{\sin(\theta)}{2\sqrt{2}}x_2(0) + \frac{1}{4}x_3(0) + \frac{\cos(\theta)}{2\sqrt{2}}x_4(0).$$

Following the mixed implementation with information structure 3) described in Section IV-A, subsystem 1 can calculate the trajectory for $x_1^1(t)$, $x_1^2(t)$, $t \in (0, T]$ based on the initial conditions. This together with real time local observation $x_1(t)$ yields $\breve{x}_1(t)$.

Subsystem 1 solves three Riccati equations to compute $P^1(t)$, $P^2(t)$, and $\breve{P}(t)$ for $t \in [0, T]$, and then applies the optimal control action given by

$$u_1(t) = -R^{-1}\Big(B^\mathsf{T}\breve{P}(t)\breve{x}_1(t) + (B-\rho E)^\mathsf{T}P^1(t)x_1^1(t) + (B+\rho E)^\mathsf{T}P^2(t)x_1^2(t)\Big)$$

according to Theorem 1. Similar implementations hold for other subsystems.

Note that if each $x_i(t) \in \mathbb{R}^{d_x}$ then $x(t) \in \mathbb{R}^{4d_x}$. A naive centralized optimal solution of the above system would involve solving a $4d_x \times 4d_x$-dimensional Riccati equation. In contrast, the above solution involves solving three $d_x \times d_x$-dimensional Riccati equations.

Moreover, these computational savings may increase with the size of the networks. For example, consider the graph $\mathcal{G}_{4c} = \mathcal{G} \otimes \mathcal{K}_c$ with $4c$ nodes, where $\mathcal{G}$ is the 4-node graph shown in Fig. 2 and $\mathcal{K}_c$ is the complete graph with c nodes and each edge weight is $\frac{1}{c}$ where c is a positive integer. The adjacency matrix of $\mathcal{G}_{4c}$ is given by $M_{4c} = M \otimes K_c$, where M and $K_c = \frac{1}{c}\mathbb{1}_{c\times c}$ are the adjacency matrices of graph $\mathcal{G}$ and $\mathcal{K}_c$ respectively. The only non-zero eigenvalue of K_c is 1. Thus, the eigenvalues of M_{4c} are the same as eigenvalues of M. Note that the Riccati equations in Theorem 1 only depend on the eigenvalues. So for all different graphs $\mathcal{G}_{4c}$ where c can be any positive integer, the Riccati equations are the same. The method proposed in Theorem 1 would require solving the same three $d_x \times d_x$-dimensional Riccati equations while a naive direct solution would require solving a $4cd_x \times 4cd_x$-dimensional Riccati equation.

As an illustration, we consider the graph $\mathcal{G}_{4c} = \mathcal{G} \otimes \mathcal{K}_c$ where $\mathcal{G}$ is given in Fig. 2 with weights $a = 2$ and $b = 1$. Recall that G and H are given by (45). As argued above, the matrix M_{4c} has two non-zero eigenvalues and the optimal control at each subsystem can be obtained by solving only 3 Riccati equations. Let us set $c = 5$. Then $M_{20} = M \otimes \frac{1}{5}\mathbb{1}_{5\times 5}$.

Example 1: We consider a network of coupled harmonic oscillators where for subsystem $i \in \mathcal{N}$, the state is given by $x_i = [\theta_i, \omega_i]^\mathsf{T}$ representing the angle and angular velocity, and the control u_i represents the input force. Thus $d_x = 2$ and $d_u = 1$. Consider Problem 1 with 20 coupled harmonic oscillators on G_{20} where the coupling matrix is the adjacency matrix M_{20} of the graph G_{20} and the parameters are

$$A = \begin{bmatrix} 0 & 10 \\ -20 & 0 \end{bmatrix}, \ B = \begin{bmatrix} 0 \\ 1.5 \end{bmatrix}, \ D = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \ R = 1,$$
$$E = \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \ Q = \begin{bmatrix} 6 & 0 \\ 0 & 6 \end{bmatrix}, \ Q_T = \begin{bmatrix} 5 & 0 \\ 0 & 5 \end{bmatrix}, \ T = 2. \tag{46}$$

The result is illustrated in Fig. 3.

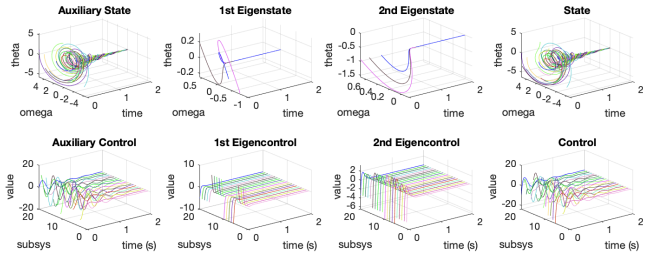


Fig. 3: Numerical example under the proposed optimal control on a network of 20 coupled harmonic oscillators with $T = 2$.

B. Adjacency matrix coupling for stochastic systems

We consider the same model as in the previous section, but assume that the system dynamics are stochastic. In particular, we consider the graph $\mathcal{G}_{20}$ in Section VII-A.

Example 2: Consider the stochastic generalization of Example 1 with $F = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$. All other parameters are the same as those in Example 1. A simulation result is given in Fig. 4.

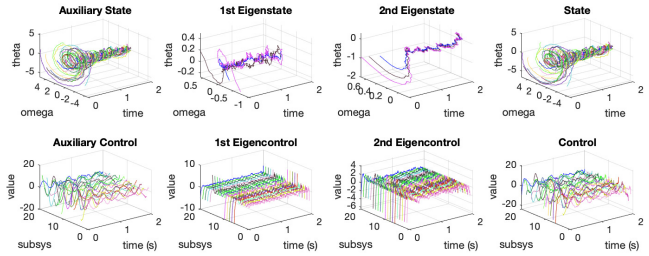


Fig. 4: Numerical example with additive noise under the proposed optimal control on a network of 20 coupled harmonics oscillators with time horizon $T = 2$.

C. Laplacian matrix coupling

We now consider examples where the coupling matrix M is the Laplacian matrix of the underlying graph $\mathcal{G}_{20}$ in Section VII-A.

Example 3: Consider Problem 1 with 20 coupled harmonic oscillators on $\mathcal{G}_{20}$ where $d_x = 2$, $d_u = 1$, the coupling matrix is the Laplacian matrix $\mathcal{L} := \operatorname{diag}(M_{20}\mathbb{1}_{20}) - M_{20}$ of the graph $\mathcal{G}_{20}$, and the parameters are given by $G = \mathcal{L}^2$, $H = I$ and (46). The graph $\mathcal{G}_{20}$ is connected and hence the rank of $\mathcal{L}$ is 19. However, there are only 5 distinct non-zero eigenvalues. Therefore, the solution following Theorem 1 requires solving $5 + 1$ decoupled scalar Riccati equations (see Remark 2). In contrast, a direct centralized solution requires solving a 20×20 dimensional matrix Riccati equation. Note that, the solution to the auxiliary Riccati equation is $\breve{P}(t) = 0$ for all $t \in [0, T]$, which implies the control signal in the auxiliary direction should alway be zero (see the auxiliary control in Fig. 5). The simulation result is illustrated in Fig. 5.

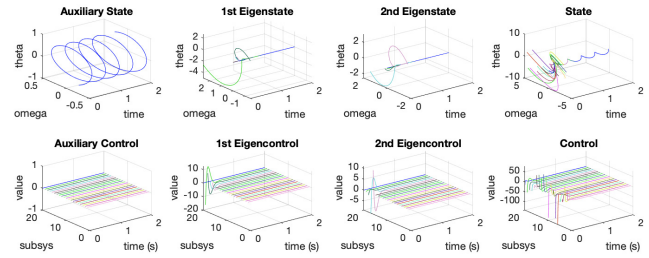


Fig. 5: Numerical example with Laplacian matrix coupling under the proposed optimal control on a network of 20 coupled harmonics oscillators over the time horizon $T = 2$.

VIII. Conclusion

We consider the optimal control of network-coupled subsystems where the dynamics and cost couplings depend on three symmetric coupling matrices that share the same set of eigenvectors. The main idea of a low-dimensional decomposition is to project the state $x(t)$ into L orthogonal eigendirections where L denotes the number of linearly independent eigendirections associated with non-zero eigenvalue triples of the coupling matrices. This projection generates L eigenstates $\{x^\ell(t)\}_{\ell=1}^L$ and an auxiliary state $\breve{x}(t) = x(t) - \sum_{\ell=1}^L x^\ell(t)$. A similar decomposition is obtained for the control inputs. These $L + 1$ components are decoupled both in dynamics and cost. Therefore, the optimal control input for each component can be obtained by solving decoupled Riccati equations.

The proposed approach requires solving at most $L + 1$ Riccati equations, each of dimension $d_x \times d_x$. If, furthermore, some of the non-zero eigenvalue triples are repeated and the coupling matrix has only L_{dist} (with $L_{\text{dist}} \leq L \leq n$) distinct non-zero eigenvalue triples, then the proposed approach only requires solving $L_{\text{dist}} + 1$ decoupled Riccati equations. In contrast, a naive centralized solution requires solving an $nd_x \times nd_x$-dimensional Riccati equation. Thus, even when $L_{\text{dist}} = n$, the proposed approach leads to considerable computational savings. These savings improve significantly when $L_{\text{dist}} \ll n$, as is the case for adjacency matrices for many real-world networks.

The proposed approach requires spectral decompositions of the coupling matrices. For some matrices, spectral decompositions can be obtained analytically [53], e.g., when M is the adjacency or Laplacian matrix of a complete graph or those of an undirected circulant graph. For others, the spectral decompositions can be approximated via graphons ([38], [40]). However, in general, the spectral decomposition will need to be computed numerically, which typically has a complexity of $\mathcal{O}(n^3)$ (e.g., using QR iterations with Householder transformation [54, p. 213]). Even when the spectral decomposition needs to be computed numerically, the approach proposed in this paper leads to computational savings. Solving an algebraic Riccati equation (in the infinite horizon setting) with state dimension d_x has a complexity[1] of $\mathcal{O}(d_x^3)$. Hence the complexity of our solution, including the spectral decomposition, is $\mathcal{O}(n^3) + \mathcal{O}(L_{\text{dist}} d_x^3)$. In contrast, the complexity of a naive centralized solution is $\mathcal{O}(n^3 d_x^3)$.

Future directions of this work include: 1) models where the subsystems have different local dynamics, 2) models where the coupling matrices only admit an approximate low-rank representations, and 3) models with non-linear local dynamics. Some generalizations of the coupling structure using the common invariant subspace property appear in [40]. A learning algorithm which exploits the spectral decomposition proposed in this paper appears in [56].

References

[1] S. Gao and A. Mahajan, "Networked control of coupled subsystems: Spectral decomposition and low-dimensional solutions," in *Proc. Conf. Decision and Control*, Nice, France, December 2019, pp. 4514–4520.

[2] Y. Ho and S. Mitter, *Directions in Decentralized Control, Many-Person Optimization, and Large Scale Systems*. Plenum, New York, 1976.

[3] N. Sandell, P. Varaiya, M. Athans, and M. Safonov, "Survey of decentralized control methods for large scale systems," *IEEE Trans. Autom. Control*, vol. 23, no. 2, pp. 108–128, 1978.

[4] Y.-Y. Liu, J.-J. Slotine, and A.-L. Barabási, "Controllability of complex networks," *Nature*, vol. 473, no. 7346, pp. 167–173, 2011.

[5] G.-H. Yang and S.-Y. Zhang, "Structural properties of large-scale systems possessing similar structures," *Automatica*, vol. 31, no. 7, pp. 1011–1017, 1995.

[6] G. Yan, G. Tsekenis, B. Barzel, J.-J. Slotine, Y.-Y. Liu, and A.-L. Barabási, "Spectrum of controlling and observing complex networks," *Nature Physics*, vol. 11, no. 9, pp. 779–786, 2015.

[1]There are faster algorithms without explicit complexity which rely on structures of parameter matrices of algebraic Riccati equations ([55]).

[7] F. Pasqualetti, S. Zampieri, and F. Bullo, "Controllability metrics, limitations and algorithms for complex networks," *IEEE Trans. Control Netw. Syst.*, vol. 1, no. 1, pp. 40–52, 2014.

[8] R. Olfati-Saber and R. M. Murray, "Consensus protocols for networks of dynamic agents," in *Proc. American Control Conf.*, vol. 2. IEEE, 2003, pp. 951–956.

[9] A. Jadbabaie, J. Lin, and A. S. Morse, "Coordination of groups of mobile autonomous agents using nearest neighbor rules," *IEEE Trans. Autom. Control*, vol. 48, no. 6, pp. 988–1001, 2003.

[10] K. H. Movric and F. L. Lewis, "Cooperative optimal control for multi-agent systems on directed graph topologies," *IEEE Trans. Autom. Control*, vol. 59, no. 3, pp. 769–774, 2013.

[11] M. K. Sundareshan and R. M. Elbanna, "Qualitative analysis and decentralized controller synthesis for a class of large-scale systems with symmetrically interconnected subsystems," *Automatica*, vol. 27, no. 2, pp. 383–388, 1991.

[12] G.-H. Yang and S.-Y. Zhang, "Decentralized control of a class of large-scale systems with symmetrically interconnected subsystems," *IEEE Trans. Autom. Control*, vol. 41, no. 5, pp. 710–713, 1996.

[13] D. D. Šiljak and A. Zečević, "Control of large-scale systems: Beyond decentralized feedback," *Annual Reviews in Control*, vol. 29, no. 2, pp. 169–179, 2005.

[14] T. Bian, Y. Jiang, and Z.-P. Jiang, "Decentralized adaptive optimal control of large-scale systems with application to power systems," *IEEE Trans. Ind. Electron.*, vol. 62, no. 4, pp. 2439–2447, 2015.

[15] A. Arenas, A. Díaz-Guilera, J. Kurths, Y. Moreno, and C. Zhou, "Synchronization in complex networks," *Physics Reports*, vol. 469, no. 3, pp. 93–153, 2008.

[16] H. Gang and Q. Zhilin, "Controlling spatiotemporal chaos in coupled map lattice systems," *Physical Review Letters*, vol. 72, no. 1, p. 68, 1994.

[17] R. Grigoriev, M. Cross, and H. Schuster, "Pinning control of spatiotemporal chaos," *Physical Review Letters*, vol. 79, no. 15, p. 2795, 1997.

[18] X. F. Wang and G. Chen, "Pinning control of scale-free dynamical networks," *Physica A: Statistical Mechanics and Its Applications*, vol. 310, no. 3-4, pp. 521–531, 2002.

[19] J.-S. Li, "Ensemble control of finite-dimensional time-varying linear systems," *IEEE Trans. Autom. Control*, vol. 56, no. 2, pp. 345–357, 2011.

[20] J. Lunze, "Dynamics of strongly coupled symmetric composite systems," *International Journal of Control*, vol. 44, no. 6, pp. 1617–1640, 1986.

[21] J. Grizzle and S. Marcus, "The structure of nonlinear control systems possessing symmetries," *IEEE Trans. Autom. Control*, vol. 30, no. 3, pp. 248–258, 1985.

[22] D. Madjidian and L. Mirkin, "Distributed control with low-rank coordination," *IEEE Trans. Control Netw. Syst.*, vol. 1, no. 1, pp. 53–63, 2014.

[23] J. Arabneydi and A. Mahajan, "Team-optimal solution of finite number of mean-field coupled LQG subsystems," in *Proc. Conf. Decision and Control*, Dec 2015, pp. 5308–5313.

[24] ——, "Linear quadratic mean field teams: Optimal and approximately optimal decentralized solutions," *arXiv preprint arXiv:1609.00056v2*, 2017.

[25] R. Elliott, X. Li, and Y.-H. Ni, "Discrete time mean-field stochastic linear-quadratic optimal control problems," *Automatica*, vol. 49, no. 11, pp. 3222–3233, 2013.

[26] L. Arditti, G. Como, and F. Fagnani, "Graphical games and decomposition," *arXiv preprint arXiv:2003.13123*, 2020.

[27] P. Benner, "Solving large-scale control problems," *IEEE Control Systems Magazine*, vol. 24, no. 1, pp. 44–59, 2004.

[28] G. Jing, H. Bai, J. George, and A. Chakrabortty, "Model-free optimal control of linear multi-agent systems via decomposition and hierarchical approximation," *arXiv preprint arXiv: 2008.06604*, 2020.

[29] S.-C. Chang, T.-S. Chang, and P. B. Luh, "A hierarchical decomposition for large-scale optimal control problems with parallel processing structure," *Automatica*, vol. 25, no. 1, pp. 77–86, 1989.

[30] S. C. Hamilton and M. E. Broucke, "Patterned linear systems," *Automatica*, vol. 48, no. 2, pp. 263–272, 2012.

[31] M. Huang, R. P. Malhamé, and P. E. Caines, "Large population stochastic dynamic games: closed-loop McKean-Vlasov systems and the nash certainty equivalence principle," *Communications in Information & Systems*, vol. 6, no. 3, pp. 221–252, 2006.

[32] J.-M. Lasry and P.-L. Lions, "Jeux à champ moyen. I–le cas stationnaire," *Comptes Rendus Mathématique*, vol. 343, no. 9, pp. 619–625, 2006.

[33] T. Li and J.-F. Zhang, "Asymptotically optimal decentralized control for large population stochastic multiagent systems," *IEEE Trans. Autom. Control*, vol. 53, no. 7, pp. 1643–1660, 2008.

[34] M. Aoki, "Control of large-scale dynamic systems by aggregation," *IEEE Trans. Autom. Control*, vol. 13, no. 3, pp. 246–253, 1968.

[35] J. Jiao, H. L. Trentelman, and M. K. Camlibel, "A suboptimality approach to distributed linear quadratic optimal control," *IEEE Trans. Autom. Control*, vol. 65, no. 3, pp. 1218–1225, 2020.

[36] F. Borrelli and T. Keviczky, "Distributed LQR design for identical dynamically decoupled systems," *IEEE Trans. Autom. Control*, vol. 53, no. 8, pp. 1901–1912, 2008.

[37] S. Gao and P. E. Caines, "Graphon control of large-scale networks of linear systems," *IEEE Trans. Autom. Control*, vol. 65, no. 10, pp. 4090–4105, 2020.

[38] ——, "Optimal and approximate solutions to linear quadratic regulation of a class of graphon dynamical systems," in *Proc. Conf. Decision and Control*, Nice, France, December 2019, pp. 8359–8365.

[39] G. Jing, H. Bai, J. George, and A. Chakrabortty, "Decomposability and parallel computation of multi-agent LQR," *arXiv preprint arXiv:2010.08615*, 2020.

[40] S. Gao and P. E. Caines, "Subspace decomposition for graphon LQR: Applications to VLSNs of harmonic oscillators," *IEEE Trans. Control Netw. Syst.*, 2020, doi: 10.1109/TCNS.2021.3058923.

[41] F. L. Lewis, H. Zhang, K. Hengster-Movric, and A. Das, *Cooperative control of multi-agent systems: optimal and adaptive design approaches*. Springer Science & Business Media, 2013.

[42] M. Udell and A. Townsend, "Why are big data matrices approximately low rank?" *SIAM Journal on Mathematics of Data Science*, vol. 1, no. 1, pp. 144–160, 2019.

[43] K. Zhou, H. Zha, and L. Song, "Learning social infectivity in sparse low-rank networks using multi-dimensional hawkes processes," in *Artificial Intelligence and Statistics*, 2013, pp. 641–649.

[44] S. Gao and P. E. Caines, "Spectral representations of graphons in very large network systems control," in *Proc. Conf. Decision and Control*, Nice, France, December 2019, pp. 5068–5075.

[45] B. D. MacArthur, R. J. Sánchez-García, and J. W. Anderson, "Symmetry in complex networks," *Discrete Applied Mathematics*, vol. 156, no. 18, pp. 3525–3531, 2008.

[46] D. Liberzon, *Calculus of variations and optimal control theory: a concise introduction*. Princeton University Press, 2011.

[47] M. Huang, P. E. Caines, and R. P. Malhamé, "Social optima in mean field LQG control: centralized and decentralized strategies," *IEEE Trans. Autom. Control*, vol. 57, no. 7, pp. 1736–1751, 2012.

[48] V. Kucera, "A contribution to matrix quadratic equations," *IEEE Trans. Autom. Control*, vol. 17, no. 3, pp. 344–347, 1972.

[49] M. Mesbahi and M. Egerstedt, *Graph theoretic methods in multiagent networks*. Princeton University Press, 2010.

[50] Y. Cao and W. Ren, "Optimal linear-consensus algorithms: An LQR perspective," *IEEE Trans. Syst., Man, Cybern. B*, vol. 40, no. 3, pp. 819–830, 2009.

[51] H. Van de Water and J. Willems, "The certainty equivalence property in stochastic control theory," *IEEE Trans. Autom. Control*, vol. 26, no. 5, pp. 1080–1087, 1981.

[52] J. Yong and X. Y. Zhou, *Stochastic controls: Hamiltonian systems and HJB equations*. Springer Science & Business Media, 1999, vol. 43.

[53] A. E. Brouwer and W. H. Haemers, *Spectra of graphs*. Springer Science & Business Media, 2011.

[54] J. W. Demmel, *Applied numerical linear algebra*. Society for Industrial and Applied Mathematics (SIAM), 1997.

[55] P. Benner, Z. Bujanovic, P. Kurschner, and J. Saak, "A numerical comparison of different solvers for large-scale, continuous-time algebraic riccati equations and LQR problems," *SIAM Journal on Scientific Computing*, vol. 42, no. 2, pp. A957–A996, 2020.

[56] S. Sudhakar, A. Mahajan, A. Nayyar, and Y. Ouyang, "Scalable regret for learning to control network-coupled subsysems with unknown dynamics," *submitted to IEEE Trans. Autom. Control*, 2021.

APPENDIX

A. Preliminary properties of the state decomposition

Lemma 2 *Let k be a positive integer k and $\ell, \ell' \in \{1, \dots, L\}$. Let the pair $(\bar{M}, \bar{\lambda}^\ell)$ represent (M, λ^ℓ), (M_r, λ_r^ℓ) or (M_q, λ_q^ℓ). Then, under Assumption (A0), we have the following:*

(P1) $x^\ell(t)\bar{M} = \bar{\lambda}^\ell x^\ell(t)$ *and* $u^\ell(t)\bar{M} = \bar{\lambda}^\ell u^\ell(t)$.

(P2) $x^\ell(t)\bar{M}^k = (\bar{\lambda}^\ell)^k x^\ell(t)$ *and* $u^\ell(t)\bar{M}^k = (\bar{\lambda}^\ell)^k u^\ell(t)$.

(P3) $x^\ell(t)G = q^\ell x^\ell(t)$ *and* $u^\ell(t)H = r^\ell u^\ell(t)$.

(P4) $\breve{x}(t)\bar{M} = 0$ *and* $\breve{u}(t)\bar{M} = 0$.

(P5) $\breve{x}(t)\bar{M}^k = 0$ *and* $\breve{u}(t)\bar{M}^k = 0$.

(P6) $\breve{x}(t)G = q_0\breve{x}(t)$ *and* $\breve{u}(t)H = r_0\breve{u}(t)$.

(P7) $x(t)G = q_0\breve{x}(t) + \sum_{\ell=1}^{L} q^\ell x^\ell(t)$ *and* $u(t)G = r_0\breve{u}(t) + \sum_{\ell=1}^{L} r^\ell u^\ell(t)$.

(P8) $\sum_{i\in\mathcal{N}} x_i^\ell(t)^\intercal Q x_i^{\ell'}(t) = \delta_{\ell\ell'} \sum_{i\in\mathcal{N}} x_i^\ell(t)^\intercal Q x_i^{\ell'}(t)$, *where* $\delta_{\ell\ell'}$ *is the Kronecker delta function.*

(P9) $\sum_{i\in\mathcal{N}} x_i(t)^\intercal Q x_i^\ell(t) = \sum_{i\in\mathcal{N}} x_i^\ell(t)^\intercal Q x_i^\ell(t)$ *and* $\sum_{i\in\mathcal{N}} u_i(t)^\intercal R u_i^\ell(t) = \sum_{i\in\mathcal{N}} u_i^\ell(t)^\intercal R u_i^\ell(t)$. □

PROOF We show the result for $\breve{x}(t)$. The result for $\breve{u}(t)$ follows from a similar argument.

Since $v^1, \dots, v^L$ are orthonormal, from (9) we have $v^\ell v^{\ell\intercal}\bar{M} = \bar{\lambda}^\ell v^\ell v^{\ell\intercal}$, which implies (P1). (P2) follows immediately from (P1) and (P3) follows from (P2).

(P4) follows immediately from the definition of $\breve{x}(t)$, (14) and (P1). (P5) follows immediately from (P4) and (P6) follows from (P5). (P7) follows from (18), (P3) and (P6). To prove (P8), we observe that (10) implies that

$$\sum_{i\in\mathcal{N}} x_i^\ell(t)^\intercal Q x_i^{\ell'}(t) = \sum_{i\in\mathcal{N}} v_i^\ell v^{\ell\intercal} x(t)^\intercal Q x(t) v^{\ell'} v_i^{\ell'\intercal} = \Big(\sum_{i\in\mathcal{N}} v_i^\ell v_i^{\ell'}\Big) v^{\ell\intercal} x(t)^\intercal Q x(t) v^{\ell'}. \tag{47}$$

Since $v^1, \dots, v^L$ is orthonormal, we get $\sum_{i\in\mathcal{N}} v_i^\ell v_i^{\ell'} = v^{\ell\intercal} v^{\ell'} = \delta_{\ell\ell'}$. Substituting this in (47) completes the proof of (P8). To prove (P9) observe that

$$\sum_{i\in\mathcal{N}} x_i(t)^\intercal Q x_i^\ell(t) = \sum_{i\in\mathcal{N}} x_i(t)^\intercal Q x(t) v^\ell v_i^\ell = \sum_{i\in\mathcal{N}} v_i^\ell x_i(t)^\intercal Q x(t) v^\ell = v^{\ell\intercal} x(t)^\intercal Q x(t) v^\ell. \tag{48}$$

From (47), we get that the expression in (48) is equal to $\sum_{i\in\mathcal{N}} x_i^\ell(t)^\intercal Q x_i^\ell(t)$. ■

Lemma 3 *Let P, x, and y be defined in* (20)*. Let P_i denote the i-th column of P. Then, we can write*

$$\langle x, y\rangle_P = \sum_{i\in\mathcal{N}} x_i^\intercal y P_i \quad \text{or} \quad \langle x, y\rangle_P = \sum_{j\in\mathcal{N}} P_j^\intercal x^\intercal y_j. \qquad \square$$

PROOF The result follows from the definition of $\langle x, y\rangle_P$. ■

B. Proof for Proposition 2

We consider the terms depending on $x(t)$. The term depending on $u(t)$ may be simplified in a similar manner.

From (18) and linearity of $\langle\cdot,\cdot\rangle_G$ in both arguments, we get

$$\begin{aligned}\langle x(t), Qx(t)\rangle_G &= \Big\langle \breve{x}(t) + \sum_{\ell=1}^{L} x^\ell(t), Q\Big(\breve{x}(t) + \sum_{\ell=1}^{L} x^\ell(t)\Big)\Big\rangle_G \\ &= \langle \breve{x}(t), Q\breve{x}(t)\rangle_G + 2\Big\langle \sum_{\ell=1}^{L} x^\ell(t), Q\breve{x}(t)\Big\rangle_G \\ &\quad + \Big\langle \sum_{\ell=1}^{L} x^\ell(t), Q\Big(\sum_{\ell=1}^{L} x^\ell(t)\Big)\Big\rangle_G. \end{aligned} \tag{49}$$

From Lemma 3 and (P6), the first term of (49) simplifies to

$$\langle \breve{x}(t), Q\breve{x}(t)\rangle_G = q_0 \sum_{i\in\mathcal{N}} \breve{x}_i(t)^\intercal Q \breve{x}_i(t), \tag{50}$$

and the second term simplifies to

$$\begin{aligned}\Big\langle \sum_{\ell=1}^{L} x^\ell(t), Q\breve{x}(t)\Big\rangle_G &= q_0 \sum_{i\in\mathcal{N}} \sum_{\ell=1}^{L} x_i^\ell(t)^\intercal Q \breve{x}_i(t) \\ &= q_0 \sum_{\ell=1}^{L} \sum_{i\in\mathcal{N}} x_i^\ell(t)^\intercal Q\Big(x_i(t) - \sum_{\ell'=1}^{L} x_i^{\ell'}(t)\Big) \\ &\overset{(a)}{=} q_0 \sum_{\ell=1}^{L} \sum_{i\in\mathcal{N}} \Big(x_i^\ell(t)^\intercal Q x_i^\ell(t) - x_i^\ell(t)^\intercal Q x_i^\ell(t)\Big) = 0,\end{aligned}$$

where (a) follows from (P8) and (P9). From Lemma 3 and (P3), the third term of (49) simplifies to

$$\begin{aligned}\Big\langle \sum_{\ell=1}^{L} x^\ell(t), Q\Big(\sum_{\ell=1}^{L} x^\ell(t)\Big)\Big\rangle_G &= \sum_{i\in\mathcal{N}} \sum_{\ell=1}^{L} x_i^\ell(t)^\intercal Q\Big(\sum_{\ell'=1}^{L} q^{\ell'} x_i^{\ell'}(t)\Big) \\ &= \sum_{\ell=1}^{L} \sum_{i\in\mathcal{N}} x_i^\ell(t)^\intercal Q\Big(\sum_{\ell'=1}^{L} q^{\ell'} x_i^{\ell'}(t)\Big) \\ &\overset{(b)}{=} \sum_{\ell=1}^{L} \sum_{i\in\mathcal{N}} q^\ell x_i^\ell(t)^\intercal Q x_i^\ell(t),\end{aligned} \tag{51}$$

where (b) follows from (P8). We get the result by substituting (50)–(51) in (49).

PLACE PHOTO HERE

Shuang Gao (S'14-M'19) received the B.E. degree in automation and M.S. in control science and engineering, from Harbin Institute of Technology, Harbin, China, in 2011 and 2013. He received the Ph.D. degree in electrical engineering from McGill University, Montreal, QC, Canada, in Feb. 2019. He is currently a Postdoctoral Researcher at the Department of Electrical and Computer Engineering at McGill University. His main research interests include control of network systems, optimization on networks, network modeling, and mean field games.

PLACE PHOTO HERE

Aditya Mahajan (S'06-M'09-SM'14) received B.Tech degree from the Indian Institute of Technology, Kanpur, India, in 2003, and M.S. and Ph.D. degrees from the University of Michigan, Ann Arbor, USA, in 2006 and 2008. From 2008 to 2010, he was a Postdoctoral Researcher at Yale University, New Haven, CT, USA. He has been with the department of Electrical and Computer Engineering, McGill University, Montreal, Canada, since 2010 where he is currently Associate Professor. He currently serves as Associate Editor of Springer Mathematics of Control, Signal, and Systems. He was an Associate Editor of the IEEE Control Systems Society Conference Editorial Board from 2014 to 2017. He is the recipient of the 2015 George Axelby Outstanding Paper Award, 2014 CDC Best Student Paper Award (as supervisor), and the 2016 NecSys Best Student Paper Award (as supervisor). His principal research interests include learning and control of decentralized stochastic systems.