Performance Bounds for Robust Decentralized Control

Weixuan Lin Eilyan Bitar

***Abstract*—We consider the decentralized output feedback control of stochastic linear systems, subject to robust linear constraints on both the state and input trajectories. For problems with partially nested information structures, we establish an upper bound on the minimum achievable cost by computing the optimal affine decentralized control policy as a solution to a finite-dimensional conic program. For problems with general (possibly *nonclassical*) information structures, we construct another finite-dimensional conic program whose optimal value stands as a lower bound on the minimum achievable cost. With this lower bound in hand, one can bound the suboptimality incurred by any feasible decentralized control policy. A study of a partially nested system reveals that affine policies can be close to optimal, even in the presence state/input constraints and non-Gaussian disturbances.**

I. Introduction

Problems of decentralized decision making naturally arise in the control of large-scale networked systems, such as electric power networks, transportation networks, supply chains and digital communication networks [1]. In such systems, the information sharing between controllers is often limited, because of the geographical separation between the system components, cost of communication, and limits on computation. As a consequence, there arises a need to decentralize the use of information in control design.

In general, the design of an optimal decentralized controller amounts to an infinite-dimensional, nonconvex optimization problem. The difficulty in solution derives in part from the manner in which information is shared between controllers – the so called information structure of a problem; see [2] for a survey. Considerable effort has been made to identify information structures under which the problem of decentralized control design can be recast as an equivalent convex optimization problem. For instance, partial nestedness of the information structure [3] is known to simplify the control design, as it eliminates the incentive to signal between controllers. In particular, linear policies are guaranteed to be optimal for decentralized LQG problems with partially nested information structures [3]–[6]. Closely related notions of quadratic invariance [7] and funnel causality [8] guarantee convexity of decentralized controller synthesis, which minimizes the closed-loop norm of an LTI system. There is also a body of literature offering insight on the structure of optimal decentralized controllers for problems with nonclassical information structures; we refer the reader to [9]–[12] for recent advances.

Many of the aforementioned results are reliant on the assumption of a Gaussian disturbance process and cannot directly accommodate explicit constraints on the state and input of the system. There is, however, another stream of literature, which directly addresses such issues by extending techniques from centralized model predictive control (MPC) control design to the decentralized setting [13]–[16]. Although the decentralized controllers they construct exhibit good performance in practice, they are suboptimal in general. The question as to how far from optimal such policies might be forms the basis of motivation for this paper.

The setting we consider in this paper entails the decentralized output feedback control of a discrete-time, linear time-varying system over a finite horizon; the system is subjected to coupled linear constraints on the state and input trajectories and the disturbance process is assumed to have known and bounded support described by convex conic inequalities. The problem of determining an optimal decentralized control policy for such systems is, in general, computationally intractable. In this paper, we abandon the search for optimal decentralized control policies and resort, instead, to approximation.

Relying to a great degree on recent advances in stochastic programming [17] and control [18], we explore the extent to which one might construct suboptimal (affine) policies with efficiently computable bounds on performance. Our primary contributions are two-fold. First, for problems with partially nested information structures, we show that the problem of determining an optimal affine decentralized output feedback control policy can be equivalently reformulated as a finite-dimensional conic program; whose optimal value yields an upper bound on the minimum achievable cost. Second, for problems with general (possibly *nonclassical*) information structures, we construct another finite-dimensional conic program whose optimal value is guaranteed to stand as lower bound on minimum achievable cost. To the best of our knowledge, such result is the first to offer an efficiently computable (and nontrivial) lower bound on the cost achievable by decentralized control policies in the setting considered.

The remainder of this paper is organized as follows. Section II formulates the decentralized output feedback control design problem. Section III describes a procedure for computing the optimal affine controller for problems with partially nested information structures through solution of a finite-dimensional conic program – the optimal value of which serves as a primal upper bound on the minimum achievable cost. Section IV treats problems with general in-

Supported in part by NSF grant ECCS-1351621, NSF grant CNS-1239178, PSERC under sub-award S-52, and US DoE under the CERTS initiative.

Weixuan Lin (wl476@cornell.edu) and Eilyan Bitar (eyb5@cornell.edu) are with the School of Electrical and Computer Engineering, Cornell University, Ithaca, NY, 14853, USA.

formation structures and specifies another finite-dimensional conic program whose optimal value stands as a dual lower bound on the minimum achievable cost. Section V offers a numerical analysis of a problem instance with a partially nested information structure. We omit the majority of mathematical proofs due to space constraints.

Notation: Let $\mathbf{R}$ denote the set of real numbers. Denote the transpose of a vector $x \in \mathbf{R}^n$ by x'. We use the comma operator (,) to denote vertical vector concatenation. That is, for any pair of vectors $x = (x_1, \ldots, x_n) \in \mathbf{R}^n$ and $y = (y_1, \ldots, y_m) \in \mathbf{R}^m$, we define their concatenation as $(x, y) = (x_1, \ldots, x_n, y_1, \ldots, y_m) \in \mathbf{R}^{n+m}$. Given a process $\{x(t)\}$ indexed by $t = 0, \cdots, T-1$, we denote by $x^t = (x(0), x(1), \cdots, x(t))$ its history until time t. We consider block matrices throughout the paper. Given a block matrix A whose dimension will be clear from the context, we denote by $[A]_{ij}$ its $(i,j)^{\text{th}}$ block. We denote the trace of a square matrix A by $\text{Tr}\,(A)$. Finally, we denote by $\mathcal{K}$ a proper cone (i.e., convex, closed, and pointed with an nonempty interior). Let $\mathcal{K}^*$ denote its dual cone. We write $x \succeq_{\mathcal{K}} y$ to indicate that $x - y \in \mathcal{K}$. For a matrix A of appropriate dimension, $A \succeq_{\mathcal{K}} 0$ denotes its columnwise inclusion in $\mathcal{K}$.

II. Problem Formulation

A. System Model

Consider a discrete-time, linear time-varying system consisting of N coupled subsystems whose dynamics are described by

$$x_i(t+1) = \sum_{j=1}^{N} \left(A_{ij}(t)x_j(t) + B_{ij}(t)u_j(t)\right) + G_i(t)\xi(t), \tag{1}$$

for $i = 1, \ldots, N$. The system operates for finite time $t = 0, \ldots, T-1$ and the initial condition is assumed fixed and known. We associate with each subsystem i a *local state* $x_i(t) \in \mathbf{R}^{n_x^i}$ and *local input* $u_i(t) \in \mathbf{R}^{n_u^i}$. And we denote by $\xi(t) \in \mathbf{R}^{n_\xi}$ the stochastic disturbance process whose joint probability distribution is assumed known. We denote by $y_i(t) \in \mathbf{R}^{n_y^i}$ the *local measured output* of subsystem i at time t. It is given by

$$y_i(t) = \sum_{j=1}^{N} C_{ij}(t)x_j(t) + H_i(t)\xi(t), \tag{2}$$

for $i = 1, \ldots, N$. All system matrices are assumed to be real and of compatible dimensions.

In the sequel, it will be convenient to work with a more compact representation of the system (1) and (2), given by

$$\begin{aligned} x(t+1) &= A(t)x(t) + B(t)u(t) + G(t)\xi(t) \\ y(t) &= C(t)x(t) + H(t)\xi(t). \end{aligned}$$

Here, we denote by $x(t) = (x_1(t), \ldots, x_N(t)) \in \mathbf{R}^{n_x}$, $u(t) = (u_1(t), \ldots, u_N(t)) \in \mathbf{R}^{n_u}$, and $y(t) = (y_1(t), \ldots, y_N(t)) \in \mathbf{R}^{n_y}$ the full system state, input, and output at time t, respectively. Their corresponding dimensions are given by $n_x = \sum_{i=1}^N n_x^i$, $n_u = \sum_{i=1}^N n_u^i$, and $n_y = \sum_{i=1}^N n_y^i$. We will occasionally refer to the tuple

$$\Theta := \{A(t), B(t), G(t), C(t), H(t)\}_{t=0}^{T-1},$$

when making reference to the parameterization of system (1) and (2). The system trajectories are related according to

$$x = Bu + G\xi \quad \text{and} \quad y = Cx + H\xi,$$

where x, u, ξ, and y denote the trajectories of the full system state, input, disturbance, and output, respectively.[1] We denote them by

$$\begin{aligned} x &= (x(0), \ldots, x(T)) \in \mathbf{R}^{N_x} \\ u &= (u(0), \ldots, u(T-1)) \in \mathbf{R}^{N_u} \\ \xi &= (1, \xi(0), \ldots, \xi(T-1)) \in \mathbf{R}^{N_\xi} \\ y &= (1, y(0), \ldots, y(T-1)) \in \mathbf{R}^{N_y}, \end{aligned}$$

where their dimensions are given by $N_x = n_x(T+1)$, $N_u = n_u T$, $N_\xi = 1 + n_\xi T$, and $N_y = 1 + n_y T$. Notice that in our specification of the both the disturbance and output trajectories, ξ and y, we have extended each trajectory to include a constant scalar as its initial component. Such notational convention will prove useful in simplifying the specification of affine control policies in the sequel.

B. Disturbance Model

In order to ensure well-posedness of the problem to follow, we require that the disturbance process satisfy the following conditions. First, we assume that the disturbance trajectory ξ has support Ξ that is a nonempty and compact subset of $\mathbf{R}^{N_\xi}$, representable by

$$\Xi = \{\xi \in \mathbf{R}^{N_\xi} \mid \xi_1 = 1 \text{ and } W_k\xi \succeq_{\mathcal{K}} 0,\ k = 1, \ldots, \ell\},$$

where $\mathcal{K} \subseteq \mathbf{R}^{N_\xi}$ is a proper cone and the matrices $W_k \in \mathbf{R}^{N_\xi \times N_\xi}$ are assumed known for $k = 1, \ldots, \ell$. In addition to compactness, we require that the linear hull of Ξ spans $\mathbf{R}^{N_\xi}$. Such assumption is without loss of generality. And, it is straightforward to verify that such assumption ensures that the corresponding second-order moment matrix, defined as $M := \mathbf{E}\left(\xi\xi'\right)$, is both invertible and positive definite. The fact that the second-order moment matrix M is finite-valued is a consequence of our assumption that the disturbance have compact support. We remark that our choice of disturbance model represents a departure from the standard assumption of Gaussianity in problems of decentralized stochastic control.

C. System Constraints

In characterizing the set of feasible input trajectories, we require that the input and resulting state trajectory respect the following infinite collection of linear inequality constraints,

$$F_x x + F_u u + F_\xi \xi \leq 0, \quad \text{for all} \ \ \xi \in \Xi, \tag{3}$$

[1] The block matrices $(A(t), B(t), G(t), C(t), H(t))$ for $t = 0, \ldots, T-1$ are readily constructed from the primitive problem data defining (1) and (2). The explicit specification of (B, G, C, H) can be found in Appendix A.

where $F_x \in \mathbf{R}^{m\times N_x}$, $F_u \in \mathbf{R}^{m\times N_u}$, and $F_\xi \in \mathbf{R}^{m\times N_\xi}$ are assumed arbitrary. Namely, the input u and resulting state trajectory x should satisfy the m linear inequality constraints, whatever the realization of the disturbance ξ. We will refer to input trajectories satisfying (3) as being *robustly feasible*. The challenge in determining a robustly feasible input trajectory derives in part from the need to enforce the infinitely many constraints defined in (3). This difficulty is exasperated by the fact that input trajectories will, in general, be allowed to depend causally on output trajectories through arbitrary functions belonging to an infinite-dimensional space.

D. Decentralized Control Design

At each time t, each subsystem must determine its local input based on the local information to which it has access. We describe the pattern according to which information is shared between subsystems with a directed graph $\mathcal{G}_I = (\mathcal{V}, \mathcal{E}_I)$, which we refer to as the *information graph* of the system. Here, the node set $\mathcal{V} := \{1, \ldots, N\}$ assigns a distinct node i to each subsystem i, and the directed edge set $\mathcal{E}_I$ determines the pattern of information sharing between subsystems. More precisely, we let $(i, j) \in \mathcal{E}_I$ if and only if for each time t, subsystem j has access to subsystem i's local output measurement $y_i(t)$. We include self-loops (i, i) in the edge set $\mathcal{E}_I$ to capture the implicit assumption that each subsystem i has access to its local output measurement $y_i(t)$ at each time t. We also assume that each subsystem has *perfect recall*; that is to say that each subsystem has access to its entire history of past information at any given time. We thus define the *local information* available to each subsystem i at time t as

$$z_i(t) := \{y_j^t \mid j \in N^-_{\mathcal{G}_I}(i)\}, \tag{4}$$

and restrict its local control input to be of the form

$$u_i(t) = \gamma_i(z_i(t), t) \tag{5}$$

– a causal measurable function $\gamma_i(\cdot, t)$ of its local information. Here, we take $N^-_{\mathcal{G}_I}(i) := \{j \in \mathcal{V} \mid (j, i) \in \mathcal{E}_I\}$ to denote the set of nodes belonging to the *in-neighborhood* of node i, according to the directed graph $\mathcal{G}_I$. We define the *local control policy* for subsystem i as $\gamma_i = (\gamma_i(\cdot, 0), \ldots, \gamma_i(\cdot, T-1))$. We refer to the collection of local control policies $\gamma = (\gamma_1, \ldots, \gamma_N)$ as the *decentralized control policy* and define $\Gamma(\mathcal{G}_I)$ as the family of all decentralized control policies respecting the information structure, (4) and (5), induced by the information graph $\mathcal{G}_I$. Henceforth, we will use the information graph $\mathcal{G}_I$ to denote the *information structure* associated with a decentralized control policy $\gamma \in \Gamma(\mathcal{G}_I)$.[2]

We measure the performance of a decentralized control policy γ according to the expected quadratic cost

$$J(\gamma) := \mathbf{E}^\gamma \left(x' R_x x + u' R_u u\right), \tag{6}$$

where expectation is taken with respect to the joint distribution on (x, u) induced by the choice of policy γ. The *cost matrices* $R_x \in \mathbf{R}^{N_x\times N_x}$ and $R_u \in \mathbf{R}^{N_u\times N_u}$ are assumed to be symmetric positive semidefinite and positive definite, respectively. Of interest is the characterization of decentralized control policies, which minimize the expected cost criterion (6) while respecting the informational and system constraints. We define the *decentralized control design problem* as

$$\begin{array}{rl} \text{minimize} & \mathbf{E}^\gamma \left(x' R_x x + u' R_u u\right) \\ \text{subject to} & \gamma \in \Gamma(\mathcal{G}_I) \\ & \left.\begin{array}{l} F_x x + F_u u + F_\xi \xi \le 0 \\ x = Bu + G\xi \\ y = Cx + H\xi \\ u = \gamma(y) \end{array}\right\} \forall\, \xi \in \Xi. \end{array} \tag{7}$$

The decentralized control policy $\gamma^* \in \Gamma(\mathcal{G}_I)$ is said to be *optimal* if it is robustly feasible and $J(\gamma^*) \le J(\gamma)$ for all robustly feasible policies $\gamma \in \Gamma(\mathcal{G}_I)$. We denote the *optimal value* of problem (7) by $J^* := J(\gamma^*)$ for γ^* optimal.

In general, the decentralized control design problem (7) amounts to an infinite-dimensional, nonconvex optimization problem with neither analytical nor computationally efficient solution available at present time [1], [2], [18], [19]. The difficulty in solution derives in part from the partial observation of state and decentralization of information among the different controllers. In what follows, we abandon the search for optimal decentralized control policies. We instead resort to approximation and explore the extent to which one might construct suboptimal policies with efficiently computable bounds on performance.

III. Affine Policies and a Primal Upper Bound

The partial nesting of information is known to simplify problems of decentralized control design, as it eliminates the incentive to signal between controllers [2]–[4]. In this section, we restrict our attention to finite-dimensional decentralized control policies that are *affine* in the measured output, and explore the extent to which the partial nesting of information might facilitate the efficient optimization over such restricted class of policies. We demonstrate how powerful techniques for centralized affine control design [18], [20]–[22] can be extended to decentralized systems to both compute optimal affine decentralized policies and bound their loss optimality via finite-dimensional convex optimization. In particular, our approach to bounding the loss of optimality incurred by affine decentralized controllers relies centrally on techniques developed in [17], [18].

A. Purification of Partially Nested Information

In what follows, we introduce the standard concept of *output purification* [18], [20] and show that, under a partially nested information structure, the output process and the purified output process generate the same information for each subsystem. One can thus reparameterize the decentralized control policy in the purified output process without loss of optimality. The advantage in doing so derives from the purified output process's independence from the underlying control policy. Before proceeding, we provide a formal definition of partially nested information structures using the notion of precedence, as defined by Ho and Chu [3].

[2]We remark that the class information structures considered in this paper cannot accommodate delay constraints on information sharing.

Definition 1 (Precedence)**.** Given the information structure defined by $\mathcal{G}_I$, we say subsystem j is a *precedent* to subsystem i, denoted by $j \prec i$, if there exist times $0 \leq s < t \leq T-1$ and subsystem $k \in N^-_{\mathcal{G}_I}(i)$ such that $\left[C(t)A^t_{s+1}B(s)\right]_{kj} \neq 0$.[3]

Recall that $\left[C(t)A^t_{s+1}B(s)\right]_{kj}$ denotes the $(k,j)^{\text{th}}$ block of $C(t)A^t_{s+1}B(s)$, which is an $N \times N$ block matrix. Essentially, j is *precedent* to i, if the control input at subsystem j can affect the local information available to subsystem i in the future.

Definition 2 (Partially Nested Information)**.** The information structure defined by $\mathcal{G}_I$ is *partially nested* with respect to the system Θ, if $j \prec i \implies z_j(t) \subseteq z_i(t)$ for all times $t = 0, \ldots, T-1$.

We denote by PN(Θ) the set of information graphs that are partially nested with respect to the the system Θ. We note that the above definition of partial nestedness is tailored for the information structure considered in this paper. A more general definition of partial nestedness can be found in [3], [4], and one can refer to [23] for an investigation on how partial nestedness is manifested under different information structures.

Next, we introduce the concept of output purification as defined in [20]. Given an input process $u(t)$ and its corresponding output process $y(t)$, define the sequence of purified outputs $\eta(t)$ according to

$$\begin{aligned} \bar{x}(0) &= 0, \\ \bar{x}(t+1) &= A(t)\bar{x}(t) \;+\; B(t)u(t), \\ \bar{y}(t) &= C(t)\bar{x}(t), \\ \eta(t) &= y(t) \;-\; \bar{y}(t), \end{aligned}$$

for $t = 0, \cdots, T-1$. Similar to the definition of the *local information* in (4), we define the *local purified information* available to each subsystem i at time t as

$$\zeta_i(t) = \{\eta^t_j \mid j \in N^-_{\mathcal{G}_I}(i)\}.$$

In addition, we define the trajectory of the purified output according to $\eta = (1, \eta(0), \ldots, \eta(T-1)) \in \mathbf{R}^{N_y}$. It is straightforward to establish the following relation, which reveals the purified output trajectory η to be independent of the input trajectory u. Namely,

$$\eta = P\xi,$$

where we define the matrix $P := (CG + H) \in \mathbf{R}^{N_y \times N_\xi}$.

We establish the following important Lemma, which reveals that, given a partially nested information structure, the local information $z_i(t)$ and purified local information $\zeta_i(t)$ contain the same information for each subsystem i and time t. In other words, they generate the same σ-algebras.

Lemma 1 (Equivalence of Information)**.** Let $\gamma \in \Gamma(\mathcal{G}_I)$ be any decentralized control policy. If $\mathcal{G}_I \in$ PN(Θ), then the local information $z_i(t)$ and purified local information $\zeta_i(t)$ are functions of each other for each subsystem $i = 1, \cdots, N$ and time $t = 0, \cdots, T-1$.

We defer the proof of Lemma 1 to the Appendix. A consequence of Lemma 1 is that, for problems with a partially nested information structure, a reparameterization of the decentralized control policy in the purified output process (i.e., $u = \gamma(\eta)$ for $\gamma \in \Gamma(\mathcal{G}_I)$) is without loss of optimality.

B. Primal Affine Decentralized Control Policies

In this section, we restrict our attention to decentralized control policies that are *affine* in the measured output. The ability to solve the decentralized control design problem (7) given such restriction will yield a robustly feasible suboptimal policy whose expected cost stands as an upper bound on the optimal value of (7). The restriction to affine policies, alone, does not however circumvent the problem of nonconvexity in decentralized control design, due to the possibility for signaling between controllers under general information structures. In what follows, we show that – given a partially nested information structure and restriction to affine output feedback policies – the decentralized control design problem (7) can be equivalently reformulated as a semi-infinite robust convex program. And, given our assumption that the disturbance has compact support described by conic inequalities, this semi-infinite robust convex program can be equivalently reformulated as a finite-dimensional conic optimization problem.

Given an information structure $\mathcal{G}_I$, we consider affine control policies of the form

$$u_i(t) = \bar{u}_i(t) + \sum_{s=0}^{t} \sum_{j \in \mathcal{N}^-_{\mathcal{G}_I}(i)} K_{ij}(t,s)y_j(s), \tag{8}$$

for each subsystem $i = 1, \cdots, N$ and time $t = 0, \cdots, T-1$. Here, $\bar{u}_i(t) \in \mathbf{R}^{n^i_u}$ represents the open-loop component of the control and $K_{ij}(t,s) \in \mathbf{R}^{n^i_u \times n^j_y}$ the feedback control gain. One can lift the representation in (8) to relate the output trajectory y to the input trajectory u under the linear map

$$u = Ky, \quad \text{where} \quad K \in S(\mathcal{G}_I).$$

Here, were require the gain $K \in \mathbf{R}^{N_u \times N_y}$ to belong to $S(\mathcal{G}_I)$, which we define to be the linear space of causal (lower block triangular) matrices respecting the information structure $\mathcal{G}_I$. That is, for any $K \in S(\mathcal{G}_I)$, the decentralized control policy defined by $\gamma(y) = Ky$ satisfies $\gamma \in \Gamma(\mathcal{G}_I)$. Given the restriction to decentralized control policies that are *affine* in the measured output, we have the following reformulation of the decentralized control design problem (7) as

$$\begin{aligned} \text{minimize} \quad & \mathbf{E}\left(x'R_x x + u'R_u u\right) \\ \text{subject to} \quad & K \in S(\mathcal{G}_I) \\ & \left.\begin{aligned} & F_x x + F_u u + F_\xi \xi \leq 0 \\ & x = Bu + G\xi \\ & y = Cx + H\xi \\ & u = Ky \end{aligned}\right\} \forall\; \xi \in \Xi. \end{aligned} \tag{9}$$

[3]We refer the reader to Appendix A for a definition of the matrix A^t_{s+1}.

We denote by J^p the optimal value of problem (9), which clearly holds as a *primal upper bound* on the optimal value of problem (7). Namely, $J^* \leq J^p$. The affine control design problem (9) is known to be nonconvex in the matrix variable K [18], [24], [25]. However, under the additional assumption of partially nested information, one can apply a suitable change of variables to obtain an equivalent reformulation of problem (9) as a semi-infinite robust convex program. The underlying technique is equivalent in nature to the classical Youla parameterization [26]. We have the following proposition.

Proposition 1. If $\mathcal{G}_I \in \text{PN}(\Theta)$, then the following statements are true.

(i) Let $K \in S(\mathcal{G}_I)$ and define $Q = K(I - CBK)^{-1}$. Then $Q \in S(\mathcal{G}_I)$ and $Q\eta = Ky$ for all $\xi \in \Xi$.

(ii) Let $Q \in S(\mathcal{G}_I)$ and define $K = (I + QCB)^{-1}Q$. Then $K \in S(\mathcal{G}_I)$ and $Ky = Q\eta$ for all $\xi \in \Xi$.

We refer the reader to the Appendix for a proof of the above proposition. Proposition 1 builds on Lemma 1 to reveal that if the information structure is partially nested, then any decentralized affine output feedback controller $K \in S(\mathcal{G}_I)$ can be transformed to an *equivalent* decentralized affine purified output feedback controller $Q \in S(\mathcal{G}_I)$ through an invertible nonlinear transformation, and vice versa.

Proposition 2. Let Q^* be an optimal solution to the following optimization problem,

$$\begin{array}{ll} \text{minimize} & \mathbf{E}\left(x'R_x x + u'R_u u\right) \\ \text{subject to} & Q \in S(\mathcal{G}_I) \\ & \left.\begin{array}{l} F_x x + F_u u + F_\xi \xi \leq 0 \\ x = Bu + G\xi \\ u = QP\xi \end{array}\right\} \forall\, \xi \in \Xi. \end{array} \tag{10}$$

Then $K^* = (I + Q^*CB)^{-1}Q^*$ is an optimal solution to problem (9).

Proof: By Proposition 1, we have that for any $Q \in S(\mathcal{G}_I)$, $K = (I + QCB)^{-1}Q$ satisfies $K \in S(\mathcal{G}_I)$. If Q^* solves problem (10), then by Proposition 1, $K^* = (I+Q^*CB)^{-1}Q^*$ is the affine output feedback controller that results in the same sequence of control inputs as the affine purified output feedback controller Q^*, so it is the optimal affine output feedback controller. ■

In the absence of constraints on the state and input trajectories (i.e., $F_x, F_u, F_\xi = 0$), problem (10) reduces to an unconstrained convex quadratic program. Problem (10) is in general, however, a semi-infinite convex quadratic program, as it contains infinitely many linear constraints in Q. Given our assumption that the set Ξ is described by finitely many conic inequalities, one can use techniques grounded in duality to show that problem (10) admits an equivalent reformulation as a finite-dimensional conic optimization problem. The underlying approach relies on arguments analogous to those in [17], [18].

Proposition 3. An optimal solution to problem (10) can be obtained by solving the following equivalent finite-dimensional conic optimization problem,

$$\begin{array}{ll} \text{minimize} & \text{Tr}\left(P'Q'RQPM + 2G'R_xBQPM + G'R_xGM\right) \\ \text{subject to} & Q \in S(\mathcal{G}_I) \\ & Z \in \mathbf{R}^{m \times N_\xi}, \quad \Lambda_k \in \mathbf{R}^{N_\xi \times m}, \quad \mu \in \mathbf{R}^m_+ \\ & (F_u + F_x B)QP + F_x G + F_\xi + Z = 0, \\ & Z = \mu e_1' + \sum_{k=1}^{\ell} \Lambda_k' W_k, \\ & \Lambda_k \succeq_{\mathcal{K}^*} 0, \quad k = 1, \dots, \ell, \end{array} \tag{11}$$

where $R = R_u + B'R_xB$, and $e_1 = (1, 0, \dots, 0)$ is a unit vector in $\mathbf{R}^{N_\xi}$.

The proof is straightforward, as it mirrors that of Proposition 3.2 in [18]. It is thus omitted due to space constraints. We remark that the conic optimization problem (11) can be efficiently solved for a wide range of cones $\mathcal{K}$, including polyhedral and second-order cones.

IV. A Dual Lower Bound

The restriction to decentralized control policies that are *affine* in the measured output will in general result in the loss of optimality with respect to the original problem of interest (7). That is to say that $J^p \geq J^*$. In this section, we offer a generalization of the lower bounding technique in [18] to enable the efficient computation of a lower bound on the optimal value J^*. With such bound in hand, one can estimate the suboptimality incurred by any feasible decentralized control policy. We begin this section by specifying the calculation of this lower bound for problems with partially nested information structures. In Section IV-B, we generalize the lower bound to problems with nonclassical information structures, through application of a *minimal* information relaxation.

A. Partially Nested Information Structures

We require an additional assumption on the disturbance process in order to derive the dual lower bound.

Assumption 1. The disturbance trajectory ξ is distributed according to an elliptically contoured distribution.

The family of elliptically contoured distributions is broad. It includes the multivariate Gaussian distribution, multivariate t-distribution, their truncated versions, and uniform distributions on ellipsoids. If ξ follows an elliptically contoured distribution, then the conditional expectation of ξ given a subvector of ξ, is affine in this subvector. And any linear transformation of ξ also follows an elliptically contoured distribution [27]. Such properties play an integral role in the derivation of the lower bound on the optimal value of problem (7), which we state formally in the following result. Its proof is similar to that of Proposition 4.2 in [18], with an important modification required to accommodate the decentralization of information.

Proposition 4. Let Assumption 1 hold. If $\mathcal{G}_I \in \text{PN}(\Theta)$, then the optimal value of the following problem is a lower bound on the optimal value of problem (7):

$$\begin{array}{ll} \text{minimize} & \text{Tr}\,(P'Q'RQPM + 2G'R_xBQPM + G'R_xGM) \\ \text{subject to} & Q \in S(\mathcal{G}_I), \quad Z \in \mathbf{R}^{m \times N_\xi} \\ & (F_u + F_xB)QP + F_xG + F_\xi + Z = 0, \\ & W_kMZ' \succeq_{\mathcal{K}} 0, \ k = 1, \ldots, \ell, \\ & e_1'MZ' \geq 0, \end{array} \tag{12}$$

where $R = R_u + B'R_xB$, and $e_1 = (1, 0, \ldots, 0)$ is a unit vector in $\mathbf{R}^{N_\xi}$.

Given an information graph $\mathcal{G}_I$, we denote by $J^d(\mathcal{G}_I)$ the optimal value of problem (12). It stands as a *dual lower bound* on the optimal value of problem (7). Namely,

$$J^d(\mathcal{G}_I) \leq J^* \leq J^p.$$

The reason for making explicit the dependency of $J^d(\mathcal{G}_I)$ on the information graph $\mathcal{G}_I$ will be made apparent in the sequel.

We have an immediate corollary to Propositions 3 and 4. In the absence of constraints on state or input, affine policies are optimal for problem (7), assuming Assumption 1 to hold. The fact that affine policies are optimal for partially nested LQG problems is a special case of Corollary 1. Moreover, Corollary 1 recovers the classical result of Chu [28], which proves optimality of affine policies for systems with elliptically contoured disturbance distributions.

Corollary 1 (Optimality of Affine Policies)**.** Let Assumption 1 hold. If $\mathcal{G}_I \in \text{PN}(\Theta)$ and $F_x, F_u, F_\xi = 0$, then affine policies are optimal for problem (7) and $J^d(\mathcal{G}_I) = J^* = J^p$.

B. General Information Structures

We now construct a lower bound on J^* for problems with arbitrary information structures. Our approach is simple. Given an arbitrary information graph $\mathcal{G}_I$, we identify the *smallest* directed graph $\widehat{\mathcal{G}}_I$ containing $\mathcal{G}_I$, such that $\widehat{\mathcal{G}}_I \in \text{PN}(\Theta)$. One can then apply the results of Proposition 4 to obtain a valid lower bound $J^d(\widehat{\mathcal{G}}_I) \leq J^*$. In a sense, this approach amounts to constructing an information relaxation to a partially nested structure that introduces the smallest number of additional edges to the information graph. We now describe an approach to constructing such information relaxations. First, we require a nonrestrictive assumption.

Assumption 2. For any information graph $\mathcal{G}_I = (\mathcal{V}, \mathcal{E}_I)$ and system Θ, we require that $i \prec i$ for all $i \in \mathcal{V}$.

Assumption 2 essentially requires that the control input to each subsystem will causally affect its own output. Given an information graph $\mathcal{G}_I$ and system Θ, we define the *smallest information relaxation* yielding a partially nested information structure as the optimal solution of the following problem:

$$\underset{\mathcal{G} \supseteq \mathcal{G}_I}{\text{minimize}} \ |\mathcal{E}| \quad \text{subject to} \quad \mathcal{G} \in \text{PN}(\Theta), \tag{13}$$

where $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, and $|\mathcal{E}|$ is the cardinality of the directed edge set $\mathcal{E}$. The containment condition $\mathcal{G} \supseteq \mathcal{G}_I$ requires that any feasible solution $\mathcal{G}$ to problem (13) be a supergraph of $\mathcal{G}_I$ and have vertex set $\mathcal{V} = \{1, \ldots, N\}$.

The problem we consider in (13) is similar in spirit to [29], which considers the problem of finding the smallest relaxation of the information constraint which is *quadratically invariant* when the original problem may not be. When the information constraint is defined in terms of sparsity constraints on the measurements that each controller has access to, the authors specify an iterative method that is guaranteed to recover the minimal relaxation in a fixed number of steps. Problem (13) admits an explicit solution. We first require several definitions.

Definition 3 (Precedence Graph)**.** The *precedence graph* associated with the system Θ and the information graph $\mathcal{G}_I$ is defined as the directed graph $\mathcal{G}_P(\Theta, \mathcal{G}_I) = (\mathcal{V}, \mathcal{E}_P(\Theta, \mathcal{G}_I))$, whose directed edge set is defined as

$$\mathcal{E}_P(\Theta, \mathcal{G}_I) := \{(i, j) \mid i, j \in \mathcal{V}, \ i \prec j \text{ with respect to } (\Theta, \mathcal{G}_I)\}.$$

Essentially, the precedence graph gives a directed graphical representation of the precedence relations between all subsystems, as specified in Definition 1.

Definition 4 (Transitive Closure)**.** The *transitive closure* of a directed graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ is defined as the directed graph $\overline{\mathcal{G}} = (\mathcal{V}, \overline{\mathcal{E}})$, where $(i, j) \in \overline{\mathcal{E}} \iff$ there exists a directed path in $\mathcal{G}$ from node i to node j.

We remark that the transitive closure of a directed graph is readily computed using Warshall's algorithm [30]. Equipped with these concepts, we have the following result, which offers a 'closed-form' solution for problem (13).

Proposition 5 (Information Relaxation)**.** Let Assumption 2 hold. The optimal solution to (13) is given by the *transitive closure of the precedence graph* $\mathcal{G}_P(\Theta, \mathcal{G}_I)$.

While we omit the proof of Proposition 5 due to space constraints, we mention that a key step in its derivation involves showing that $\mathcal{G}_I \in \text{PN}(\Theta)$ if and only if it equals the transitive closure of its precedence graph $\mathcal{G}_P(\Theta, \mathcal{G}_I)$ – a result which is closely related to the necessary and sufficient graph theoretic condition for quadratic invariance in [31].

Implicit in the statement of Proposition 5 is the fact that the transitive closure of the precedence graph $\mathcal{G}_P(\Theta, \mathcal{G}_I)$ not only yields a partially nested information relaxation, but also the smallest such relaxation. A trivial consequence of Proposition 5 is that $J^d\big(\overline{\mathcal{G}_P(\Theta, \mathcal{G}_I)}\big) \leq J^*$, which is a valid lower bound on the optimal value of a decentralized control problem with arbitrary information graph $\mathcal{G}_I$.

V. Case Study

Consider a linear time-invariant system consisting of $N = 5$ subsystems. Its information graph $\mathcal{G}_I$ is depicted in Figure 1. One can easily show the information structure to be partially nested, as $\mathcal{G}_I = \overline{\mathcal{G}_P(\Theta, \mathcal{G}_I)}$.

The dimensions of the state and output variables are given by $n_x^i = n_y^i = 1$ for all $i = 1, \ldots, N$. The dimensions of the input and disturbance variables are given by $n_u^1 = 2$, $n_u^2 = 3$,

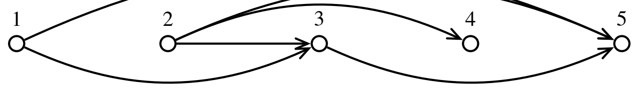

Fig. 1: The directed graph above depicts the information graph $\mathcal{G}_I$ of the test system. Although omitted, each node is also assumed to possess a self-loop.

$n_u^3 = 2$, $n_u^4 = 1$, $n_u^5 = 1$, and $n_\xi = 10$, respectively. The system is described by the following set of matrices:

$$A(t) = \begin{bmatrix} 0.8 & 0 & 0 & 0 & 0 \\ 0 & 0.9 & 0 & 0 & 0 \\ 0.2 & 0.05 & 0.6 & 0 & 0 \\ 0 & 0.05 & 0 & 1 & 0 \\ 0 & 0 & 0.4 & 0 & 1 \end{bmatrix},$$

$$B(t) = \begin{bmatrix} -1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & -1 & -1 & -1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 & -1 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & -1 \end{bmatrix},$$

$$C(t) = I_5, \; G(t) = \begin{bmatrix} I_5 & 0_{5\times 5} \end{bmatrix}, \; H(t) = \begin{bmatrix} 0_{5\times 5} & I_5 \end{bmatrix},$$

for all $t = 0, \dots, T-1$, where I_5 is the 5-by-5 identity matrix, and $0_{5\times 5}$ is a 5-by-5 matrix of all zeros. The constraints on the state and input are defined as

$$\|x\|_\infty \leq 5 \;\text{ and }\; \|u\|_\infty \leq 2, \quad \forall\, \xi \in \Xi.$$

For a given time horizon T, the support Ξ of the disturbance trajectory ξ is defined as

$$\Xi := \left\{ \xi \in \mathbf{R}^{N_\xi} \;\middle|\; \|\xi\|_2^2 \leq T+1, \; \xi_1 = 1 \right\}.$$

We assume that ξ has a uniform distribution over Ξ. It follows that the distribution of ξ is elliptically contoured, and its second moment matrix is given by

$$M = \operatorname{diag}\left(1, \frac{T}{N_\xi + 1} I_{N_\xi - 1}\right).$$

Finally, we define the system cost matrices by $R_x = \operatorname{diag}(R_x(0), \dots, R_x(T))$ and $R_u = \operatorname{diag}(R_u(0), \dots, R_u(T-1))$, where $R_x(t) = \operatorname{diag}(5, 5, 15, 5, 15)$ and $R_u(t) = \operatorname{diag}(1, 0.05, 1, 0.05, 0.05, 1, 0.05, 1, 1)$ for $t = 0, \dots, T$.

A. Numerical Results

We let the time horizon range from $T = 5$ to 10, and plot the primal upper bound J^p and the dual lower bound J^d as a function of the time horizon T in Figure 2. Roughly, both J^p and J^d appear to grow linearly with the horizon T. Although a clear gap is observed between the primal upper bound and the dual lower bound, this gap is small relative to the dual lower bound, indicating that affine control policies are close to optimal for this particular problem instance.

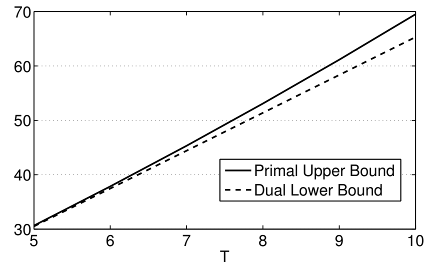

Fig. 2: A plot of the primal upper bound (J^p) and dual lower bound (J^d) as a function of the horizon T.

REFERENCES

[1] N. R. Sandell Jr, P. Varaiya, M. Athans, and M. G. Safonov, "Survey of decentralized control methods for large scale systems," *Automatic Control, IEEE Transactions on*, vol. 23, no. 2, pp. 108–128, 1978.

[2] A. Mahajan, N. C. Martins, M. C. Rotkowitz, and S. Yuksel, "Information structures in optimal decentralized control," in *Decision and Control (CDC), 2012 IEEE 51st Annual Conference on*. IEEE, 2012, pp. 1291–1306.

[3] Y. C. Ho and K.-h. Chu, "Team decision theory and information structures in optimal control problems–part i," *Automatic Control, IEEE Transactions on*, vol. 17, no. 1, pp. 15–22, 1972.

[4] A. Gattami, "Optimal decisions with limited information," Ph.D. dissertation, Lund University, 2007.

[5] A. Rantzer, "Linear quadratic team theory revisited," in *American Control Conference, 2006*. IEEE, 2006, pp. 5–pp.

[6] ——, "A separation principle for distributed control," in *Decision and Control, 2006 45th IEEE Conference on*. IEEE, 2006, pp. 3609–3613.

[7] M. Rotkowitz and S. Lall, "A characterization of convex problems in decentralized control," *Automatic Control, IEEE Transactions on*, vol. 51, no. 2, pp. 274–286, 2006.

[8] B. Bamieh and P. G. Voulgaris, "A convex characterization of distributed control problems in spatially invariant systems with communication constraints," *Systems & Control Letters*, vol. 54, no. 6, pp. 575–583, 2005.

[9] A. Mahajan and A. Nayyar, "Sufficient statistics for linear control strategies in decentralized systems with partial history sharing," 2014.

[10] A. Nayyar, A. Mahajan, and D. Teneketzis, "Optimal control strategies in delayed sharing information structures," *Automatic Control, IEEE Transactions on*, vol. 56, no. 7, pp. 1606–1620, 2011.

[11] ——, "Decentralized stochastic control with partial history sharing: A common information approach," *Automatic Control, IEEE Transactions on*, vol. 58, no. 7, pp. 1644–1658, 2013.

[12] N. R. Sandell Jr and M. Athans, "Solution of some nonclassical LQG stochastic decision problems," *Automatic Control, IEEE Transactions on*, vol. 19, no. 2, pp. 108–116, 1974.

[13] A. Bemporad, M. Heemels, and M. Johansson, *Networked control systems*. Springer, 2010, vol. 406.

[14] E. Camponogara, D. Jia, B. H. Krogh, and S. Talukdar, "Distributed model predictive control," *Control Systems, IEEE*, vol. 22, no. 1, pp. 44–52, 2002.

[15] D. Jia and B. H. Krogh, "Distributed model predictive control," in *American Control Conference, 2001. Proceedings of the 2001*, vol. 4. IEEE, 2001, pp. 2767–2772.

[16] T. Keviczky, F. Borrelli, and G. J. Balas, "Decentralized receding horizon control for large scale dynamically decoupled systems," *Automatica*, vol. 42, no. 12, pp. 2105–2115, 2006.

[17] D. Kuhn, W. Wiesemann, and A. Georghiou, "Primal and dual linear decision rules in stochastic and robust optimization," *Mathematical Programming*, vol. 130, no. 1, pp. 177–209, 2011.

[18] M. J. Hadjiyiannis, P. J. Goulart, and D. Kuhn, "An efficient method to estimate the suboptimality of affine controllers," *Automatic Control, IEEE Transactions on*, vol. 56, no. 12, pp. 2841–2853, 2011.

[19] J. N. Tsitsiklis and M. Athans, "On the complexity of decentralized decision making and detection problems," *Automatic Control, IEEE Transactions on*, vol. 30, no. 5, pp. 440–446, 1985.

[20] A. Ben-Tal, S. Boyd, and A. Nemirovski, "Control of uncertainty-affected discrete time linear systems via convex programming," Minerva Optimization Center, Technion, Haifa, Israel, Tech. Rep., 2005.

[21] D. Van Hessem and O. Bosgra, "A conic reformulation of model predictive control including bounded and stochastic disturbances under state and input constraints," in *Decision and Control, 2002, Proceedings of the 41st IEEE Conference on*, vol. 4. IEEE, 2002, pp. 4643–4648.
[22] C. Wang, C.-J. Ong, and M. Sim, "Model predictive control using segregated disturbance feedback," *Automatic Control, IEEE Transactions on*, vol. 55, no. 4, pp. 831–840, 2010.
[23] A. Nayyar and L. Lessard, "Optimal control for LQG systems on graphs—part i: Structural results," *arXiv preprint arXiv:1408.2551*, 2014.
[24] A. Ben-Tal, S. Boyd, and A. Nemirovski, "Extending scope of robust optimization: Comprehensive robust counterparts of uncertain problems," *Mathematical Programming*, vol. 107, no. 1-2, pp. 63–89, 2006.
[25] J. Skaf and S. P. Boyd, "Design of affine controllers via convex optimization," *Automatic Control, IEEE Transactions on*, vol. 55, no. 11, pp. 2476–2487, 2010.
[26] D. C. Youla, H. Jabr, J. J. Bongiorno Jr *et al.*, "Modern wiener-hopf design of optimal controllers–part ii: The multivariable case," *Automatic Control, IEEE Transactions on*, vol. 21, no. 3, pp. 319–338, 1976.
[27] S. Cambanis, S. Huang, and G. Simons, "On the theory of elliptically contoured distributions," *Journal of Multivariate Analysis*, vol. 11, no. 3, pp. 368–385, 1981.
[28] K. C. Chu, "Estimation and decision for linear systems with elliptical random processes," *IEEE Transactions on Automatic Control*, no. 5, pp. 499–505, 1973.
[29] M. C. Rotkowitz and N. C. Martins, "On the nearest quadratically invariant information constraint," *Automatic Control, IEEE Transactions on*, vol. 57, no. 5, pp. 1314–1319, 2012.
[30] S. Warshall, "A theorem on boolean matrices," *Journal of the ACM (JACM)*, vol. 9, no. 1, pp. 11–12, 1962.
[31] J. Swigart and S. Lall, "A graph-theoretic approach to distributed control over networks," in *Decision and Control, 2009 held jointly with the 2009 28th Chinese Control Conference. CDC/CCC 2009. Proceedings of the 48th IEEE Conference on*. IEEE, 2009, pp. 5409–5414.

APPENDIX

A. Matrix Definitions

The block matrices (B, G, C, H) are given by:

$$B := \begin{bmatrix} 0 & & & & \\ A_1^1 B(0) & 0 & & & \\ A_1^2 B(0) & A_2^2 B(1) & 0 & & \\ \vdots & & & \ddots & \\ \vdots & & & & 0 \\ A_1^T B(0) & A_2^T B(1) & \cdots & \cdots & A_T^T B(T-1) \end{bmatrix}$$

$$G := \begin{bmatrix} A_0^0 x(0) & & & & \\ A_0^1 x(0) & A_1^1 G(0) & & & \\ A_0^2 x(0) & A_1^2 G(0) & A_2^2 G(1) & & \\ \vdots & \vdots & & \ddots & \\ A_0^T x(0) & A_1^T G(0) & A_2^T G(1) & \cdots & A_T^T G(T-1) \end{bmatrix}$$

$$C := \begin{bmatrix} 0 & & & \\ C(0) & 0 & & \\ & \ddots & \ddots & \\ & & C(T-1) & 0 \end{bmatrix}$$

$$H := \mathrm{diag}(1, H(0), \ldots, H(T-1)),$$

where $A_s^t := \prod_{r=s}^{t-1} A(r)$ for $s < t$, and $A_t^t = I$.

B. Proof of Lemma 1

We prove by induction in t that for any $\gamma \in \Gamma(\mathcal{G}_I)$, $z_i(t)$ and $\zeta_i(t)$ are functions of each other for all $i = 1, \ldots, N$, $t = 0, \ldots, T-1$. For the base step $t = 0$, we have that $y(0) = \eta(0)$, so the above claim is true.

Assume that the claim is true for times $0, \ldots, t-1$. We now show that the claim is also true for time t. It suffices to show that for all $i \in \{1, \ldots, N\}$ and $k \in N^-_{\mathcal{G}_I}(i)$, $y_k(t)$ is a function of $\zeta_i(t)$, and $\eta_k(t)$ is a function of $z_i(t)$.

Using the definition of precedents, we can write $y_k(t)$ as

$$y_k(t) = \left(\sum_{s=0}^{t-1} \sum_{j:\, j \prec i} \left[C(t) A^t_{s+1} B(s) \right]_{kj} u_j(s) \right) + \eta_k(t). \tag{14}$$

for each $i \in \{1, \ldots, N\}$ and $k \in N^-_{\mathcal{G}_I}(i)$. Hence, to show that $y_k(t)$ is a function of $\zeta_i(t)$, it suffices to show that for all $j \prec i$ and $s \leq t-1$, it holds that $u_j(s)$ is a function of $\zeta_i(s)$. Since $z_j(s)$ is the information available to subsystem j at time s, we have that $u_j(s)$ is a function of $z_j(s)$. And, given the assumption that $\mathcal{G}_I \in \mathrm{PN}(\Theta)$, we have that $j \prec i$ indicates that $z_j(s) \subseteq z_i(s)$. It follows that $u_j(s)$ is a function of $z_i(s)$. Moreover, by the induction hypothesis, $z_i(s)$ is a function of $\zeta_i(s)$ for $s \leq t-1$. It follows that $u_j(s)$ is a function of $\zeta_i(s)$ for all $j \prec i$ and $s \leq t-1$. This finishes the proof that $y_k(t)$ is a function of $\zeta_i(t)$ for each $i \in \{1, \ldots, N\}$ and $k \in N^-_{\mathcal{G}_I}(i)$.

The proof that $\eta_k(t)$ is a function of $z_i(t)$ for all $i \in \{1, \ldots, N\}$ and $k \in N^-_{\mathcal{G}_I}(i)$ is identical. Hence, the claim is also true for time t, which completes the proof by induction.

C. Proof of Proposition 1

We only prove part (i) of the proposition, as the proof for part (ii) is identical. Let $K \in S(\mathcal{G}_I)$ and define $u = Ky$ for all $\xi \in \Xi$. Using the fact that system trajectories are related according to $x = Bu + G\xi$ and $y = Cx + H\xi$, we have that

$$y = CBu + (CG + H)\xi = CBKy + \eta.$$

It follows that $\eta = (I - CBK)y$. The controller $u = Ky$ can be rewritten as

$$u = K(I - CBK)^{-1}\eta,$$

where the invertibility of $(I - CBK)$ is guaranteed, since CBK is strictly block lower triangular. Let $Q = K(I - CBK)^{-1}$. It follows that $u = Ky = Q\eta$ for all $\xi \in \Xi$.

To show that $Q \in S(\mathcal{G}_I)$, we note that Lemma 1 ensures that $z_i(t)$ is a function of $\zeta_i(t)$ under any decentralized controller $\gamma \in \Gamma(\mathcal{G}_I)$. If $u = Ky$ for some $K \in S(\mathcal{G}_I)$, then $u_i(t)$ is a function of $z_i(t)$, and hence a function of $\zeta_i(t)$. Since $Q\eta = Ky$, we see that under the affine controller $u = Q\eta$, $u_i(t)$ only depends on $\zeta_i(t)$, so $Q \in S(\mathcal{G}_I)$.