2009 American Control Conference
Hyatt Regency Riverfront, St. Louis, MO, USA
June 10-12, 2009



# Distributed Estimation and Control under Partially Nested Pattern

*Invited Paper*

Ather Gattami
Laboratory for Information and Decision Systems
Massachusetts Institute of Technology
Cambridge, MA 02139, USA
E-mail: ather@mit.edu

***Abstract*— In this paper we study distributed estimation and control problems over graphs under partially nested information patterns. We show a duality result that is very similar to the classical duality result between state estimation and state feedback control with a classical information pattern.**

***Index Terms*— Distributed Estimation and State Feedback Control, Duality.**

## I. Introduction

Control with information structures imposed on the decision maker(s) have been very challenging for decision theory researchers. Even in the simple linear quadratic static decision problem, it has been shown that complex nonlinear decisions could outperform any given linear decision (see [10]). Important progress was made for the stochastic static team decision problems in [6] and [7]. New information structures were explored in [5] for the stochastic linear quadratic finite horizon control problem. The distributed stochastic linear quadratic team problem was revisited in [4], which generalizes previous results for tractability of the dynamic team decision problem with information constraints. An analog deterministic version of the stochastic team decision problem was solved in [4], which also showed that for the finite and infinite horizon linear quadratic minimax control problem with bounds on the information propagation, the optimal decision is linear and can be found by solving a linear matrix inequality. In [8], the stationary state feedback stochastic linear quadratic control problem was solved using state space formulation, under the condition that all the subsystems have a *common past*. With common past, we mean that all subsystems have information about the *global* state from some time step in the past. The problem was posed as a finite dimensional convex optimization problem. The time-varying and stationary output feedback version was solved in [3].

## II. preliminaries

### A. Notation

Let $\mathbb{R}$ be the set of real numbers. For a stochastic variable $x$, $x \sim \mathcal{N}(m, X)$ means that $x$ is a Gaussian variable with $\mathbf{E}\{x\} = m$ and $\mathbf{E}\{(x-m)(x-m)^T\} = X$.

$M_i$, or $[M]_i$, denotes either block column $i$ or block row $i$ of a matrix $M$ with proper dimensions, which should follow from the context. For a matrix $A$ partitioned into blocks, $[A]_{ij}$ denotes the block of $A$ in block position $(i, j)$. For vectors $v_k, v_{k-1}, ..., v_0$, we define $v_{[0,k]} := (v_k, v_{k-1}, ..., v_0)$.

The forward shift operator is denoted by $\mathbf{q}$. That is $x_{k+1} = \mathbf{q}x_k$, where $\{x_k\}$ is a given process. A causal linear time-invariant operator $T(\mathbf{q})$ is given by its generating function $T(\mathbf{q}) = \sum_{t=0}^{\infty} T_t \mathbf{q}^{-t}$, $T_t \in \mathbb{R}^{p\times m}$. A causal linear time-varying operator $T(k, \mathbf{q})$ is given by its generating function $T(k, \mathbf{q}) = \sum_{t=0}^{\infty} T_t(k)\mathbf{q}^{-t}$, $T_t(k) \in \mathbb{R}^{p\times m}$. A transfer matrix in terms of state-space data is denoted

$$\left(\begin{array}{c|c} A & B \\ \hline C & D \end{array}\right) := C(\mathbf{q}I - A)^{-1}B + D.$$

## III. Graph Theory

We will present in brief some graph theoretical definitions and results that could be found in the graph theory or combinatorics literature (see for example [2]). A (simple) graph $\mathcal{G}$ is an ordered pair $\mathcal{G} := (\mathcal{V}, \mathcal{E})$ where $\mathcal{V}$ is a set, whose elements are called *vertices* or *nodes*, $\mathcal{E}$ is a set of pairs (unordered) of distinct vertices, called *edges* or *lines*. The vertices belonging to an edge are called the ends, endpoints, or end vertices of the edge. The set $\mathcal{V}$ (and hence $\mathcal{E}$) is taken to be finite in this paper. The order of a graph is $|\mathcal{V}|$ (the number of vertices). A graph's size is $|\mathcal{E}|$ , the number of edges. The degree of a vertex is the number of other vertices it is connected to by edges. A *loop* is an edge with both ends the same.

A directed graph or digraph $\mathcal{G}$ is a graph where $\mathcal{E}$ is a set of ordered pairs of vertices, called *directed* edges, arcs, or arrows. An edge $e = (v_i, v_j)$ is considered to be directed from $v_i$ to $v_j$; $v_j$ is called the head and $v_i$ is called the tail of the edge.

A *path* or *walk* $\Pi$ in a graph of length $m$ from vertex $u$ to $v$ is a sequence $e_1 e_2 \cdots e_m$ of $m$ edges such that the head of $e_m$ is $v$ and the tail of $e_1$ is $u$, and the head of $e_i$ is the tail of $e_{i+1}$, for $i = 1, ..., m-1$. The first vertex is called the start vertex and the last vertex is called the end vertex. Both of them are called end or terminal vertices of the walk. If also $u = v$, then we say that $\Pi$ is a *closed walk* based at $u$. A directed graph is *strongly connected* if for every pair of vertices $(v_i, v_j)$ there is a walk from $v_i$ to $v_j$.

The *adjacency matrix* of a finite directed graph $\mathcal{G}$ on $n$ vertices is the $n \times n$ matrix where the nondiagonal entry $a_{ij}$ is the number of edges from vertex $i$ to vertex $j$, and the diagonal entry $a_{ii}$ is the number of loops at vertex $i$

978-1-4244-4524-0/09/$25.00 ©2009 AACC

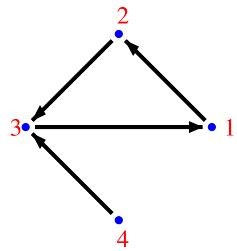


Fig. 1. An example of a graph $\mathcal{G}$.

(the number of loops at every node is defined to be one, unless another number is given on the graph). For instance, the adjacency matrix of the graph in Figure 1 is

$$A = \begin{pmatrix} 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 1 \end{pmatrix}.$$

A graph $\mathcal{G}$ with adjacency matrix $A$ is *isomorphic* to another graph $\mathcal{G}'$ with adjacency matrix $A'$ if there exists a permutation matrix $P$ such that

$$PAP^T = A'$$

The matrix $A$ is said to be *reducible* if there exists a permutation matrix $P$ such that

$$PAP^T = \begin{pmatrix} E & F \\ 0 & G \end{pmatrix} \tag{1}$$

where $E$ and $G$ are square matrices. If $A$ is not reducible, then it is said to be *irreducible*.

*Proposition 1:* A matrix $A \in \mathbb{Z}_2^{n\times n}$ is irreducible if and only if its corresponding graph $\mathcal{G}$ is strongly connected.

*Proof:* If $A$ is reducible, then from (1) we see that the vertices can be divided in two subsets; one subset belongs to the rows of $E$ and the other belongs to the rows of $G$. The latter subset is closed, because there is no walk from the second subset to the first one. Hence, the graph is not strongly connected.

Now, suppose that $\mathcal{G}$ is not strongly connected. Then there exists an isolated subset of vertices. Permute the vertices of $\mathcal{G}$ such that the vertices in the isolated subset comes last in the enumeration of $\mathcal{G}$. Then we see that the same permutation with $A$ gives a block triangular form as in (1). ■

*Proposition 2:* Consider an arbitrary finite graph $\mathcal{G}$ with adjacency matrix $A$. Then there is a permutation matrix $P$ and a positive integer $r$ such that

$$PAP^T = \begin{pmatrix} A_1 & A_{12} & \cdots & A_{1j} \\ 0 & A_2 & \cdots & A_{2j} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & A_r \end{pmatrix}, \tag{2}$$

where $A_1, ..., A_r$ are adjacency matrices of strongly connected graphs.

*Proof:* If $A$ is strongly connected, then it is irreducible according to Proposition 1, and there is nothing to do. Suppose now that $A$ is reducible. Then Proposition 1 gives that there is a permutation matrix $P_1$ such that $A_1 = P_1AP_1^T$ with

$$A_1 = \begin{pmatrix} E & F \\ 0 & G \end{pmatrix},$$

and $E$ and $G$ are square matrices. Now if $E$ and $G$ are irreducible, then we are done. Otherwise we repeat the same argument with $E$ and/or $G$. Since the graph is finite, we can only repeat this procedure a finite number of times, and hence there is some positive integer $r$ where this procedure stops. Then we arrive at a sequence of permutation matrices $P_1, ..., P_r$, such that $(P_r \cdots P_1)A(P_r \cdots P_1)^T$ has a block triangular structure given by (2) with $A_1, ..., A_r$ irreducible, and hence adjacency matrices of strongly connected graphs. Taking $P = P_r \cdots P_1$ completes the proof. ■

Now let $\omega : \mathcal{E} \to R$ be a *weight function* on $\mathcal{E}$ with values in some commutative ring $R$ (we can take $R = \mathbb{C}$ or a polynomial ring over $\mathbb{C}$). If $\Pi = e_1e_2 \cdots e_m$ is a walk, then the *weight* of $\Pi$ is defined by $w(\Pi) = w(e_1)w(e_2)\cdots w(e_m)$.

Let $\mathcal{G}$ be a finite directed graph, with $|\mathcal{V}| = p$. In this case, letting $i, j \in \{1, ..., p\}$ and $n \in \mathbb{N}$, define

$$A_{ij}(n) = \sum_{\Pi} \omega(\Pi),$$

where the sum is over all walks $\Pi$ in $\mathcal{G}$ of length $n$ from $v_i$ to $v_j$. In particular $A_{ij}(0) = \delta(i-j)$. Define a $p \times p$ matrix $A$ by

$$A_{ij} = \sum_{e} \omega(e),$$

where the sum is over all edges $e$ with $v_i$ and $v_j$ as the head and tail of $e$, respectively. In other words, $A_{ij} = A_{ij}(1)$. The matrix $A$ is called the adjacency matrix of $\mathcal{G}$, with respect to the weight function $\omega$.

The following proposition can be found in [9]:

*Proposition 3:* Let $n \in \mathbb{N}$. Then the $(i, j)$-entry of $A^n$ is equal to $A_{ij}(n)$.

*Proof:* This is immediate from the definition of matrix multiplication. Specifically, we have

$$[A^n]_{ij} = \sum A_{ii_1}A_{i_1i_2} \cdots A_{i_{n-1}j},$$

where the sum is over all sequences $(i_1, ..., i_{n-1}) \in \{1, ..., p\}^{n-1}$. The summand is 0 if there is no walk $e_1e_2 \cdots e_n$ from $v_i$ to $v_j$ with $v_{i_k}$ as the tail of $e_k$ $(1 < k \leq n)$ and $v_{i_{k-1}}$ as the head of $e_k$ $(1 \leq k < n)$. If such a walk exists, then the summand is equal to the sum of the weights of all such walks, and the proof follows. ■

*Corollary 1:* Let $\mathcal{G}$ be a graph with adjacency matrix $A \in \mathbb{Z}_2^{n\times n}$. Then there is a walk of length $k$ from node $v_i$ to node $v_j$ if and only if $[A^k]_{ij} \neq 0$. In particular, if $[A^{n-1}]_{ij} = 0$, then $[A^k]_{ij} = 0$ for all $k \in \mathbb{N}$.

A particularly elegant result for the matrices $A_{ij}(n)$ is that the generating function $g_{ij}(\lambda) = \sum_n A_{ij}(n)\lambda^n$ is

$$\begin{aligned} g_{ij}(\lambda) &= \sum_n A_{ij}(n)\lambda^n \\ &= \sum_n [A^n]_{ij}\lambda^n \end{aligned}$$

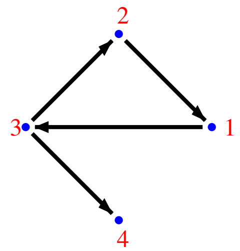


Fig. 2. The adjoint of the graph $\mathcal{G}$ in Figure 1.

Thus, we can see that the generating matrix function

$$G(\lambda) = (I - \lambda A)^{-1},$$

is such that $[G(\lambda)]_{ij} = g_{ij}(\lambda)$.

The *adjoint* graph of a finite directed graph $\mathcal{G}$ is denoted by $\mathcal{G}^*$, and it is the graph with the orientation of all arrows in $\mathcal{G}$ reversed. If the adjacency matrix of $\mathcal{G}$ with respect to the weight function $\omega$ is $A$ then the adjacency matrix of $\mathcal{G}^*$ is $A^*$.

*Example 1:* Consider the graph $\mathcal{G}$ in Figure 1. The adjacency matrix (in $\mathbb{Z}_2^{4\times 4}$) of this graph is

$$A = \begin{pmatrix} 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 1 \end{pmatrix}.$$

The adjoint graph $\mathcal{G}^*$ is given by Figure 2. It is easy to verify that the adjacency matrix (in $\mathbb{Z}_2^{4\times 4}$) of $\mathcal{G}^*$ is

$$A^* = \begin{pmatrix} 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 0 & 0 & 1 & 1 \end{pmatrix}.$$

To see if there is a walk of length 2 or 3 between any two nodes in $\mathcal{G}$, we calculate $A^2$ and $A^3$:

$$A^2 = \begin{pmatrix} 1 & 1 & 2 & 1 \\ 2 & 1 & 1 & 0 \\ 1 & 2 & 1 & 2 \\ 0 & 0 & 0 & 1 \end{pmatrix}, \quad A^3 = \begin{pmatrix} 2 & 3 & 3 & 3 \\ 3 & 2 & 3 & 1 \\ 3 & 3 & 3 & 3 \\ 0 & 0 & 0 & 1 \end{pmatrix}.$$

Using Corollary 1, we can see that there is no walk of length 2 from node 2 to node 4 since $[A^2]_{24} = 0$. On the other hand, there is a walk of length 3 since $[A^3]_{24} = 1 \neq 0$. There is also a walk of length 3 from node 2 to node 3, since we have assumed that every node has a loop. An example of such a walk is node 2 $\rightarrow$ node 1 $\rightarrow$ node 1 $\rightarrow$ node 3. Note also that since $[A^3]_{4i} = 0$ for $i = 1, 2, 3$, there is *no walk* that leads from 4 to any of the nodes 1, 2, or 3.

*Example 2:* Consider the matrix

$$A = \begin{pmatrix} A_{11} & 0 & A_{13} & 0 \\ A_{21} & A_{22} & 0 & 0 \\ 0 & A_{32} & A_{33} & A_{34} \\ 0 & 0 & 0 & A_{44} \end{pmatrix}. \tag{3}$$

The sparsity structure of $A$ can be represented by the graph given in Figure 1. Hence, if there is an edge from node $i$ to node $j$, then $A_{ij} \neq 0$. Mainly, the block structure of $A$ is the same as the adjacency matrix of the graph in Figure 1.

*Proposition 4:* Let $\mathcal{G}_1 = (\mathcal{V}_1, \mathcal{E}_1)$ and $\mathcal{G}_2 = (\mathcal{V}_2, \mathcal{E}_2)$ be two graphs with corresponding weighting functions $\omega_1 : \mathcal{E}_1 \rightarrow R$ and $\omega_2 : \mathcal{E}_2 \rightarrow R$, respectively. Also, let their adjacency matrices be represented by the generating functions $G_1$ and $G_2$, respectively. Define a graph $\mathcal{G}_3$ corresponding to $G_3 = G_1 G_2$. Then, if any two of the functions $G_1, G_2, G_3$ has the same sparsity structure, then $G_1, G_2, G_3$ all have the same structure (that is the graphs $\mathcal{G}_1$, $\mathcal{G}_2$, and $\mathcal{G}_3$ are identical).

*Proof:* The proof is simple by realizing that the graph $\mathcal{G}_3$ corresponding to $G_3$ is identical to $\mathcal{G}_1$ and $\mathcal{G}_2$, and has a weighting function $\omega_3 = \omega_1 + \omega_2$. ∎

### *A. Systems over Graphs*

Consider linear systems $\{G_i(\mathbf{q})\}$ with state space realization

$$\begin{aligned} x_i(k+1) &= \sum_{j=0}^{N} A_{ij} x_j(k) + B_i u_i(k) + w_i(k) \\ y_i(k) &= C_i x_i(k), \end{aligned} \tag{4}$$

for $i = 1, ..., N$. Here, $A_{ii} \in \mathbb{R}^{n_i \times n_i}$, $A_{ij} \in \mathbb{R}^{n_i \times n_j}$ for $j \neq i$, $B_i \in \mathbb{R}^{n_i \times m_i}$, and $C_i \in \mathbb{R}^{p_i \times n_i}$. Here, $w_i$ is the disturbance and $u_i$ is the control signal entering system $i$. The systems are interconnected as follows. If the state of system $j$ at time step $k$ ($x_j(k)$) affects the state of system $i$ at time step $k+1$ ($x_i(k+1)$), then $A_{ij} \neq 0$, otherwise $A_{ij} = 0$.

This block structure can be described by a graph $\mathcal{G}$ of order $N$, whose adjacency matrix is $A$, with respect to some weighting function $\omega$. The graph $\mathcal{G}$ has an arrow from node $j$ to $i$ if and only if $A_{ij} \neq 0$. The transfer function of the interconnected systems is given by $G(\mathbf{q}) = C(\mathbf{q}I - A)^{-1}B$. Then, the system $G^T(\mathbf{q})$ is equal to $B^T(\mathbf{q}I - A^T)^{-1}C^T$, and it can be represented by a graph $\mathcal{G}^*$ which is the adjoint of $\mathcal{G}$, since the adjacency matrix of $\mathcal{G}^*$ is $A^* = A^T$. The block diagram for the transposed interconnection is simply obtained by reversing the orientation of the interconnection arrows. This property was observed in [1].

*Example 3:* Consider four interconnected systems with the global system matrix $A$ given by

$$A = \begin{pmatrix} A_{11} & 0 & A_{13} & 0 \\ A_{21} & A_{22} & 0 & 0 \\ 0 & A_{32} & A_{33} & A_{34} \\ 0 & 0 & 0 & A_{44} \end{pmatrix}. \tag{5}$$

The interconnection can be represented by the graph $\mathcal{G}$ in Figure 1. The state of system 2 is affected directly by the state of system 1, and this is reflected in the graph by an arrow from node 1 to node 2. It is also reflected in the $A$ matrix, where $A_{21} \neq 0$. On the other hand, system 1 is not affected direclty by the state of system 2 since $A_{12} = 0$, and therefore there is no arrow from node 1 to node 2. The adjoint matrix $A^T$ is given by

$$A^T = \begin{pmatrix} A_{11}^T & A_{21}^T & 0 & 0 \\ 0 & A_{22}^T & A_{32}^T & 0 \\ A_{13}^T & 0 & A_{33}^T & 0 \\ 0 & 0 & A_{34}^T & A_{44}^T \end{pmatrix}. \qquad (6)$$

The interconnection structure for the transposed system can be described by the adjoint of $\mathcal{G}$ in Figure 2.

### B. Information Pattern

The information pattern that will be considered in this paper is the partially nested, which was first discussed in [5] and that we will briefly describe in this section. A more modern treatment can be found in [4], which our presentation builds on.

Consider the interconnected systems given by equation (4). Now introduce the *information set*

$$\mathbb{I}^k = \{\{y_1(t), ..., y_N(t)\}_{t=0}^k, \{\mu_1^t, ..., \mu_N^t\}_{t=0}^k\},$$

the set of all input and output signals up to time step $k$. Let the control action of system $i$ at time step $k$ be a decision function $u_i(k) = \mu_i^k(\mathbb{I}_i^k)$, where $\mathbb{I}_i^k \subset \mathbb{I}_k$. We assume that every system $i$ has access to its own output, that is $y_i(k) \in \mathbb{I}_i^k$. Write the state $x(k)$ as

$$x(k) = \sum_{t=1}^{k} A^{k-t}(Bu(t-1) + w(t-1)),$$

For $t < k$, the decision $\mu_j^t(\mathbb{I}_j^t) = u_j(t)$ will affect the output $y_i(k) = [Cx(k)]_i$ if $[CA^{k-t-1}B]_{ij} \neq 0$. Thus, if $\mathbb{I}_j^t \not\subset \mathbb{I}_i^k$, decision maker $j$ has at time $t$ an incentive to encode information about the elements that are not available to decision maker $i$ at time $k$, a so called *signaling incentive* (see [5], [4] for further reading).

*Definition 1:* We say that the information structure $\mathbb{I}_i^k$ is *partially nested* if $\mathbb{I}_j^t \subset \mathbb{I}_i^k$ for $[CA^{k-t-1}B]_{ij} \neq 0$, for all $t < k$.

It's not hard to see that the information structure can be defined recursively, where the recursion is reflected by the interconnection graph (which is in turn reflected by the structure of the system matrix $A$). Let $\mathcal{G}$ represent the interconnection graph. Denote by $\mathcal{J}_i$ the set of indexes that includes $i$ and its neighbours $j$ in $\mathcal{G}$. Then, $\mathbb{I}_i^k = y_i(k) \cup_{j \in \mathcal{J}_i} \mathbb{I}_j^{k-1}$. In words, the information available to the decision function $\mu_i^k$ is what its neighbours know from one step back in time, what its neighbours' neighbours know from two steps back in time, etc...

## IV. Problem Description

### A. Distributed State Feedback Control

The problem we're considering here is to find the optimal distributed state feedback control law $u_i(k) = \mu_i^k(\mathbb{I}_i^k)$, for $i = 1, ..., N$ that minimizes the quadratic cost

$$J_M(x, u) := \sum_{k=1}^{M} \mathbf{E}\|Cx(k) + Du(k)\|^2,$$

with respect to the system dynamics

$$\begin{aligned} x_i(k+1) &= \sum_{j=1}^{N} A_{ij}x_j(k) + B_iu_i(k) + w_i(k) \\ y_i(k) &= x_i(k), \end{aligned} \qquad (7)$$

$w(k) \sim \mathcal{N}(0, I)$, and $x(k) = 0$ for all $k \leq 0$. Without loss of generality, we assume that $B_i$ has full column rank, for $i = 1, ..., N$ (and hence has a left inverse). In [5], it has been shown that if $\{\mathbb{I}_i^k\}$ are described by a partially nested information pattern, then every decision function $\mu_i^k$ can be taken to be a linear map of the elements of its information set $\mathbb{I}_i^k$. Hence, the controllers will be assumed to be linear:

$$u_i(k) = [K_t(k, \mathbf{q})]_i x(k) = \sum_{t=0}^{\infty} [K_t(k)]_i \mathbf{q}^{-t} x(k). \qquad (8)$$

The information pattern is reflected in the parameters $K_t(k)$, where $[K_t(k)]_{ij} = 0$ if $[A^t]_{ij} = 0$. Thus, the sparsity structure of the coefficients of $K(k, \mathbf{q})$ is the same as that of $(I - \mathbf{q}^{-1}A)^{-1} = I + A\mathbf{q}^{-1} + A^2\mathbf{q}^{-2} + \cdots$. Hence, our objective is to minimize the quadratic cost $J(x, u) \to$ min, subject to (7) and sparsity constraints on $K_t(k)$ that are reflected by the dynamic interconnection matrix $A$. In particular, we are interested in finding a solution where every $K_t(k)$ is "as sparse as it can get" without loosing optimality. To summarize, the problem we are considering is:

$$\begin{aligned} &\inf_{u} \quad \sum_{i=1}^{M} \sum_{k=1}^{N} \mathbf{E}\|z_i(k)\|^2 \\ &\text{subject to} \quad x_i(k+1) = \sum_{j=1}^{N} A_{ij}x_j(k) + B_iu_i(k) + w_i(k) \\ &\qquad z(k) = Cx(k) + Du(k) \\ &\qquad u_i(k) = \sum_{t=0}^{\infty} [K_t(k)]_i \mathbf{q}^{-t} x(k) \\ &\qquad [K_t(k)]_{ij} = 0 \ \text{ if } \{[A^{k-t-1}B]_{ij} = 0, t < k \\ &\qquad \text{or } t = k, i \neq j\} \\ &\qquad w(k) = x(k) = x(0) = 0 \ \text{ for all } k < 0 \\ &\qquad w(k) \sim \mathcal{N}(0, I) \ \text{ for all } k \geq 0 \\ &\qquad \text{for } i = 1, ..., N \end{aligned} \qquad (9)$$

### B. Distributed State Estimation

Consider $N$ systems given by

$$\begin{aligned} x(k+1) &= Ax(k) + Bw(k) \\ y_i(k) &= C_ix_i(k) + D_iw(k), \end{aligned} \qquad (10)$$

for $i = 1, ..., N$, $w(k) \sim \mathcal{N}(0, I)$, and $x(k) = 0$ for all $k \leq 0$. Without loss of generality, we assume that $C_i$ has full row rank, for $i = 1, ..., N$. The problem is to find optimal distributed estimators $\mu_i^k$ in the following sense:

$$\inf_{\mu} \quad \sum_{i=1}^{M} \sum_{k=1}^{N} \mathbf{E}\|x_i(k) - \mu_i^k(\mathbb{I}_i^{k-1})\|^2 \qquad (11)$$

In a similar way to the distributed state feedback problem, the information pattern is the partially nested, which is reflected by the interconnection of the graph. The linear decisions are optimal, hence we can assume that

$$\begin{aligned}\hat{x}_i(k) := \mu_i^k(\mathbb{I}_i^{k-1}) &= [L_t(k,\mathbf{q})]_i y_i(k-1)\\ &= \sum_{t=0}^{\infty}[L_t(k)]_i \mathbf{q}^{-t} y_i(k-1).\end{aligned} \tag{12}$$

Then, our problem becomes

$$\begin{aligned}&\inf_{\hat{x}} \quad \sum_{i=1}^{M}\sum_{k=1}^{N}\mathbf{E}\|x_i(k)-\hat{x}_i(k)\|^2\\ &\text{subject to} \quad x(k+1) = Ax(k)+Bw(k)\\ &\qquad y_i(k) = C_i x_i(k) + D_i w(k)\\ &\qquad \hat{x}_i(k) = \sum_{t=0}^{\infty}[L_t(k)]_i \mathbf{q}^{-t} y_i(k-1)\\ &\qquad [L_t(k)]_{ij} = 0 \text{ if } \{[CA^{k-t-1}]_{ij} = 0, t<k\\ &\qquad \text{or } t=k, i\neq j\}\\ &\qquad w(k) = x(k) = x(0) = 0 \text{ for all } k<0\\ &\qquad w(k) \sim \mathcal{N}(0,I) \text{ for all } k\geq 0\\ &\qquad \text{for } i=1,...,N\end{aligned} \tag{13}$$

## V. Duality of Estimation and Control under Sparsity Constraints

In [4], it was shown that the distributed state feedback control problem is the dual to the distributed estimation problem under the partially nested information pattern. It showed exactly how to *transform* a state feedback problem to an estimation problem. Note that since we are considering linear controllers, we can equivalently consider linear controllers of the disturbance rather than the state. It can be done since

$$\begin{aligned}x(k) &= (\mathbf{q}I - A - BK(k,\mathbf{q}))^{-1}w(k)\\ &= (I - A\mathbf{q}^{-1} - BK(k,\mathbf{q})\mathbf{q}^{-1})^{-1}\mathbf{q}^{-1}w(k),\end{aligned} \tag{14}$$

which clearly yields

$$w(k-1) = (I - A\mathbf{q}^{-1} - BK(k,\mathbf{q})\mathbf{q}^{-1})x(k). \tag{15}$$

Thus, we will set

$$u(k) = -R(k,\mathbf{q})w(k-1) = -\sum_{t=0}^{\infty}R_t(k)\mathbf{q}^{-t}w(k-1).$$

The structure of $R(k,\mathbf{q})$ is inherited from that of $(I-\mathbf{q}^{-1}A)^{-1}$. To see this, note that $R(k,\mathbf{q}) = K(k,\mathbf{q})(I - A\mathbf{q}^{-1} - BK(k,\mathbf{q}))^{-1}$. $(I - A\mathbf{q}^{-1})$ clearly has the same sparsity structure as that of $(I-\mathbf{q}^{-1}A)^{-1}$. $K(k,\mathbf{q})$ does also have the same sparsity structure as $(I-\mathbf{q}^{-1}A)^{-1}$ (by definition of the information constraints on the controller), and since $B$ is block diagonal, so does $BK(k,\mathbf{q})$. Applying Proposition 4 twice, shows that $(I - A\mathbf{q}^{-1} - BK(k,\mathbf{q}))^{-1}$ and $K(k,\mathbf{q})(I - A\mathbf{q}^{-1} - BK(k,\mathbf{q}))^{-1}$ has the same sparsity structure as $(I-\mathbf{q}^{-1}A)^{-1}$.

Now each term in the quadratic cost of (9) is given by

$$\begin{aligned}&\mathbf{E}\|Cx(k)+Du(k)\|^2 =\\ &\mathbf{E}\|C(\mathbf{q}I-A)^{-1}w(k) -\\ &[C(\mathbf{q}I-A)^{-1}B + D]R(k,\mathbf{q})\mathbf{q}^{-1}w(k)\|^2 =\\ &\mathbf{E}\|(\mathbf{q}I-A^T)^{-1}C^T w(k) -\\ &R^T(k,\mathbf{q})\mathbf{q}^{-1}[B^T(\mathbf{q}I-A^T)^{-1}C^T + D^T]w(k)\|^2\end{aligned} \tag{16}$$

where the last equality is obtained from transposing which doesn't change the value of the norm. Introduce the state space equations

$$\begin{aligned}\bar{x}(k+1) &= A^T\bar{x}(k) + C^T w(k)\\ y(k) &= B^T\bar{x}(k) + D^T w(k)\end{aligned} \tag{17}$$

and let

$$\hat{x}(k) = R^T(k,\mathbf{q})y(k-1).$$

Then comparing with (16) we see that

$$\begin{aligned}\mathbf{E}\|Cx(k)+Du(k)\|^2 &= \mathbf{E}\|\bar{x}(k)-\hat{x}(k)\|^2\\ &= \sum_{i=1}^{N}\mathbf{E}\|\bar{x}_i(k)-\hat{x}_i(k)\|^2.\end{aligned} \tag{18}$$

Hence, we have transformed the control problem to an estimation problem, where the parameters of the estimation problem are the transposed parameters of the control problem:

$$\begin{aligned}A &\leftrightarrow A^T\\ B &\leftrightarrow C^T\\ C &\leftrightarrow B^T\\ D &\leftrightarrow D^T\end{aligned} \tag{19}$$

The solution of the control problem described as a feedforward problem, $R(k,\mathbf{q})$, is equal to $L^T(k,\mathbf{q})$, where $L(k,\mathbf{q}) = R^T(k,\mathbf{q})$ is the solution of the corresponding dual estimation problem. The information constraints on $L(k,\mathbf{q})$ follow easily from that of $R(k,\mathbf{q})$ by transposing (that is, looking at the adjacency matrix of the dual graph). We can now state

*Theorem 1:* Consider the distributed state feedback linear quadratic problem (9), with state space realization

$$\left(\begin{array}{c|cc} A & I & B\\ \hline C & 0 & D\\ I & 0 & 0\end{array}\right)$$

and solution $u(k) = \sum_{t=0}^{\infty}R_t(k)\mathbf{q}^{-t}w(k-1)$, and the distributed estimation problem (13) with state space realization

$$\left(\begin{array}{c|cc} A^T & C^T & 0\\ \hline I & 0 & 0\\ B^T & D^T & 0\end{array}\right)$$

and solution $\hat{x}(k) = \sum_{t=0}^{\infty}L_t(k)\mathbf{q}^{-t}y(k-1)$. Then

$$R_t(k) = L_t^T(k).$$

## VI. CONCLUSION

We show that distributed estimation and control problems are dual under partially nested information pattern using a novel system theoretic formulation of dynamics over graphs.

Future work will be examining controller order reduction of the optimal controllers without loosing optimality by exploiting the structure of the problem.

## REFERENCES


[1] B. Bernhardsson. *Topics in Digital and Robust Control of Linear Systems*. PhD thesis, Lund University, 1992.

[2] R. Diestel. *Graph Theory*. Springer-Verlag, 2005.

[3] A. Gattami. Generalized linear quadratic control theory. In *45th IEEE Conference on Decision and Control*, 2006.

[4] A. Gattami. *Optimal Decisions with Limited Information*. PhD thesis, Lund University, 2007.

[5] Y.-C. Ho and K.-C. Chu. Team decision theory and information structures in optimal control problems-part i. *IEEE Trans. on Automatic Control*, 17(1), 1972.

[6] J. Marschak. Elements for a theory of teams. *Management Sci.*, 1:127–137, 1955.

[7] R. Radner. Team decision problems. *Ann. Math. Statist.*, 33(3):857–881, 1962.

[8] A. Rantzer. Linear quadratic team theory revisited. In *ACC*, 2006.

[9] Richard Stanley. *Enumerative Combinatorics, Volume I*. Cambridge University Press, 1997.

[10] H. S. Witsenhausen. A counterexample in stochastic optimum control. *SIAM Journal on Control*, 6(1):138–147, 1968.