Control and estimation problems under partially nested information pattern

The MIT Faculty has made this article openly available. Please share how this access benefits you. Your story matters.

<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>As Published</td>
<td><a href="http://dx.doi.org/10.1109/CDC.2009.5400652">http://dx.doi.org/10.1109/CDC.2009.5400652</a></td>
</tr>
<tr>
<td>Publisher</td>
<td>Institute of Electrical and Electronics Engineers</td>
</tr>
<tr>
<td>Version</td>
<td>Final published version</td>
</tr>
<tr>
<td>Citable Link</td>
<td><a href="http://hdl.handle.net/1721.1/59325">http://hdl.handle.net/1721.1/59325</a></td>
</tr>
<tr>
<td>Terms of Use</td>
<td>Article is made available in accordance with the publisher’s policy and may be subject to US copyright law. Please refer to the publisher’s site for terms of use.</td>
</tr>
<tr>
<td>Detailed Terms</td>
<td></td>
</tr>
</tbody>
</table>
Control and Estimation Problems under Partially Nested Information Pattern

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 [12]). Important progress was made for the stochastic static team decision problems in [7] and [8]. New information structures were explored in [6] for the stochastic linear quadratic finite horizon control problem. Team problems where revisited [1] and [10], where it was shown that if information propagation was fast enough, the dynamic team problem can be recast as a convex optimization problem using Youla parametrization. In [9], 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 [4].

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 \( \mathbb{E}(x) = m \) and \( \mathbb{E}((x - m)(x - m)^T) = X \).

\( M_i \), 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 \( q \). That is \( x_{k+1} = q x_k \), where \( \{ x_k \} \) is a given process. A causal linear time-invariant operator \( T(q) \) is given by its generating function

\[
T(q) = \sum_{t=0}^{\infty} T(t) q^{-t}, \quad T_t \in \mathbb{R}^{p \times m}.
\]

A transfer matrix in terms of state-space data is denoted

\[
\begin{pmatrix} A & B \\ C & D \end{pmatrix} := C(qI - A)^{-1} B + D.
\]

B. 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 [3]). A (simple) graph \( G \) is an ordered pair \( G := (V, E) \) where \( V \) is a set, whose elements are called vertices or nodes, \( E \) is a set of pairs (unordered) of distinct vertices, called edges or lines. The set \( V \) (and hence \( E \)) is taken to be finite in this paper. The order of a graph is \( |V| \) (the number of vertices). A graph’s size is \( |E| \), the number of edges. A loop is an edge which starts and ends with the same node.

A directed graph or digraph \( G \) is a graph where \( E \) 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.

The adjacency matrix of a finite directed graph \( G \) 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 \( i \) (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
\[
A = \begin{pmatrix}
1 & 0 & 1 & 0 \\
1 & 1 & 0 & 0 \\
0 & 1 & 1 & 1 \\
0 & 0 & 0 & 1 \\
\end{pmatrix}.
\]

For a graph with adjacency matrix \( A \), we write the generating function
\[
G(\lambda) = (I - \lambda A)^{-1} = \sum_{t \geq 0} A^t \lambda^t.
\]

With the sparsity structure of a given generating function \( G(\lambda) = \sum_{t \geq 0} G(t) \lambda^t \), we mean the sparsity structure of \( G(t) \), \( t \geq 0 \).

**Proposition 1:** Suppose that \( G_1(q) \) and \( G_2(q) \) have a sparsity structure given by \( (I - q^{-1}A)^{-1} \), \( A \in \mathbb{Z}_{+}^{n \times n} \), respectively. Then the sparsity structure of \( G_1(q)G_2(q) \) is given by \( (I - q^{-1}A)^{-1} \).

**Proof:** First note that for any matrix \( A_1 \) with sparsity structure given by \( A \), we have that the sparsity structure of \( A_1 q^{-1} G_1(q) \) is the same as that of \( (I - q^{-1}A)^{-1} \). By considering the formal series of the generating functions of \( G_1(q) \) and \( G_2(q) \), we get that the sparsity structure of \( G_1(q)G_2(q) \) is given by \( (I - q^{-1}A)^{-1} \).

The **adjoint** graph of a finite directed graph \( G \) is denoted by \( G^* \), and it is the graph with the orientation of all arrows in \( G \) reversed. If the adjacency matrix of \( G \) is \( A \) then the adjacency matrix of \( G^* \) is \( A^* \).

**Example 1:** Consider the graph \( G \) in Figure 1. The adjacency matrix (in \( \mathbb{Z}_{+}^{4 \times 4} \)) 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 \( G^* \) is given by Figure 2. It is easy to verify that the adjacency matrix (in \( \mathbb{Z}_{+}^{4 \times 4} \)) of \( G^* \) is
\[
A^* = \begin{pmatrix}
1 & 1 & 0 & 0 \\
0 & 1 & 1 & 0 \\
1 & 0 & 1 & 0 \\
0 & 0 & 1 & 1 \\
\end{pmatrix}.
\]

**Example 2:** Consider the matrix
\[
A = \begin{pmatrix}
A_{11} & 0 & A_{13} & 0 \\
A_{21} & A_{22} & 0 & 0 \\
0 & A_{32} & A_{33} & A_{34} \\
0 & 0 & 0 & A_{44} \\
\end{pmatrix}.
\]

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.

**C. Systems over Graphs**

Consider linear systems \( \{P_i(q)\} \) with state space realization
\[
x_i(k+1) = \sum_{j=0}^{N} A_{ij} x_j(k) + B_i u_i(k) + w_i(k)
\]
\[
y_i(k) = C_i x_i(k),
\]
for \( i = 1, ..., N \). Here, \( A_{ij} \in \mathbb{R}^{n_i \times n_i} \), \( A_{ij} \in \mathbb{R}^{n_j \times n_j} \) 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, \( u_i \) is the disturbance and \( w_i \) 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 \( G \) of order \( N \), whose adjacency matrix is \( A \). The graph \( G \) 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 \( P(q) = C(qI - A)^{-1}B \). Then, the system \( P^T(q) \) is equal to \( B^T(qI - A^T)^{-1}C^T \), and it can be represented by a graph \( G^* \), which is the adjoint of \( G \), since the adjacency matrix of \( G^* \) 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 [2].

**Example 3:** Consider four interconnected systems with the global system matrix \( A \) given by
\[
A = \begin{pmatrix}
A_{11} & 0 & A_{13} & 0 \\
A_{21} & A_{22} & 0 & 0 \\
0 & A_{32} & A_{33} & A_{34} \\
0 & 0 & 0 & A_{44} \\
\end{pmatrix}.
\]

The interconnection can be represented by the graph \( G \) 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 directly 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
\[
A^T = \begin{pmatrix}
A_{11}^T & A_{13}^T & 0 & 0 \\
0 & A_{22}^T & A_{32}^T & 0 \\
A_{13}^T & 0 & A_{33}^T & 0 \\
0 & 0 & A_{44}^T & A_{44}^T \\
\end{pmatrix}.
\]

The interconnection structure for the transposed system can be described by the adjoint of \( G \) in Figure 2.
D. Information Pattern

The information pattern that will be considered in this paper is the partially nested, which was first discussed in [6] and that we will briefly describe in this section. A more modern treatment can be found in [5], which our presentation builds on.

Consider the interconnected systems given by equation (2). Now introduce the information set

$$\mathbb{I}^k = \{u_{[0,k]}, y_{[0,k]}, w_{[0,k]}\},$$

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 \(u_i(k) = \mu^k_{i}(I^k_{i})\), where \(I^k_{i} \subseteq \mathbb{I}^k\). We assume that every system \(i\) has access to its own output, that is \(y_i(k) \in \mathbb{I}^k_i\). 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 \(\mu^t_{i}(I^t_{i}) = u_i(t)\) will affect the output \(y_i(t) = [C x(t)]_i\) if \([CA^{k-t-1}B]_{ij} \neq 0\). Thus, if \(I^t_{i} \not\subseteq I^{k-1}_{i}\), 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 [6], [5] for further reading).

Definition 1: We say that the information structure \(I^k_i\) is partially nested if \(I^t_{i} \subset I^{k-1}_{i}\) 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 \(J_i\) the set of indexes that includes \(i\) and its neighbours \(j\) in \(\mathcal{G}\). Then, \(I^k_{i} = y_i(k) \cup_{j \in \mathcal{J}_i} I^{k-1}_{j}\). In words, the information available to the decision function \(\mu^k_{i}\) is what its neighbours know from one step back in time, what its neighbours’ neighbours know from two steps back in time, etc...

III. PROBLEM DESCRIPTION

A. Distributed State Feedback Control

The problem we are considering here is to find the optimal distributed state feedback control law \(u_i(k) = \mu^k_{i}(I^k_{i})\), for \(i = 1, ..., N\) that minimizes the quadratic cost

$$J(x, u) := \lim_{M \to \infty} \frac{1}{M} \sum_{k=1}^{M} \mathbb{E}[\|Cx(k) + Du(k)\|^2],$$

with respect to the system dynamics

$$x_i(k + 1) = \sum_{j=1}^{N} A_{ij} x_j(k) + B_i u_i(k) + w_i(k)$$

$$y_i(k) = x_i(k),$$

\(w(k) \sim \mathcal{N}(0, I)\) for all \(k\), 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 [6], it has been shown that if \(\{I^k_{i}\}\) are described by a partially nested information pattern, then every decision function \(\mu^k_{i}\) can be taken to be a linear map of the elements of its information set \(I^k_{i}\). Hence, the controllers will be assumed to be linear:

$$u_i(k) = [K(q)]_i x(k) = \sum_{t=0}^{\infty} [K(t)]_i q^{-t} x(k).$$

The information pattern is reflected in the parameters \(K(q)\), where \([K(t)]_i = 0\) if \([A^t]_{ij} = 0\), and \(A \in \mathbb{R}^{N \times N}\) is the adjacency matrix of the interconnection graph. Thus, the block sparsity structure of \(K(q)\) is the same as the sparsity structure of \((I - q^{-1}A)^{-1} = I + Aq^{-1} + A^2q^{-2} + \cdots\). Hence, our objective is to minimize the quadratic cost, \(J(x, u) \to \min\), subject to (5) and sparsity constraints on \(K(i)\) that are reflected by the dynamic interconnection matrix \(A\). In particular, we are interested in finding a solution where every \(K(k)\) is “as sparse as it can get” without losing optimality. To summarize, the problem we are considering is:

$$\inf_{\mu} \lim_{M \to \infty} \frac{1}{M} \sum_{k=1}^{M} \sum_{i=1}^{N} \mathbb{E}[\|z_i(k)\|^2]$$

subject to

$$x(k + 1) = Ax(k) + Bu(k) + w(k)$$

$$z(k) = Cx(k) + Du(k)$$

$$u_i(k) = \sum_{t=0}^{\infty} [K(t)]_i q^{-t} x(k)$$

$$[K(t)]_ij = 0 \text{ if } ([A^t]_{ij} = 0, t < k \text{ or } t = k, i \neq j)$$

$$w(k) = x(k) = x(0) = 0 \text{ for all } k < 0$$

$$w(k) \sim \mathcal{N}(0, I) \text{ for all } k \geq 0$$

for \(i = 1, ..., N\).

B. Distributed State Estimation

Consider \(N\) systems given by

$$x_i(k + 1) = Ax_i(k) + Bu(k) + w_i(k)$$

$$y_i(k) = C_i x_i(k) + D_i w_i(k),$$

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 \(\mu^k_i\) in the following sense:

$$\inf_{\mu} \lim_{M \to \infty} \frac{1}{M} \sum_{k=1}^{M} \sum_{i=1}^{N} \mathbb{E}[\|x_i(k) - \mu^k_i(I^k_{i})\|^2]$$

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

$$\hat{x}_i(k) := \mu^k_i(I^k_{i}) = [L(q)]_i y(k - 1)$$

$$= \sum_{t=0}^{\infty} [L(t)]_i q^{-t} y(k - 1).$$
Then, our problem becomes
\[
\begin{align*}
\inf_{\hat{x}} & \lim_{M \to \infty} \frac{1}{M} \sum_{k=1}^{M} \sum_{i=1}^{N} \mathbb{E}[\|x_i(k) - \hat{x}_i(k)\|^2] \\
\text{subject to} & \ x(k+1) = Ax(k) + Bu(k) \\
& \ y(k) = Cx(k) + Du(k) \\
& \ \hat{x}_i(k) = \sum_{j=0}^{\infty} [L(t)]_{ij} q^{-t} y(k-1) \\
& \ [L(t)]_{ij} = 0 \text{ if } \{|A^{k-1}\}_{ij} = 0, t < k \text{ or } t = k, i \neq j \}
\end{align*}
\]

which clearly yields
\[
w(k-1) = (I - Aq^{-1} - BK(q)q^{-1})x(k).
\]

Thus, we will set
\[
u(k) = -G(q)w(k-1) = -\sum_{t=0}^{\infty} G(t)q^{-t}w(k-1).
\]

The structure of $G(q)$ is inherited from that of $(I - q^{-1}A)^{-1}$. To see this, note that $u_i(k-1) = x_i(k) - \sum_{j=1}^{N} A_{ij}x_j(k-1) - B_i y_i(k)$. The disturbance $w_i(k-1)$ is available to decision maker $i$. It can be generated from $\mathbb{I}_i^k$ since $\{x_i(k)\}, \mathbb{I}_i^{k-1} \subseteq \mathbb{I}_i^k$, and $x_j(k-1) \in \mathbb{I}_i^k$ if $A_{ij} \neq 0$, for $j = 1, ..., N$, and $u_i(k-1) = \mu_i(\mathbb{I}_i^{k-1})$. Since the information sets $\mathbb{I}_i^k$ are partially nested, the information about $w_i(k-1)$ will also be partially nested. It can also be shown algebraically. We have that $u(k) = K(q)x(k) = K(q)(I - Aq^{-1} - BK(q)q^{-1})w(k-1)$, and thus $G(q) = K(q)(I - Aq^{-1} - BK(q)q^{-1})^{-1}$. Applying Proposition II-B to the formal power series of $G(q)$,
\[
G(q) = \sum_{t \geq 0} K(q)(A + BK(q))^{t} q^{-t}
\]
we see that $G(q)$ has the same sparsity structure as $(1 - Aq^{-1})^{-1}$.

Now each term in the quadratic cost of (7) is given by
\[
\begin{align*}
\mathbb{E}[Cx(k) + Du(k)]^2 &= 0 \\
\mathbb{E}[C(qI - A)^{-1}w(k)-] &= \|C(qI - A)^{-1} - [C(qI - A)^{-1}B + D]q^{-1}\|_2 \\
& = \|[qI - A^{-1}]^{-1}B - G^T(q)q^{-1}[B^T(qI - A^{-1})C^T + D^T]\|_2^2
\end{align*}
\]

where the third equality is obtained from transposing which doesn’t change the value of the norm, and the last is obtained by introducing $\{\bar{w}(k)\}$ as an uncorrelated Gaussian process. Introduce the state space equation
\[
\begin{align*}
\bar{x}(k+1) &= A^T \bar{x}(k) + C^T \bar{w}(k) \\
y(k) &= B^T \bar{x}(k) + D^T \bar{w}(k)
\end{align*}
\]

and let
\[
\hat{x}(k) = G^T(q)y(k-1).
\]

Then comparing with (14) we see that
\[
\mathbb{E}[Cx(k) + Du(k)]^2 = \mathbb{E}[\bar{x}(k) - \hat{x}(k)]^2
\]

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:
\[
A \leftrightarrow A^T \\
B \leftrightarrow C^T \\
C \leftrightarrow B^T \\
D \leftrightarrow D^T
\]

Consider the following control (feedforward) problem:
\[
\begin{align*}
\inf_u & \lim_{M \to \infty} \frac{1}{M} \sum_{k=1}^{M} \sum_{i=1}^{N} \mathbb{E}[z_i(k)]^2 \\
\text{subject to} & \ x(k+1) = Ax(k) + Bu(k) + w(k) \\
& \ z(k) = Cx(k) + Du(k) \\
& \ u_i(k) = \sum_{t=0}^{\infty} [G(t)]_{ij} q^{-t} w(k-1) \\
& \ [G(t)]_{ij} = 0 \text{ if } \{|A^{k-1}\}_{ij} = 0, t < k \text{ or } t = k, i \neq j \}
\end{align*}
\]

The solution of the control problem described as a feedforward problem, $G(q)$, is equal to $L^T(q)$, where $L(q)$ is the solution of the corresponding dual estimation problem. The information constraints on $L(q)$ follow easily from that of
\( G(q) \) by transposing (that is, looking at the adjacency matrix of the dual graph, see Section II-C). We can now state

**Theorem 1:** Consider the distributed feedforward linear quadratic problem (7), with state space realization

\[
\begin{bmatrix}
A & I & B \\
C & 0 & D \\
0 & -I & 0
\end{bmatrix}
\]

and solution \( u(k) = \sum_{t=0}^{\infty} G(t)q^{-t}w(k-1) \), and the distributed estimation problem (11) with state space realization

\[
\begin{bmatrix}
A^T & C^T & 0 \\
I & 0 & -I \\
B^T & D^T & 0
\end{bmatrix}
\]

and solution \( \hat{x}(k) = \sum_{t=0}^{\infty} L(t)q^{-t}y(k-1) \). Then \( G(t) = L^T(t) \).

V. 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.

VI. ACKNOWLEDGEMENT

The author would like to thank the reviewers for their valuable comments.

REFERENCES


