Solving Least Squares Problems on Partially Ordered Sets

Alexandre Amice and Pablo A. Parrilo

***Abstract*— We study a general class of least-squares problems structured according to a partially ordered set (poset). This is a fundamental optimization problem underlying the design of structured controllers on directed acyclic graphs or posets. We show that the optimality conditions of this problem yield a structured linear system, with sparsity pattern determined by a derived poset known as the poset of intervals. In general, this system could be relatively dense, and thus standard sparse linear algebra techniques may fail to provide significant reduction in computational complexity. Nonetheless, for a broad class of posets called multitrees identified in [1] we show that performing elimination according to an order defined by the poset intervals progressively decouples variables, reducing the arithmetic complexity of solving the problem.**

I. Introduction

The control of large-scale, networked dynamical systems via decentralized controllers is a central challenge in modern engineering. Due to known hardness results concerning the design of optimal, decentralized controllers for even seemingly simple systems [2], identifying information structures that enable efficient computation is an important current research direction.

Information structures modelled by directed acyclic graphs (DAGs) or, equivalently, partially ordered sets (posets) have been shown to provide a rich class of decentralized systems amenable to control design. Controllers compatible with an information constraint defined by a poset are known as poset causal controllers [3] and prior work has shown that in the case of linear dynamics with quadratic costs, the optimal controller is linear [4]. Moreover, Youla domain techniques relying on quadratic invariance [5] or Systems Level Synthesis [6] enable posing the search for linear, poset causal controllers as a structurally constrained least squares program of the form

$$\min_{X \in \mathcal{I}(\mathcal{P})} \|A - CXB\|^2 \tag{1}$$

where A, B, C, X are all constrained to be blockwise in the incidence algebra of a given poset $\mathcal{P}$ (defined formally in Section II-A).

Laboratory for Information and Decision Systems, Massachusetts Institute of Technology. Email: {amice, parrilo}@mit.edu

The authors gratefully acknowledge funding from Air Force Research Lab Award FA8750-19-2-1000

For linear dynamical systems, it is possible to solve (or approximate) (1) using techniques such as vectorization [7] and finite truncations of infinite dimensional operators. However, such methods may not be computationally efficient and provide no insight into the structure of the optimal controller. These drawbacks emphasize the need for solution techniques in the vein of [3], [8], [9] which exploit the order-theoretic structure to provide direct, efficient, and interpretable solutions to (1).

Thus far, this line of work has succeeded in describing solution procedures for control problems of the form (1) for arbitrary posets, in the state feedback case [3]. Results for the output feedback case are limited to totally ordered sets (chain graphs) and two-layer broadcast architectures [8], [9], [10]. Additionally, it has been shown that the optimal solution in these settings demonstrates intuitive, recursive structures intimately related to combinatorial aspects of the underlying information graph [1], [11]. Nonetheless, a general solution technique for the output feedback case remains unknown. This paper seeks to make some progress towards this goal.

At its heart, efficiently solving (1) is a problem in structured linear algebra. Therefore, in this work, we investigate the combinatorial and algebraic structure underlying the poset-structured least-squares problems.

Our main contributions are:

- A characterization of the optimality conditions as a set of sparse linear equations, with sparsity pattern related to the *poset of intervals* of $\mathcal{P}$ (Theorem 1).
- An identification of a key separability property among the variables of the linear system of Theorem 1 which increases the sparsity after ones step of elimination is performed (Theorem 2).
- An explicit, sequential elimination strategy for solving the linear system of Theorem 1. When $\mathcal{P}$ is a multitree (defined in Section IV-C) this elimination strategy leads to progressive sparsification of the linear system (Theorem 3), reducing the number of operations required to solve the system.

The rest of the paper is organized as follows. In Section II, we introduce the necessary preliminaries about posets and recall some standard notions from the sparse linear algebra literature. In Section III, we formally introduce the problem. We describe our main results in Section IV.

A. Notation

Throughout we will use calligraphic letters $\mathcal{X}$, capital roman letters X, and lowercase roman letters x to denote sets, matrices, and vectors respectively. Operators will be denoted in boldface font $\boldsymbol{X}$. When it is clear from context, scalar values can be written either upper or lowercase. When X is a block matrix, we use subscript X_{ij} to denote the i,j^{th} block. Finally, we define $\mathcal{T}^c$ as the complement of a set $\mathcal{T}$ in a base set $\mathcal{S}$.

II. Mathematical Preliminaries

A. Posets

Definition 1. *A* partially ordered set *(poset) $\mathcal{P} = (\mathcal{S}, \sqsubseteq)$ is a set $\mathcal{S}$ and a binary relation $\sqsubseteq$ satisfying:*

1) *$i \sqsubseteq i$ (reflexivity)*
2) *$i \sqsubseteq j$ and $j \sqsubseteq i$ implies that $i = j$ (antisymmetry)*
3) *$i \sqsubseteq j$ and $j \sqsubseteq k$ implies that $i \sqsubseteq k$ (transitivity)*

We also define the relationship $i \sqsubseteq\sqsupseteq j$ if either $i \sqsubseteq j$ or $j \sqsubseteq i$. The symbols $\sqsubset, \sqsupseteq, \sqsupset, \not\sqsubseteq, \not\sqsupseteq, \not\sqsubseteq\not\sqsupseteq$ are defined as expected. In this paper we will assume that $\mathcal{P}$ is finite of size n (i.e. $|\mathcal{S}| = n$). Moreover, when the underlying set is clear, we will abuse notation and write $i \in \mathcal{P}$ and $\mathcal{T} \subseteq \mathcal{P}$ to denote membership in or subset relation to the set underlying $\mathcal{P}$ with inheritance of the order relation $\sqsubseteq$. Finally, we will also assume that $\mathcal{P}$ is connected: $\forall\ i, j \in \mathcal{P}$ there exists a finite sequence k_α such that $j \sqsubseteq\sqsupseteq k_1 \sqsubseteq\sqsupseteq \ldots \sqsubseteq\sqsupseteq k_l \sqsubseteq\sqsupseteq i$.

We note that for every partial order, we can always assign a linear (total) order consistent with the partial order. Such an ordering is called a linear extension.

Definition 2. *An* ordering *σ of the poset $\mathcal{P} = (\mathcal{S}, \sqsubseteq)$ is a bijection $\sigma : \mathcal{S} \to \{1, \ldots, n\}$. σ is called a* linear extension *of $\mathcal{P}$ if $j \sqsubseteq i \implies \sigma(j) \le \sigma(i)$.*

Throughout this paper, we will assume that a fixed linear extension has been chosen for every poset.

Example 1. *Posets are frequently drawn as Hasse diagrams where a directed edge from j to i is drawn if $j \sqsubset i$ and $\nexists k$ such that $j \sqsubset k \sqsubset i$. The $\sqsubseteq$ relationships are the directed transitive closure of this graph. In Figure 1a we draw the Hasse diagram of a poset with 5 nodes $\{1, 2, 3, 4, 5\}$ satisfying the relationships:*

- $1 \sqsubseteq 3 \sqsubseteq 4$
- $2 \sqsubseteq 3 \sqsubseteq 4$
- $1 \sqsubseteq 3 \sqsubseteq 5$
- $2 \sqsubseteq 3 \sqsubseteq 5$
- $1 \not\sqsubseteq\not\sqsupseteq 2$
- $4 \not\sqsubseteq\not\sqsupseteq 5$

We now introduce some common subsets of a poset.

Definition 3.

- ***Maximal (minimal) node:*** *i is maximal (minimal) if there does not exist $j \in \mathcal{P}$ such that $i \sqsubset j$ ($j \sqsubset i$)*
- ***Interval:*** *$[j, i] := \{k \in \mathcal{S} \mid j \sqsubseteq k \sqsubseteq i\}$*
- ***Downstream (upstream) set:*** *$\downarrow i := \{k \in \mathcal{S} \mid i \sqsubseteq k\}$, ($\uparrow i := \{k \in \mathcal{S} \mid k \sqsubseteq i\}$)*
- ***Set of nodes with common descendants:*** *$\rangle i \langle = \{k \in \mathcal{P} \mid \downarrow i \cap \downarrow k \neq \emptyset\}$*

Given a poset $\mathcal{P}$, its intervals inherit a natural ordering, given by setwise inclusion. The resulting poset is known as the poset of intervals $\text{Int}(\mathcal{P})$.

Definition 4. *Let $\mathcal{S} = \{[j, i] \mid i, j \in \mathcal{P},\ [j, i] \neq \emptyset\}$. Define $\preceq$ as $[j, i] \preceq [l, k]$ if $[j, i] \subseteq [l, k]$. The tuple $\text{Int}(\mathcal{P}) = (\mathcal{S}, \preceq)$ is the* poset of intervals *of $\mathcal{P}$.*

Since we have assumed that $\mathcal{P}$ is finite, then $\text{Int}(\mathcal{P})$ is also finite. We will denote $|\text{Int}(\mathcal{P})| =: N \le \binom{n+1}{2}$.

An example of an ordered relationship in $\text{Int}(\mathcal{P})$ for the poset $\mathcal{P}$ in Figure 1a is given by

$$\{3, 4\} \subseteq \{2, 3, 4\} \implies [3, 4] \preceq [2, 4]$$

However, since $5 \notin \{2, 3, 4\}$ then $[3, 5] \not\preceq\not\succeq [2, 4]$. The Hasse diagram for $\text{Int}(\mathcal{P})$ is shown in Figure 1c.

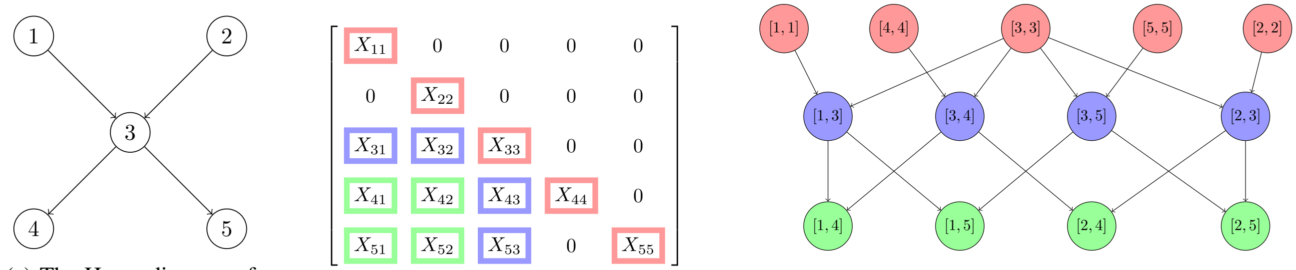
Finally, we recall the standard notion of the incidence algebra of a poset $\mathcal{P}$.

Definition 5. *Let $\mathcal{P}$ be a poset and $\mathbb{Q}$ a ring. The set of all functions $f : \mathcal{P} \times \mathcal{P} \to \mathbb{Q}$ such that $f(i, j) = 0$ if $j \not\sqsubseteq i$ is called the* incidence algebra of $\mathcal{P}$ over $\mathbb{Q}$ *and is denoted by $\mathcal{I}(\mathcal{P})$.*

Since $\mathcal{P}$ is finite, we can arrange the values of f into a lower triangular $\mathbb{Q}$-valued matrix with the rows and columns indexed by $\mathcal{P}$ in the fixed linear extension. This lower triangular matrix will have a sparsity pattern given by the order relationship of the poset. A generic example of an element of $\mathcal{I}(\mathcal{P})$ for the poset in Figure 1a is given in Figure 1b. Addition, scalar multiplication, and multiplication of $f, g \in \mathcal{I}(\mathcal{P})$ are defined to be compatible with the usual matrix operations. Under these definitions, $\mathcal{I}(\mathcal{P})$ forms an associative algebra [12].

Definition 6. *We say a matrix $Y \in \mathcal{I}(\mathcal{P})$ if Y has a sparsity pattern compatible with $\mathcal{I}(\mathcal{P})$. We say a matrix $Y \in \mathcal{I}(\mathcal{P})$* blockwise *if there is a partitioning of Y into n^2 blocks such that $Y_{ij} = 0$ if $j \not\sqsubseteq i$ or equivalently if $[j, i] \notin \text{Int}(\mathcal{P})$.*

Remark 1. *If $Y \in \mathcal{I}(\mathcal{P})$ blockwise, there is a natural pairing between the block Y_{ij} and the interval $[j, i]$. We choose to make Y lower triangular in order to be consistent with the poset-causal controller literature [3].*



(a) The Hasse diagram of the poset in Example 1

(b) A matrix $X \in \mathcal{I}(\mathcal{P})$.

(c) The Hasse diagram of Int($\mathcal{P}$) for the poset in Figure 1a.

Fig. 1: An example of a poset, a matrix X in $\mathcal{I}(\mathcal{P})$, and the associated poset of intervals. Notice that the X_{ij} block is in one-to-one correspondence with the element $[j, i] \in \text{Int}(\mathcal{P})$. The order of Int($\mathcal{P}$) naturally defines an order to solve the linear system defining the optimality conditions of (P1). Eliminating first the green nodes, then the blue nodes, and finally the red nodes reduces the number of arithmetic operations required to solve (P1)
.

B. Linear Systems and Graphs

As we will see in Section IV-A, solving problem (1) reduces to solving a sparse linear system. A (dense) system of M linear equations $\boldsymbol{L}_1(y) = b_1, \ldots, \boldsymbol{L}_M(y) = b_M$ in M variables $y_1, \ldots, y_M$ can be solved via (Gaussian) elimination with $\mathcal{O}(M^3)$ arithmetic operations. Using techniques from sparse linear algebra, one can typically improve on this complexity [13]. In this section we recall some standard notions from this literature that we will use later.

We begin by introducing the notion of a block symmetric linear system.

Definition 7. *Let $\boldsymbol{L} : \mathbb{R}^{\sum_{i=1}^M m_i} \to \mathbb{R}^{\sum_{i=1}^M m_i'}$ be a block-structured linear operator. Consider the linear system $\boldsymbol{L}(y) = b$ or equivalently the equations*

$$\boldsymbol{L}_i(y) = \sum_{j=1}^{M} \boldsymbol{L}_{ij}(y_j) = b_i \tag{2}$$

where y_j are block variables of size m_j. The system is symmetric *if the matrix representing $\boldsymbol{L}$ is symmetric.*

Given a linear system, it is always possible to reduce the number of variables via *elimination*. In the following definition, we make precise the notion of elimination that we use throughout this paper – this is essentially the Schur complement.

Definition 8. *Let $\mathcal{S} \subseteq \{1, \ldots, M\}$ and consider the partitioned linear system:*

$$\begin{aligned} \boldsymbol{L}_{\mathcal{S},\mathcal{S}}(y_{\mathcal{S}}) + \boldsymbol{L}_{\mathcal{S},\mathcal{S}^c}(y_{\mathcal{S}^c}) &= b_{\mathcal{S}} \\ \boldsymbol{L}_{\mathcal{S}^c,\mathcal{S}}(y_{\mathcal{S}}) + \boldsymbol{L}_{\mathcal{S}^c,\mathcal{S}^c}(y_{\mathcal{S}^c}) &= b_{\mathcal{S}^c} \end{aligned}$$

After eliminating variables $y_{\mathcal{S}^c}$, the resulting eliminated linear system *$\boldsymbol{L}^{\mathcal{S}}(y_{\mathcal{S}}) = b^{\mathcal{S}}$ is:*

$$\begin{aligned} &\left(\boldsymbol{L}_{\mathcal{S},\mathcal{S}} - \boldsymbol{L}_{\mathcal{S},\mathcal{S}^c}\boldsymbol{L}_{\mathcal{S}^c,\mathcal{S}^c}^{-1}\boldsymbol{L}_{\mathcal{S}^c,\mathcal{S}}\right)(y_{\mathcal{S}}) = \\ &\qquad\qquad b_{\mathcal{S}} - \boldsymbol{L}_{\mathcal{S},\mathcal{S}^c}\boldsymbol{L}_{\mathcal{S}^c,\mathcal{S}^c}^{-1} b_{\mathcal{S}^c} \end{aligned} \tag{3}$$

Remark 2. *When solving a linear system via elimination, elements from the set $\mathcal{S}$ are typically removed one at a time. While the order in which $\mathcal{S}$ is reduced changes the number of operations required to perform elimination, the end result is unique. That is to say, if $\mathcal{S} = \mathcal{T} \sqcup \mathcal{U} \sqcup \mathcal{W}$ is the disjoint union of three sets, then*

$$\begin{aligned} (\boldsymbol{L}^{\mathcal{T}\sqcup\mathcal{U}})^{\mathcal{T}}(y_{\mathcal{T}}) &= \left(b^{\mathcal{T}\sqcup\mathcal{U}}\right)^{\mathcal{T}} \\ (\boldsymbol{L}^{\mathcal{T}\sqcup\mathcal{W}})^{\mathcal{T}}(y_{\mathcal{T}}) &= \left(b^{\mathcal{T}\sqcup\mathcal{W}}\right)^{\mathcal{T}} \\ \boldsymbol{L}^{\mathcal{T}}(y_{\mathcal{T}}) &= b^{\mathcal{T}} \end{aligned}$$

are all the same equations [14]. Moreover, if $\boldsymbol{L}$ is symmetric, then $\boldsymbol{L}^{\mathcal{T}}$ will also be symmetric.

The complexity of solving a linear system $\boldsymbol{L}(y) = b$ via elimination is closely related to the interdependence between its variables. This interdependence can be described in terms of an undirected graph.

Definition 9. *Consider a block symmetric linear system of M equations in M block variables $\boldsymbol{L}(y) = b$. The* graph of a linear system *$\mathcal{G}(\boldsymbol{L}) = (\mathcal{V}(\boldsymbol{L}), \mathcal{E}(\boldsymbol{L}))$ is an undirected graph with vertex set $\mathcal{V}(\boldsymbol{L}) = \{1, \ldots, M\}$. The pair (i, j) forms an edge in $\mathcal{G}(\boldsymbol{L})$ if $\boldsymbol{L}_{ij}(\cdot) \neq 0$.*

Classical work in sparse linear algebra characterizes the complexity of solving systems by studying the edgeset of the sequence of graphs $\mathcal{G}(\boldsymbol{L}^{\mathcal{S}})$ as $\mathcal{S}$ is reduced. The number of arithmetic operations can be dramatically reduced by performing elimination in an order that adds a minimal number of edges to $\mathcal{G}(\boldsymbol{L}^{\mathcal{S}})$ as $\mathcal{S}$ shrinks [13], [15], [16]. These good orderings are completely characterized by $\mathcal{G}(\boldsymbol{L})$, which makes it possible to abstract away the underlying linear operators.

III. Problem Formulation

In this section, we formally introduce the problem studied in this paper.

Problem 1. *Let $\mathcal{P}$ be a poset and $A, B, C \in \mathcal{I}(\mathcal{P})$ blockwise. We assume that B has full row rank and C has full column rank. We consider the least-squares optimization problem*

$$\min_{X \in \mathcal{I}(\mathcal{P})} \|A - CXB\|_F^2 \,. \tag{P1}$$

From this point on, the norm involved will always be the Frobenius norm. Therefore we drop the subscript F.

Remark 3. *Problem* (P1) *appears simpler than the $\mathcal{H}_2$ control problem in [1], [9], [10] as it is finite dimensional. However, standard machinery such as Hilbert's projection theorem enables the solution technique we study here to directly generalize to the control case [9].*

In principle, it is always possible to solve (P1) using vectorization [7], i.e., solving the problem

$$\min_x \left\| \mathbf{vec}(A) - (B^T \otimes C)Ex \right\|_2 \tag{4}$$

where E is an $n^2 \times N$ matrix satisfying $Ex = \mathbf{vec}(X)$ and $\mathbf{vec}(X)$ is the vector obtained by stacking the columns of X into a vector. However, solving the problem in this manner via the normal equations involves the dense inversion of a large matrix of size $|N| \times |N|$, which completely ignores the structure of the problem and can be ill-conditioned [17]. Moreover, this technique results in a dramatic growth in the size of the state dimension when applied to the control problem and rapidly becomes intractable for even modestly sized problems [8].

Instead, we develop a solution strategy for solving (P1) which preserves the underlying sparsity implied by the incidence algebra throughout, resulting in much smaller, and in some cases decoupled, linear systems.

IV. Main Results

In this section, we present our main results. In Section IV-A we present the optimality conditions of (P1) as a set of linear equations with sparsity given by the common descendent relationship in the poset $\text{Int}(\mathcal{P})$. In subsequent sections, we study the graph of this linear system and its evolution as elimination is performed.

In Section IV-B, we show that elimination of X_{ij} removes some of the interdependencies between upstream (in the poset $\text{Int}(\mathcal{P})$) variables, leading to a *sparser* graph for the eliminated system than what is predicted by the standard theory. In Section IV-C, we specialize to multitree posets. For these posets, we use both the result of Section IV-B and the structure of the graph of the optimality conditions to define an order that maximizes the sparsity of intermediate systems during elimination.

A. Optimality Conditions

Our first theorem characterizes the sparsity pattern of the optimality conditions of (P1) implied by the poset and tensor product structure.

Theorem 1. *Let $\mathcal{P}$ be a poset of size n and* $\text{Int}(\mathcal{P})$ *its associated poset of intervals of size N. The solution to* (P1) *is given by the system of N linear equations:*

$$\boldsymbol{E}(X) = b, \tag{P2}$$

where the i, j^{th} equation $\boldsymbol{E}_{ij}(X) = b_{ij}$ is

$$\begin{aligned}
&\left(\sum_{s \sqsupseteq i} C_{si}^T C_{si}\right) X_{ij} \left(\sum_{t \sqsubseteq j} B_{jt} B_{jt}^T\right) + \\
&\sum_{\substack{[l,k] \in \curvearrowleft [j,i] \curvearrowright \\ [l,k] \neq [j,i]}} \left(\sum_{\substack{s \sqsupseteq i \\ s \sqsupseteq k}} C_{si}^T C_{sk}\right) X_{kl} \left(\sum_{\substack{t \sqsubseteq j \\ t \sqsubseteq l}} B_{lt} B_{jt}^T\right) = \\
&\qquad \sum_{[t,s]: t \sqsubseteq j, i \sqsubseteq s} C_{si}^T A_{st} B_{jt}^T
\end{aligned} \tag{5}$$

This system is symmetric and has a unique solution.

Immediate from (5) is the following corollary.

Corollary 1. *Two nodes form an edge in $\mathcal{G}(\boldsymbol{E})$ only if they have a common descendant in* $\text{Int}(\mathcal{P})$.

Using Corollary 1, we can use the Hasse diagram of $\text{Int}(\mathcal{P})$ shown in Figure 2a to construct the adjacency matrix of the graph $\mathcal{G}(\boldsymbol{E})$ shown in Figure 3. For example, the node $[2, 5]$ is a common descendant of the groups of nodes $\{[2, 2], [2, 3]\}$ and $\{[5, 5], [3, 5]\}$. Therefore, these nodes all share edges in $\mathcal{G}(\boldsymbol{E})$.

Proof: Since (P1) is equivalent to (4), we can write the first order optimality conditions in two equivalent ways:

$$E^T(BB^T \otimes C^T C)Ex = E^T(B^T \otimes C)\,\mathbf{vec}(A) \tag{6}$$

$$C^T AB^T - C^T CXBB^T \in \mathcal{I}(\mathcal{P})^c \text{ blockwise} \tag{7}$$

where $Y \in \mathcal{I}(\mathcal{P})^c$ blockwise if $Y_{ij} = 0$ whenever $j \sqsubseteq i$.

Symmetry of the first order optimality conditions is immediate from the symmetry of $E^T(BB^T \otimes C^TC)E$.

The expression for (5) is simply the i, j^{th} entry of (7) for $j \sqsubseteq i$. To see this expansion, we note that since $B, C \in \mathcal{I}(\mathcal{P})$:

$$\left(C^TC\right)_{ik} = \sum_{\substack{s \sqsupseteq i \\ s \sqsupseteq k}} C_{si}^T C_{sk}, \quad \left(BB^T\right)_{lj} = \sum_{\substack{t \sqsubseteq j \\ t \sqsubseteq l}} B_{lt} B_{jt}^T$$

Expanding the product of the second term of (7) yields

$$\left(C^TCXBB^T\right)_{ij} = \sum_k \sum_l \left(C^TC\right)_{ik} X_{kl} \left(BB^T\right)_{lj}$$

A term in this sum is zero unless $[l, k]$ is a non-empty interval, there exists s such that $s \sqsupseteq i$ and $s \sqsupseteq k$, and there exists $t \sqsubseteq l$ and $t \sqsubseteq j$. This implies that $[l, k]$ must share common descendants with $[j, i]$.

The expression for the right hand side of (5) is similarly obtained by considering the non-zero elements implied by the incidence algebra.

Uniqueness of the solution (P2) follows from uniqueness of the solution to (P1) due to the rank assumptions on B and C. $\square$

Theorem 1 demonstrates that solving (P1) can be reduced to solving the structured set of linear equations (P2). In subsequent sections, we will investigate aspects of this structure to characterize the sequence of graphs of the eliminated systems.

B. Separation in the Optimality Conditions

When $\mathcal{G}(\boldsymbol{L})$ is very sparse, techniques from [13], [15], [16] can substantially reduce the complexity of solving $\boldsymbol{L}(y) = b$. This is achieved by characterizing the evolution of $\mathcal{G}(\boldsymbol{L}^{\mathcal{S}})$ as elimination is performed. Unfortunately, though (P2) is quite structured, $\mathcal{G}(\boldsymbol{E})$ can be very dense. For example, when $\mathcal{P}$ is a chain graph, $\mathcal{G}(\boldsymbol{E})$ is a complete graph and therefore the results of [13], [15], [16] do not reduce the complexity of solving (P2).

In [13], [15], [16], the edges of $\mathcal{G}(\boldsymbol{L}^{\mathcal{S}})$ are entirely characterized from the graph $\mathcal{G}(\boldsymbol{L})$ which abstracts away the underlying operator. In contrast, the following theorem demonstrates that the tensor product structure of $\boldsymbol{E}$ can also have an effect on $\mathcal{G}(\boldsymbol{E}^{\mathcal{S}})$. In particular, it shows that after one step of elimination, the system can in fact be *sparser* than what would be predicted by ignoring the underlying operator.

Theorem 2. *Consider the interval $[j, i]$ and two upstream intervals $[j, q] \preceq [j, i]$ and $[p, i] \preceq [j, i]$. Then after eliminating X_{ij} from* (P2)*, X_{qj} vanishes from equation $[p, i]$ and symmetrically X_{ip} vanishes from equation $[j, q]$ i.e.*

$$\boldsymbol{E}_{qj}^{\{[j,i]\}^c}(\cdot) = \boldsymbol{E}_{ip}^{\{[j,i]\}^c}(\cdot) = 0$$

and so $([p, i], [j, q]) \notin \mathcal{E}(\boldsymbol{E}^{\{[j,i]\}^c})$.

Theorem 2 demonstrates that $\boldsymbol{E}$ contains a notable separability property: the elimination of downstream (in the poset Int($\mathcal{P}$)) variables from the system can reduce the interdependence of upstream variables.

For instance, in Figure 2a, the green nodes $\{[2, 2], [2, 3]\}$ and the blue nodes $\{[5, 5], [3, 5]\}$ share an edge in $\mathcal{G}(\boldsymbol{E})$ due to their common descendant relationship to $[2, 5]$. Elimination of X_{52} from (P2) eliminates the node $[2, 5]$ from the graph as in Figure 2b. Theorem 2 implies that the edges between the blue and green nodes are also eliminated.

Proof: Using (5), we can write X_{ij} as a function of the remaining variables. From this expression, notice that the only term that depends on X_{ip} is:

$$-\left(\sum_{s \sqsupseteq i} C_{si}^T C_{si}\right)^{-1} \left(\sum_{s \sqsupseteq i} C_{si}^T C_{si}\right) * \\ X_{ip} \left(\sum_{\substack{t \sqsubseteq j \\ t \sqsubseteq l}} B_{pt} B_{jt}^T\right) \left(\sum_{t \sqsubseteq j} B_{jt} B_{jt}^T\right)^{-1} \quad (8)$$

If we consider the dependence of $\boldsymbol{E}_{qj}(X) = b_{qj}$ on X_{ij} and X_{ip} we have the expression:

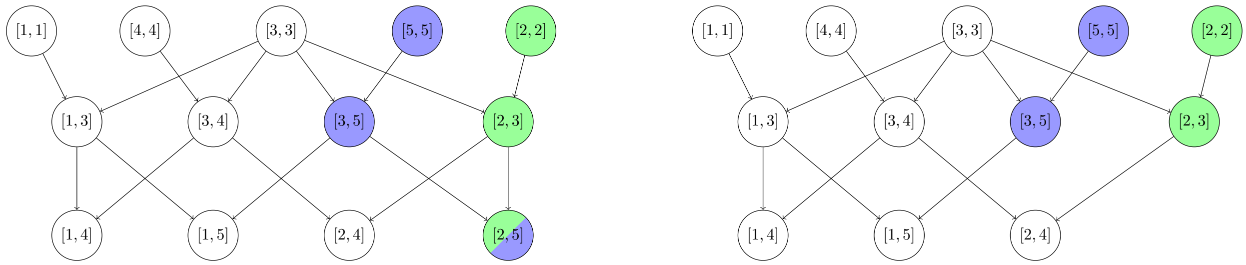
$$\left(\sum_{s \sqsupseteq i} C_{sq}^T C_{si}\right) X_{ij} \left(\sum_{t \sqsubseteq j} B_{jt} B_{jt}^T\right) + \\ \left(\sum_{s \sqsupseteq i} C_{sq}^T C_{si}\right) X_{ip} \left(\sum_{\substack{t \sqsubseteq j \\ t \sqsubseteq p}} B_{pt} B_{jt}^T\right) \quad (9)$$

The result of substituting (8) into (9) is:

$$-\left(\sum_{s \sqsupseteq i} C_{sq}^T C_{si}\right) X_{ip} \left(\sum_{\substack{t \sqsubseteq j \\ t \sqsubseteq l}} B_{pt} B_{jt}^T\right) + \\ \left(\sum_{s \sqsupseteq i} C_{sq}^T C_{si}\right) X_{ip} \left(\sum_{\substack{t \sqsubseteq j \\ t \sqsubseteq p}} B_{pt} B_{jt}^T\right) = 0 \quad (10)$$

Therefore $\boldsymbol{E}_{qj}^{\{[j,i]\}^c}(\cdot) = 0$. The proof that $\boldsymbol{E}_{ip}^{\{[j,i]\}^c}(\cdot) = 0$ is similar. Therefore, $[j, q]$ and $[p, i]$ do not share an edge in $\mathcal{G}(\boldsymbol{E}^{\{[j,i]\}^c})$. $\square$

While Theorem 2 characterizes the disappearance of some edges after a *single* step of elimination, it is insufficient to predict whether this sparsification continues



(a) The node $[2,5]$ is a common descendant of the nodes $\{[2,2],[2,3]\}$ (highlighted in green) and $\{[5,5],[3,5]\}$ (highlighted in blue). By Corollary 1, these nodes all share edges in $\mathcal{G}(\boldsymbol{E})$.

(b) Theorem 2 asserts that after elimination of $[2,5]$, the blue and green nodes no longer share an edge $\mathcal{E}(\boldsymbol{E}^{\{[2,5]\}^c})$ due to the elimination of $[2,5]$.

Fig. 2

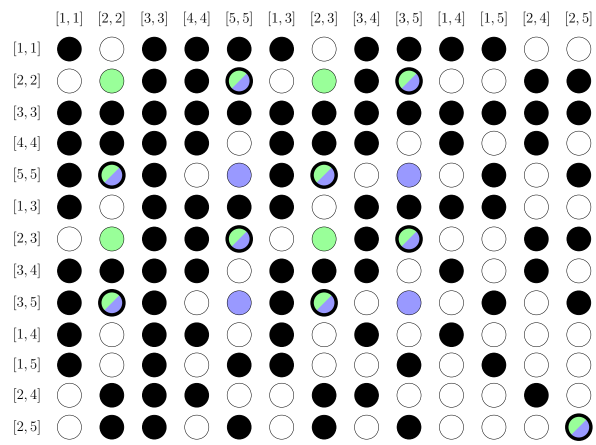


Fig. 3: The adjacency matrix of $\mathcal{G}(\boldsymbol{E})$. A filled circle represents an edge in $\mathcal{G}(\boldsymbol{E})$ and an empty circle represents the lack of an edge. The edges implied by $[2,2] \preceq [2,3]$ and $[5,5] \preceq [3,5]$ are highlighted in green and blue respectively. The edges between $\{[2,2],[2,3]\}$ and $\{[5,5],[3,5]\}$ implied by the common descendant $[2,5]$ are highlighted in blue and green.

after multiple steps. Moreover, it fails to characterize the entire graph $\mathcal{G}(\boldsymbol{E}^{\mathcal{S}})$. In the next section, we will show that this progressive sparsification can be continued when $\mathcal{P}$ is a multitree.

C. Sparsification During Elimination in Multitree Posets

In Theorem 2, we showed that elimination of downstream variables can decrease the interdependence between upstream variables in the system (P2). In this section, we demonstrate how an additional assumption on $\mathcal{P}$ can further structure (P2) in such a way that we can completely characterize the entire sequence of edges of the graphs of the eliminated system $\mathcal{G}(\boldsymbol{E}^{\mathcal{S}})$.

We begin by formally defining multitree posets:

Definition 10 ([1, Definition 1])**.** *Let $\mathcal{P}$ be a poset. Nodes $i,j,k,l \in \mathcal{P}$ form a diamond if $i \sqsubseteq j \sqsubseteq l$ and $i \sqsubseteq k \sqsubseteq l$ but $j \not\sqsubseteq\not\sqsupseteq k$. $\mathcal{P}$ is a* multitree *if it contains no diamonds. Equivalently $\mathcal{P}$ is a multitree if every interval is totally ordered.*

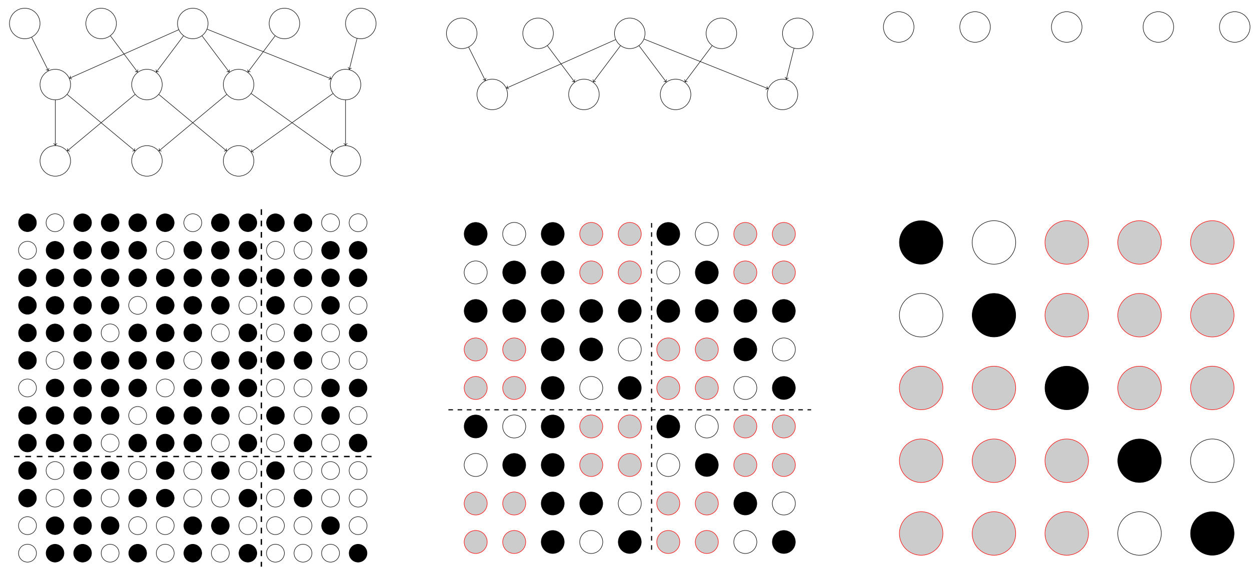
This class of posets was studied in [1] where it was shown that the optimal poset causal controllers of such posets enjoy attractive recursive properties. Here, we show that the multitree assumption can be leveraged to decrease the complexity of solving (P2). This occurs because the assumption that $\mathcal{P}$ is a multitree imposes substantial structure on $\text{Int}(\mathcal{P})$ and in turn $\boldsymbol{E}$. One property that we use here is:

Proposition 1. *Let $\mathcal{P}$ be a multitree and let $[l,k]$ and $[t,s]$ be two nodes in* $\text{Int}(\mathcal{P})$ *with $t \not\sqsubseteq l$ whose only common descendant is $[j,i]$. Then $l = j$ and $s = i$.*

Proof: Since $\mathcal{P}$ is a multitree, $[j,i]$ is a chain and so l,k,s,t are all mutually comparable. Therefore without loss of generality we can assume that $j \sqsubseteq l \sqsubseteq s \sqsubseteq i$. If $s \sqsubset i$ then $[l,i]$ contains both intervals and similarly if $l \sqsupset j$ then $[j,s]$ contains both intervals. In either case, the assumption that $[l,k]$ and $[t,s]$ share only one common descendant is violated. $\square$

In the following theorem, we will discuss the structure of the equations after a linearly ordered subset of variables is eliminated. We fix a linear extension σ of $\text{Int}(\mathcal{P})$ and perform elimination in the reverse of this order. After r steps of elimination, the remaining variables can be associated to the subposet

$$\text{Int}(\mathcal{P})^{(r)} := \{[l,k] \in \text{Int}(\mathcal{P}) \mid \sigma([l,k]) \leq N-r\} \quad (11)$$



(a) (b) (c)

Fig. 4: Evolution of the graph $\mathcal{G}(\boldsymbol{E}^{(r)})$. The black nodes represent nonzero elements, the white nodes correspond to structural zeros inherited from the structure of Int$(\mathcal{P})$, and the gray nodes vanish due to the additional cancellations from Theorem 2. Theorem 3 asserts that $\mathcal{G}(\boldsymbol{E}^{(r)})$ can always be constructed by inspection of the remaining Hasse diagram. The rows and columns to the right and below of the dashed lines in the first two frames correspond to the variables eliminated.

and the eliminated system will be given by

$$\boldsymbol{E}^{(r)}(X_{(r)}) = b^{(r)} := \boldsymbol{E}^{\text{Int}(\mathcal{P})^{(r)}}(X_{\text{Int}(\mathcal{P})^{(r)}}) = b^{\text{Int}(\mathcal{P})^{(r)}} \tag{12}$$

We now present our main theorem on the sequence $\mathcal{G}(\boldsymbol{E}^{(r)})$ of (P2).

Theorem 3. *Let $\mathcal{P}$ be a multitree and fix a linear extension σ for* Int$(\mathcal{P})$*. Then two nodes form an edge in $\mathcal{G}(\boldsymbol{E}^{(r)})$ only if they share a common descendant in* Int$(\mathcal{P})^{(r)}$.

Theorem 3 fully characterizes the edgeset of the graph $\mathcal{G}(\boldsymbol{E}^{(r)})$ during the entire course of elimination in multitrees. This improves upon the result in Theorem 2 which only characterizes the effect of a single step of elimination for only some of the edges of $\mathcal{G}(\boldsymbol{E})$. This characterization is similar to the main theorems of [13], [15], [16]. However, Theorem 3 provides a strict improvement over these works by predicting that $\mathcal{G}(\boldsymbol{E}^{(r)})$ is in fact *sparser* than these other work predict. This progressive sparsification of the graph is visualized in Figure 4. The increased sparsity is a consequence of leveraging both the tensor product structure of (P2) and the graph structure of $\mathcal{G}(\boldsymbol{E}^{(r)})$.

Proof: Let $[j,i] = \sigma^{-1}(N)$. $[j,i]$ is a maximal node of Int$(\mathcal{P})$ and so $\rangle[j,i]\langle = \uparrow [j,i]$. Therefore, the neighbors of $[j,i]$ form a clique in $\mathcal{G}(\boldsymbol{E})$. If $\boldsymbol{E}$ were a generic linear system, [13] predicts that $\mathcal{E}(\boldsymbol{E}^{(1)})$ is exactly the edgeset of the induced subgraph of $\mathcal{V}(\boldsymbol{E}^{(1)})$ in $\mathcal{G}(\boldsymbol{E})$. However, consider nodes $[j,p],[q,i] \in \mathcal{V}(\boldsymbol{E}^{(1)})$ whose only common descendant is node $[j,i]$. By Theorem 2, $[j,p]$ and $[q,i]$ do not share an edge in $\mathcal{G}(\boldsymbol{E}^{(1)})$. By Proposition 1, these are the only possible sets of nodes whose only common descendant is $[j,i]$. Therefore, two nodes share an edge in $\mathcal{G}(\boldsymbol{E}^{(1)})$ only if they share a common descendant in Int$(\mathcal{P})^{(1)}$.

We can use the same argument to continue the induction. Suppose $\mathcal{E}(\boldsymbol{E}^{(r-1)})$ is given by the common descendant relationship of the poset Int$(\mathcal{P})^{(r-1)}$. Since $[j,i] = \sigma^{-1}(N-r+1)$ is maximal in Int$(\mathcal{P})^{(r-1)}$ then the neighbors of $[j,i]$ again form a clique in $\mathcal{G}(\boldsymbol{E}^{(r-1)})$. Therefore, after elimination of $[j,i]$, $\mathcal{E}(\boldsymbol{E}^{(r)}) \subseteq \mathcal{E}(\boldsymbol{E}^{(r-1)})$. Using Theorem 2 and the fact that the resulting structure of $\boldsymbol{E}^{(r)}(X_{(r)})$ is a set property of the remaining variables (Remark 2) we have that $\mathcal{E}(\boldsymbol{E}^{(r)}))$ is given by the common descendant relationship of Int$(\mathcal{P})^{(r)}$. $\square$

V. Conclusion

In this work, we investigate the algebraic structure of the optimality conditions for a general least squares problem structured according to a partially ordered set. This is a fundamental optimization problem underlying the design of output feedback controllers operating on plants with directed, acyclic information structures.

We demonstrated that the optimality conditions of this problem are a linear system which inherit the sparsity pattern given by the common descendant relationship of the poset of intervals. This relationship can result in a relatively dense linear system and thus standard results from sparse linear algebra predict little, if any, computational savings.

By leveraging the particular algebraic structure of the optimality conditions, we established that elimination of downstream variables can sparsify this linear system. When the underlying poset is a multitree, this progressive sparsification can be achieved at every step of elimination when the variables are eliminated according to the reverse of a linear extension of the poset of intervals. We completely characterize the edgeset of the sequence of graphs of the eliminated system which has implications for the complexity of solving the poset-structured least squares problem.

Future work will characterize the exact computational implications of this result and investigate the generalization to the $\mathcal{H}_2$ problem as well as broader classes of posets containing diamonds.

References

[1] A. Nayyar and L. Lessard, "Structural results for partially nested LQG systems over graphs," in *2015 American Control Conference (ACC)*, Chicago, IL, USA, 2015, pp. 5457–5464.

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

[3] P. Shah and P. A. Parrilo, "$\mathcal{H}_2$-Optimal decentralized control over posets: A state-space Solution for state-feedback," *IEEE Transactions on Automatic Control*, vol. 58, no. 12, pp. 3084–3096, Dec. 2013, arXiv: 1111.1498.

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

[5] L. Lessard and S. Lall, "An algebraic approach to the control of decentralized systems," *IEEE Transactions on Control of Network Systems*, vol. 1, no. 4, pp. 308–317, Dec. 2014.

[6] J. Anderson, J. C. Doyle, S. H. Low, and N. Matni, "System level synthesis," *Annual Reviews in Control*, vol. 47, pp. 364–393, 2019.

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

[8] L. Lessard and S. Lall, "Optimal control of two-player systems with output feedback," *IEEE Transactions on Automatic Control*, vol. 60, no. 8, pp. 2129–2144, Aug. 2015.

[9] T. Tanaka and P. A. Parrilo, "Optimal output feedback architecture for triangular LQG problems," in *2014 American Control Conference*. IEEE, 2014, pp. 5730–5735.

[10] L. Lessard, "Decentralized LQG control of systems with a broadcast architecture," in *2012 IEEE 51st IEEE Conference on Decision and Control (CDC)*, Dec. 2012, pp. 6241–6246.

[11] P. Shah and P. Parrilo, "An optimal controller architecture for poset-causal systems," *IEEE Conference on Decision and Control and European Control Conference*, pp. 5522–5528, Dec. 2011, arXiv: 1111.7221.

[12] R. P. Stanley, *Enumerative Combinatorics*. Cambridge Studies in Advanced Mathematics, 2011, vol. I, second edition.

[13] R. E. Tarjan, "Graph theory and Gaussian elimination," *Sparse Matrix Computations*, pp. 3–22, 1976.

[14] D. E. Crabtree and E. V. Haynsworth, "An identity for the Schur complement of a matrix," *Proceedings of the American Mathematical Society*, vol. 22, no. 2, pp. 364–366, 1969.

[15] T. F. Coleman, A. Edenbrandt, and J. R. Gilbert, "Predicting fill for sparse orthogonal factorization," *Journal of the ACM (JACM)*, vol. 33, no. 3, pp. 517–532, 1986.

[16] T. A. Davis, *Direct methods for sparse linear systems*. SIAM, 2006.

[17] L. Eldén and I. Skoglund, *Algorithms for the regularization of ill-conditioned least squares problems with tensor product structure, and application to space-variant image restoration*. Universitetet i Linköping/Tekniska Högskolan, 1982.