Communication Delay Co-Design in $\mathcal{H}_2$ Distributed Control Using Atomic Norm Minimization

Nikolai Matni

Abstract—It has been shown that the $\mathcal{H}_2$ distributed control problem subject to quadratically invariant delay constraints admits a finite dimensional solution. When these delay constraints are induced by a strongly connected communication graph, this problem can be solved by decomposing the controller into a centralized but delayed component, and a distributed finite impulse response component, the latter of which can be solved for via a linearly constrained quadratic program. In this paper, we propose an atomic norm minimization based variant of this quadratic program that can be used for the co-design of a distributed controller and the communication graph on which it is to be implemented, allowing for a principled trade off between the closed loop performance of the system and the complexity of the resulting communication graph. We additionally show that the resulting co-designed solution satisfies many desirable properties: the communication graph is strongly connected, induces a quadratically invariant constraint set, and the closed loop norm of the designed system satisfies a pre-specified bound.

I. INTRODUCTION

Distributed optimal control problems arise when several controllers must coordinate their actions subject to information sharing constraints in order to control an underlying plant that is comprised of several interacting subsystems. Such optimal control problems subject to so-called non-classical information constraints are in general intractable, cf. [1], with convexity of the underlying optimization being roughly determined by the amount and speed of information sharing between controllers relative to the underlying propagation of dynamics between subsystems [2], [3], [4], [5], [6].

In [5], the broadest known characterization of such convex distributed optimal control problems was made in terms of quadratically invariant information sharing constraints. We provide a brief survey of results most relevant to our paper in the following, and refer the reader to the tutorial paper [7] for an overview of the current state of the art in optimal distributed control subject to information constraints.

A particular class of distributed control problems that has received a significant amount of attention is that of optimal $\mathcal{H}_2$ (or LQG) control subject to quadratically invariant delay constraints. In this case, the information constraints can be interpreted as arising from a communication graph, in which edge weights between nodes correspond to the delay required to transmit information between them. Intuitive and simple to verify conditions for quadratic invariance in terms of the communication delays between controllers and the propagation delays of dynamics between subsystems are provided in [6]—these results will be key to our development, and will be reviewed in Section III.

When these conditions for quadratic invariance are met the resulting convex distributed problem has been solved via vectorization [5], semi-definite programming (SDP) [8], [9], and most recently using an extension of spectral factorization [10], [11]. We also note that analogous results exist for spatially invariant distributed systems subject to delay constraints: in particular, [4] provides a sufficient condition (namely funnel-causality) for the convexity of the distributed control problem subject to communication constraints, and [12] and [13] provide principled means of computing sub-optimal controllers for such problems.

An assumption common to the aforementioned controller synthesis results is that a communication graph topology has already been designed that is well suited for distributed optimal control. Solving for the optimal (with respect to graph complexity and controller performance) communication network is inherently combinatorial in nature, and tractable methods for computing exact solutions are unlikely. An approach that has seen much success in similar problems in other fields has been to employ convex relaxations in order to approximately recover such solutions.

The machine learning and statistics communities have been particularly fruitful in applying this idea. It is often known a priori that the solution to an inference problem should be structurally “simple” — it has been shown that this simple structure can often be approximately, and sometimes exactly, recovered by minimizing an appropriately chosen convex penalty function. Well known examples include the $\ell_1$-norm to induce sparse solutions, and the nuclear norm to induce low-rank solutions (cf. [14], and the references therein). In [15], this notion of “simplicity” was formalized and generalized in terms of atomic norms.

Representatives of the use of such ideas in the control literature for the design of information sharing constraints have only recently begun to emerge. These include the use of $\ell_1$-regularization to design sparse $\mathcal{H}_2$ optimal feedback gains [16], [17], sparse treatment therapies [18], and sparse consensus [19], [20] and synchronization [21] topologies. In addition to design applications, these ideas have also been

1What we refer to as communication delays are called transmission delays in [6].

2This result was subsequently generalized in [6].
successfully applied in the context of system identification [22].

In order to apply these types of techniques, the problem to be solved must ultimately be reduced to a finite dimensional convex program. In [10], [11], it is shown that the distributed $\mathcal{H}_2$ problem subject to delay constraints induced by a strongly connected communication graph can in fact be reduced to a finite dimensional linearly constrained quadratic program. In particular, the optimal controller is solved for by decomposing it into a centralized, but delayed, component (thus the need for strong connectivity) and a distributed finite impulse response (FIR) component. It is this FIR component that is solved for in the quadratic program.

In a preliminary version of this work [23], we exploited the fact that the entire distributed nature of the problem is captured in this FIR element, and borrowed ideas from atomic norm minimization to propose a convex program that induced sparsity patterns that are consistent with how information propagates through communication graphs. In particular, we identified an appropriate atomic norm for inducing the desired structure in the FIR element of the controller, and formulated the distributed controller and communication graph co-design problem as a finite dimensional second order cone program (SOCP).

Contributions: This manuscript differs significantly from our previous work. In particular, we establish a rigorous framework linking the delay based formalism presented in [6] to the sparse subspace framework of [10], [11], allowing us to discuss the quality of the co-designed solution in a precise way. We also develop a computationally more efficient variant of the previously used atomic norm, which we term the graph enhancement norm. We then show that our approach based on the graph enhancement norm is able to identify a constraint set that corresponds to one induced by a strongly connected communication graph, and that is quadratically invariant.

Further, as is shown in Section VI, in order for our algorithm to be tractable, a certain degree of conservatism is introduced into the communication graph design process (this was informally alluded to in [23]) – we also make this statement more precise.

Article structure: In Section II, we fix notation and introduce the $\mathcal{H}_2$ distributed optimal control problem subject to delays. We then review the results of [6] in Section III, in which a characterization of quadratic invariance is provided in terms of the communication delays of a system. We also link these delays to an underlying communication graph, and to the subspace constraint that they induce. In Section IV, we integrate the discussion of the two previous sections to formally state the distributed control problem considered in [10]. Sections V and VI contain the main results of the paper, namely a graph enhancement norm based co-design algorithm, and a discussion of the properties of the resulting solution. In Section VII, we show how the graph enhancement norm can be implemented computationally as well as numerical examples illustrating the usefulness of the approach. Finally, we end with conclusions and directions for future work in Section VIII. Proofs of all intermediate results can be found in Appendix A.

II. Preliminaries

A. Notation

If $\mathcal{M}$ is a subspace of an inner product space, we denote the orthogonal projection onto $\mathcal{M}$ by $\mathbb{P}_\mathcal{M}$. We let $\otimes$ denote the Kronecker product, and $\text{vec}(\cdot) : \mathbb{R}^{p \times q} \rightarrow \mathbb{R}^{pq}$ be the vectorization operator that maps a matrix to a vector through the stacking of its columns. For a $n$ block by $n$ block matrix $M$, we define the block support operator $\text{bsupp} (\cdot) : \mathbb{R}^{p \times q} \rightarrow [\mathbb{Z}_{+}]_{1 \times n}$, where $[\mathbb{Z}_{+}]$ is the set of non-negative integers, as $(\text{bsupp} (M))_{kl} = 1$ if the $(k,l)$ block of $M$ is non-zero, and 0 otherwise. For a complex matrix $M \in \mathbb{C}^{p \times q}$, we denote its conjugate transpose by $M^\dagger$, and reserve $M^*$ to denote when $M$ is the solution to an optimization problem.

We extend the standard Banach space $\ell_2^n$ to the extended space

$$\ell_2^d := \{ f : \mathbb{Z}_+ \rightarrow \mathbb{R}^d | f_T \in \ell_2^n \text{ for all } T \in \mathbb{Z}_+ \},$$

where $f_T$ is the truncation of $f$ to its first $T$ elements.

B. Operator Theoretic Preliminaries

Let $\mathbb{D} = \{ z \in \mathbb{C} : |z| < 1 \}$ be the unit disc of complex numbers, and let $\overline{\mathbb{D}}$ be its closure. A function $G : (\mathbb{C} \cup \{ \infty \}) \setminus \overline{\mathbb{D}} \rightarrow \mathbb{C}^{p \times q}$ is in $\mathcal{H}_2$ if it can be expanded as $G(z) = \sum_{i=0}^{\infty} \frac{1}{\tau_i} G_i$, where $G_i \in \mathbb{C}^{p \times q}$ and $\sum_{i=0}^{\infty} \text{Tr}(G_i G_i^\dagger) < \infty$. Define the conjugate of $G$ by $G(z)^\sim = \sum_{i=0}^{\infty} z^i G_i^\dagger$. $\mathcal{H}_2$ is a Hilbert space with inner product given by

$$\langle G, H \rangle = \frac{1}{2\pi} \int_0^{2\pi} \text{Tr}(G(e^{j\theta})H(e^{j\theta})^\sim) d\theta = \sum_{i=0}^{\infty} \text{Tr}(G_i H_i^\dagger),$$

where the last equality follows from Parseval’s identity.

A function $G : (\mathbb{C} \cup \{ \infty \}) \setminus \overline{\mathbb{D}} \rightarrow \mathbb{C}^{p \times q}$ is in $\mathcal{H}_\infty$ if it is analytic, bounded and has a well-defined limit $G(e^{j\theta}) \in \mathbb{C}^{p \times q}$ almost everywhere on the unit circle. Note that as we are working in discrete time, $\mathcal{H}_\infty = \mathcal{H}_2$, as they both correspond to the space of transfer matrices with no poles outside of the unit disk $\mathbb{D}$ – as a matter of convention we will refer to this space as $\mathcal{H}_\infty$.

Let $\mathcal{R}_p$ denote the space of proper real transfer matrices. We append the prefix $\mathcal{R}$ to a space if we wish to restrict it to $\mathcal{R}_p$ – for example, $\mathcal{R} \mathcal{H}_\infty$ denotes the space of proper real transfer matrices that are also in $\mathcal{H}_\infty$.

Finally, as our development will depend on finite dimensional subspaces of $\mathcal{H}_\infty$, we define for any $D \geq 1$ the space of strictly proper finite impulse response (FIR) transfer matrices $\mathcal{X}^{p \times q}_D := \oplus_{p=1}^{D} \mathbb{C}^{p \times q}$, and thus $\mathcal{R} \mathcal{X}^{p \times q}_D = \oplus_{p=1}^{D} \frac{1}{2\pi} \mathbb{R}^{p \times q}$. We will often drop the dimension labeling superscripts $(p,q)$ and duration labeling subscript $D$ when they are clear from context.

C. $\mathcal{H}_2$ optimal control subject to delays

Let $P$ be a stable discrete-time plant given by

$$P = \begin{bmatrix} A & B_1 & B_2 \\ C_1 & 0 & D_{12} \\ C_2 & D_{21} & 0 \end{bmatrix} = \begin{bmatrix} P_{11} & P_{12} \\ P_{21} & P_{22} \end{bmatrix}$$ (3)
with inputs of dimension $p_1, p_2$ and outputs of dimension $q_1, q_2$. It will be convenient to view $P_{22}$ as a map from an input space $U = \ell_2^{p_2}$ to an output space $Y = \ell_2^{q_2}$ for reasons that will be made clear in the next section.

Finally, we note that we restrict attention to stable plants for clarity of exposition, but will end with an outline of how our ideas easily extend to unstable plants using the results from [11] in Section V-D2 — it is for this reason that we work with the extended spaces rather than the standard $\ell_2$ Banach spaces.

The distributed control problem of interest is to design a controller $K \in S$ so as to minimize the closed loop $H_2$ norm of the system:

$$\begin{align*}
\text{minimize} & \quad ||P_{11} + P_{12}K(I - P_{22}K)^{-1}P_{21}||_{H_2} \\
\text{s.t.} & \quad K \in S, \ K \text{ stabilizes } P
\end{align*}$$

where $S$ is a closed subspace constraint encoding the distributed nature of the controller. In order to formulate this problem as a convex model matching problem, we require the notion of a quadratically invariant constraint set [5].

**Definition 1:** A set $S$ is quadratically invariant under $P_{22}$ if

$$KP_{22}K \in S \text{ for all } K \in S. \quad (5)$$

If the constraint set $S$ is quadratically invariant, then $K \in S$ if and only if

$$Q := K(I - P_{22}K)^{-1} \in S. \quad (6)$$

Applying this fact and the standard Youla re-parametrization to (4) leads to the equivalent convex model matching problem:

$$\begin{align*}
\text{minimize} & \quad ||P_{11} + P_{12}QP_{21}||_{H_2} \\
\text{s.t.} & \quad Q \in S \cap \mathcal{RH}_\infty
\end{align*}$$

Although the constraint set $S$ can be used to encode constraints arising from various types of distributed architectures [7], we focus on those corresponding to delay patterns induced by strongly connected communication graphs. As our results are dependent on the relationship between quadratic invariance, delay patterns, communication graphs and the subspace constraints that they induce, we devote the next section to presenting results on the quadratic invariance of constraint sets defined by delays.

### III. Quadratic Invariance and Delays

We follow the formalism introduced in [6], and consider a general plant (3) comprised of $n$ subsystems, each with its own controller. Letting $N := \{1, \ldots, n\}$, we then associate with each subsystem $i \in N$ a set of possible control actions $U^i = \ell_2^{p_2}$, and a set of possible output measurements $Y^i = \ell_2^{q_2}$, and define the overall control and measurement spaces as

$$U := \ell_2^{1 \times \cdots \times U^i} \quad \text{and} \quad Y := \ell_2^{1 \times \cdots \times Y^i}.$$ 

Then, for any pair of subsystems $(i, j) \in N \times N$, we have that the $(i, j)^{th}$ block of $P_{22}$ is the mapping from the control action at node $j$ to the measurement at node $i$, i.e. $P_{22}^{ij} : U_j \rightarrow Y_i$. Similarly, the component of the controller at subsystem $i$ that uses available information from subsystem $j$ is given by $K^{ij} : Y_j \rightarrow U_i$.

Under this setup, we can then form the overall measurement and control vectors as

$$y = [(y_1)^T \cdots (y_n)^T]^T, \quad u = [(u_1)^T \cdots (u_n)^T]^T,$$

leading to the natural partitions of $P_{22}$ and $K$

$$P_{22} = \begin{bmatrix} P_{11}^{11} & \cdots & P_{11}^{1n} \\ \vdots & \ddots & \vdots \\ P_{n1}^{11} & \cdots & P_{n1}^{nn} \end{bmatrix}, \quad K = \begin{bmatrix} K^{11} & \cdots & K^{1n} \\ \vdots & \ddots & \vdots \\ K^{n1} & \cdots & K^{nn} \end{bmatrix}. \quad (8)$$

We define the Delay $(\cdot)$ operator of a causal linear time-invariant map $G$, with impulse response elements $\{G_r\}_{r=0}^{\infty}$, as

$$\text{Delay}(G) = \min \{\tau \geq 0 | G_r \neq 0\}, \quad (9)$$

and the propagation delay $p_{ij}$ from a controller at subsystem $j$ to the measured output at subsystem $i$ as

$$p_{ij} := \text{Delay} \left( P_{22}^{ij} \right), \quad (10)$$

that is the amount of time before a control action taken at $j$ can affect a measured output at subsystem $i$.

#### A. Communication Delay Induced Constraints

As mentioned previously, we are concerned with constraints induced by communication delays between controllers. In particular, define the communication delay $t_{kl}$ between subsystems $k$ and $l$ as the minimum amount of time before the controller of subsystem $k$ may use the measured outputs $y^l$ from subsystem $l$. Given these constraints, we can then naturally define the subspace of admissible controllers $S$ such that $K \in S$ if and only if

$$\text{Delay}(K_{kl}) \geq t_{kl}, \text{ for all } (k, l) \in N \times N. \quad (11)$$

**Remark 1:** It is possible to further break these communication delays into a pure communication delay and a computational delay — although informative, this further refinement is not necessary for our development, and as such we will content ourselves with dealing exclusively with communication delays.

We further assume that communication delays satisfy a triangle inequality, namely that

$$t_{kl} + t_{ij} \geq t_{kj}, \text{ for all } (k, i, j) \in N \times N \times N. \quad (12)$$

As is argued in [6], this is indeed a reasonable assumption as is posed on the communication delays of the system, as it corresponds to information being routed between controllers according to shortest path times — any violation of this triangle inequality could accordingly be rectified via a shortest path algorithm.

With these definitions in place, we can now state the main result from [6], which gives a natural characterization of quadratically invariant constraint sets in terms of the communication and propagation delays of the system.

**Theorem 1 (cf. Theorem 3 in [6]):** Suppose that $P_{22}$ and $S$ are defined as above, with propagation and communication delays $\{p_{ij}\}$ and $\{t_{kl}\}$, respectively, and that the communication delays satisfy the triangle inequality (12). If

$$p_{ij} \geq t_{ij}, \text{ for all } (i, j) \in N \times N \quad (13)$$

then $S$ is quadratically invariant under $P_{22}$. 


B. Communication Delays and Graphs

The communication delays implicitly define a communication network between controllers. In this section, we define a certain class of communication delay patterns that are easily associated with an underlying directed graph, and make explicit connections between the support of the impulse response elements of $K$, the underlying communication graph, and the communication delay pattern that it induces.

In order to do so, we first introduce some concepts from algebraic graph theory. Let $G = (V, E)$ be a graph, where $V$ is the set of nodes, and $E$ the edge set. Following [24], we define the adjacency matrix $L(G)$ of a directed graph $G$ as the integer matrix with rows and columns indexed by the vertices of $G$, such that $(L(G))_{kl}$ is equal to the number of edges from $k$ to $l$, typically 0 or 1. When the underlying graph is clear from context, we will simply denote $L(G)$ by $L$.

A walk of length $r$ in a directed graph $G$ is a sequence of (not necessarily distinct) vertices $(v_0, v_1, \ldots, v_r)$ such that $(v_t, v_{t+1}) \in E$ for all $t = 0, \ldots, r - 1$. We will then have the following useful result.

Lemma 1 (cf. Lemma 8.1.2 in [24]): Let $G$ be a directed graph with adjacency matrix $L$. The number of walks from $l$ to $k$ in $G$ of length $r$ is given by $(L^r)_{lk}$.

This previous lemma hints at a natural link between a communication graph topology and the communication delays that it induces, namely the minimum length of a walk between two nodes. In order to formalize this idea, we consider strongly connected directed communication graphs $G_c = (\mathcal{N}, \mathcal{E}_c)$, in which we associate to each node a subsystem $c$, and to each edge $(i, j) \in \mathcal{E}_c$ a communication delay of 1. Note that we assume that $(i, i) \in \mathcal{E}_c$ for all $i \in \mathcal{N}$, i.e. that there are self-loops at each node.

We then define the communication delay from node $l$ to node $k$ induced by a graph $G_c$ as

$$t_{kl}(G_c) := \min \{ \tau \geq 1 \mid (L^{\tau-1})_{kl} \neq 0 \}$$

(14)

that is to say the communication delay is given by 1 plus the length of the shortest walk from node $l$ to node $k$; we also adopt the convention that $t_{kl} = 1$ if and only if $k = l$. This definition, and in particular the addition of one to each delay, is such that the resulting constraint set is consistent with our imposition that controllers be strictly proper. We denote the constraint set collectively defined by $t_{kl}(G_c)$, $(k, l) \in \mathcal{N} \times \mathcal{N}$ by $S(G_c)$.

In order to rely on the results developed in [6], we first need to show that communication delays as defined by (14) satisfy the triangle inequality (12). Intuitively, this should be true, as these delays are in essence defined by the shortest path between nodes – we formalize this observation in the next lemma.

Lemma 2: The communication delays defined by (14) satisfy the triangle inequality (12).

Proof: See Appendix A.

Further, by our assumption that the graph is strongly connected, there exists $D(G_c) \geq 0$ such that $t_{kl}(G_c) \leq D(G_c) + 1$ for all $(k, l) \in \mathcal{N} \times \mathcal{N}$. Said another way, $D(G_c)$ is the lowest amount of time for which a local measurement has not yet been transmitted to all other nodes.

Example 1: Consider an $N$-player chain communication graph topology (illustrated for $N = 3$ in Figure 1). In this case, the adjacency matrix $L$ is tridiagonal with $L_{kl} = 1$ for all $|l - k| \leq 1$. Consequently, we can define $G_c = (\mathcal{N}, \mathcal{E}_c)$, where $\mathcal{E}_c = \{(k, l) \in \mathcal{N} \times \mathcal{N} \mid |l - k| = 1\}$ leading to communication delays satisfying $t_{kl} = 1 + |l - k|$. Thus we have that $D(G_c) = N - 1$. If the plant’s dynamics are such that $p_{ij} \geq t_{ij}$ for all $(i, j) \in \mathcal{N} \times \mathcal{N}$, then the constraint set that these delay patterns induce is quadratically invariant under $P_{22}$.

We now point out a very useful property of constraint sets induced by such delay patterns that follows almost immediately from our previous discussion.

Lemma 3: Let $P_{22}$ be a plant as in (8), and $\{p_{ij}\}$ be the propagation delays it defines. Let $G_c = (\mathcal{N}, \mathcal{E}_c)$ be a strongly connected communication graph, and let $S(G_c)$ be the constraint set that it defines, as described above. Further let $G'_c = (\mathcal{N}, \mathcal{E}'_c)$, with $\mathcal{E}'_c \supset \mathcal{E}_c$, be a graph constructed from $G_c$ by adding additional edges to it. If $t_{ij}(G_c) \leq p_{ij}$ for all $(i, j) \in \mathcal{N} \times \mathcal{N}$, then both $S(G_c)$ and $S(G'_c)$ are QI under $P_{22}$.

Proof: Via the definition of the communication delays given in (14), it is apparent that $\mathcal{E}'_c \supset \mathcal{E}_c$ implies that $t_{kl}(G'_c) \leq t_{kl}(G_c)$. Furthermore, by Lemma 2 these delays satisfy the triangle inequality which means that by Theorem 1, it is sufficient that $t_{kl}(G'_c) \leq t_{kl}(G_c) \leq p_{kl}$ for all $(k, l) \in \mathcal{N} \times \mathcal{N}$ for both $S(G_c)$ and $S(G'_c)$ to be QI under $P_{22}$. This inequality then follows immediately from the hypotheses of the Lemma, proving the result.

C. Communication Delays and Sparsity

We end this section with a final characterization of the constraint set $S(G_c)$ – in particular, we focus on the sparsity patterns that it enforces on the impulse response elements $K_\tau$ of the controller to be designed, $K$, as these patterns play a key role in both the reduction of the distributed model matching problem (6) to a finite dimensional quadratic program (QP), and in the construction of appropriate convex penalty functions for the co-design of communication delay patterns.

Fig. 1: The 3-player chain communication topology.
To that end, we write

\[ S(G_c) = \begin{bmatrix} \frac{1}{2^1p} R_p & \cdots & \frac{1}{2^1p} R_p \\ \vdots & \ddots & \vdots \\ \frac{1}{2^n1p} R_p & \cdots & \frac{1}{2^n1p} R_p \end{bmatrix} \]  

(15)

Recalling that for all \((k, l) \in N \times N\), we have that \(t_{kl}(G_c) \leq D(G_c) + 1\), we define the finite dimensional subspace \(\mathcal{Y}(G_c) \subseteq \mathcal{R}X_D(G_c)\) as

\[ \mathcal{Y}(G_c) := \bigoplus_{\tau = 1}^{D(G_c)} \frac{1}{z} \mathcal{Y}_\tau, \]

(16)

with each \(\mathcal{Y}_\tau(G_c)\) defined via the block-wise assignments

\[ \mathcal{Y}_{\tau}^{kl}(G_c) := \begin{cases} \mathbb{R}^{p_2, k \times q_2, l} & \text{if } \tau \geq t_{kl} \\ \{0\} & \text{otherwise.} \end{cases} \]

(17)

and observe that \(S(G_c)\) then admits the decomposition

\[ S(G_c) = \mathcal{Y}(G_c) \bigoplus_{\tau = 1}^{D(G_c) + 1} \frac{1}{z} \mathcal{Y}_\tau. \]

(18)

It then follows that \(K \in S(G_c)\) if and only if

\[ K_0 = 0 \]

\[ K_{\tau}^{kl} \in \begin{cases} \mathcal{Y}^{kl}_\tau \mathcal{R}^{p_2, k \times q_2, l} & \text{if } 1 \leq \tau \leq D(G_c) \\ \{0\} & \text{if } \tau > D(G_c) \end{cases} \]

(19)

Notice in particular that the impulse response elements \(K_{\tau}^{kl}\) are only constrained for \(0 \leq \tau \leq D(G_c)\) – this is due to the strongly connected nature of the communication graph. In particular, a local measurement \(y(t)\) is available to all other nodes in the system at time \(t + D(G_c) + 1\). This decomposition of the constraint space into a locally constrained finite dimensional component \(\mathcal{Y}(G_c)\) and an unconstrained but delayed component is key in reducing the distributed model matching problem (6) to a finite dimensional quadratic program.

We also point out that the entire distributed nature of the controller is captured in the local finite dimensional constraint set \(\mathcal{Y}(G_c)\) – this observation will be key in formulating our co-design problem.

**Example 2:** For the 3-player chain illustrated in Figure 1, with communication graph \(G_c\), constructed as in Example 1, the resulting constraint set \(\mathcal{Y}(G_c) \subseteq \mathcal{R}X_2\) is given by

\[ \mathcal{Y}(G_c) = \frac{1}{z} \begin{bmatrix} 0 & 0 & 0 \\ 0 & \ast & 0 \\ 0 & 0 & \ast \end{bmatrix} \bigoplus \frac{1}{z^2} \begin{bmatrix} \ast & \ast & 0 \\ \ast & \ast & \ast \\ 0 & \ast & \ast \end{bmatrix} \]

(20)

where \(\ast\) is used to denote real subspaces of appropriate dimension.

**IV. DISTRIBUTED MODEL MATCHING SUBJECT TO DELAYS**

We now return focus to the distributed model matching problem (6), with the additional assumption that our constraint set \(S\) admits a decomposition as in (18) – i.e. that is a constraint set induced by a strongly connected communication graph \(G_c\) as described in the previous section.

Although this problem admits several solutions [5], [8], [9], we follow the one presented in [10], as it reduces the model matching problem to a quadratic program.

### A. Reduction to a Quadratic Program

To ensure the existence of stabilizing solutions to the appropriate Riccati equations (note that stabilizability and detectability of \((A, B_1, C_1)\) is implied by the assumption of a stable plant), we assume that \(D_0^T D_0 > 0, D_1^T D_1 > 0, C_0^T D_0 = 0, B_1^T D_0 = 0\).

As the solution to the distributed model matching problem relies on it, we first present the classical solution to what we call the \((D + 1)\)-delayed model matching problem, in which the model matching parameter \(Q\) is constrained to lie in \(z^{-(D + 1)}\mathcal{H}_\infty\).

Although this delay pattern is not one induced by an underlying graph as previously described, it is nonetheless compatible with the decomposition (18) by setting \(\mathcal{Y}_\tau = 0_2^{p_2 \times q_2}\) for all \(\tau \leq D\), or equivalently, by imposing that \(t_{kl} = D + 1\) for all \((k, l) \in N \times N\).

Let \(X, Y\) be the stabilizing solutions to the following Riccati Equations

\[ X = C_1^T C_1 + A^T X A - (A^T X B_2)\Omega^{-1}(A^T X B_2)^\top \\ Y = B_1^T B_1 + AY A^\top - (AY C_2^\top)\Psi^{-1}(AY C_2^\top)^\top \]

(21)

where \(\Omega := D_1^T D_1 + B_2^T B_2, \Psi := D_21^T D_21 + C_2 Y C_2^\top\).

Define the regulator and filter gains, respectively, as

\[ K = -\Omega^{-1}(B_2^T X A), \quad F = -(AY C_2^\top)\Psi^{-1} \]

and the auxiliary transfer matrix \(T\) by

\[ T = \Omega^{1/2} \begin{bmatrix} A & F \\ K & 0 \end{bmatrix} \Psi^{1/2}. \]

(23)

Finally, let \(W_L\) and \(W_R\) be left and right spectral factors for \(P_{12}^T P_{12}\) and \(P_{21}^T P_{21}\) such that

\[ P_{12}^T P_{12} = W_L^\top W_L^{-1}, \quad P_{21}^T P_{21} = W_R^{-1} W_R^\top. \]

(24)

**Theorem 2 (cf. Theorem 2 in [10]):** The optimal solution to the \(D\)-delayed model matching problem

\[ \min_Q \quad ||P_{11} + P_{12} Q P_{21}||_{\mathcal{H}_2} \text{ s.t. } Q \in z^{-(D+1)}\mathcal{H}_\infty \]

(25)

is given by

\[ Q_{D+1} = -W_L P_{11}^T \frac{1}{z^{D+1}} H_\infty (T) W_R \]

(26)

With this controller at our disposal, we may now construct the solution to the distributed model matching problem when the constraint set is induced by a strongly connected communication graph \(G_c\), i.e. when \(Q\) is constrained to lie in a subspace \(S(G_c)\) admitting the decomposition (18).

The approach is to decompose the controller into a locally constrained finite impulse response component \(V \in \mathcal{Y}(G_c)\) and a delayed but unconstrained component \(U \in z^{-(D + 1)}\mathcal{H}_\infty\) as will be seen this latter component is a sum of the \(D\)-delayed solution and a correction term that takes into account the fact that local actions are being taken.

**Theorem 3 (cf. Theorem 3 in [10]):** Let \(G_c\) be a strongly connected communication graph, \(S(G_c)\) the constraint set that it induces, and let \(D(G_c)\) be as defined above. Then the optimal solution to (6) with constraint set \(S(G_c)\) is given by

\[ Q^* = U^* + V^* \]

(27)
where \( V^* \in \mathcal{Y}(G_c) \) is the unique minimizer of
\[
\|G(V)\|_{H_2}^2 + 2 \langle G(V), T \rangle
\]
with \( G(V) = \mathbb{P}_X \mathcal{X}(G_c)(W_L^{-1} V W_R^{-1}) \), and
\[
U^* = Q_D(G_c)_{1} - \ldots,
\]
\[
W_L \mathbb{P} \frac{1}{z} \mathcal{H}_\infty (W_L^{-1} V^* W_R^{-1}) W_R \in \frac{1}{z} \mathcal{H}_\infty \mathcal{H}_\infty,
\]
with \( Q_D(G_c)_{1} \) given as in (26).

The square of the optimal cost is then
\[
\|P_{11} + P_{12} Q D(G_c)_{1} + P_{21} \|_{H_2}^2 + \|G(V^*)\|_{H_2}^2 + 2 \langle G(V^*), T \rangle.
\]
(30)

The key aspect of the solution presented in the previous theorem is that, via Parseval’s identity, we can formulate (28) as a finite dimensional quadratic program. To that end, let \( H := W_L^{-1} \) and \( J := W_R^{-1} \), and note that the relevant transfer matrices then admit the expansions
\[
H = \sum_{\tau=0}^{\infty} \frac{1}{\tau^2} H_\tau = \Omega^2 \left( I - \frac{1}{\tau^2} \sum_{\tau=0}^{\infty} \frac{1}{\tau} K A^T B_2 \right),
\]
\[
J = \sum_{\tau=0}^{\infty} \frac{1}{\tau^2} J_\tau = \left( I - \frac{1}{\tau^2} \sum_{\tau=0}^{\infty} \frac{1}{\tau} C_2 A^T F \right) \Psi^2,
\]
\[
T = \sum_{\tau=1}^{\infty} \frac{1}{\tau^2} T_\tau = \frac{1}{\tau^2} \sum_{\tau=1}^{\infty} \frac{1}{\tau} \Omega^2 K A^T F \Psi^2,
\]
(31)
and
\[
V = \sum_{\tau=1}^{D(G_c)} \frac{1}{\tau^2} V_\tau, \quad G(V) = \sum_{\tau=1}^{D(G_c)} \frac{1}{\tau^2} G_\tau(V)
\]
with
\[
G_\tau(V) = \sum_{j,i,k \geq 0, i+k \leq \tau} H_j V_k J_i.
\]
(33)

Note that the FIR nature of \( G(V) \) follows from its definition as the projection of a transfer matrix onto \( \mathcal{X}(G_c) \). We also note that the linear operator \( G \), as defined, is bijective under the assumptions made at the beginning of the section.

**Lemma 4**: If \( D_{12}^T D_{12} > 0 \) and \( D_{21}^T D_{21} > 0 \), then the linear operator \( G : \mathcal{X}_D \rightarrow \mathcal{X}_D \), as defined by (32) and (33), is invertible for any integer \( D > 0 \).

**Proof**: See Appendix A.

**Lemma 5 (Reduction to a QP, cf. Lemma 4 in [10]):**

The impulse response elements \( \{V_\tau\}_{\tau=1}^{D(G_c)} \) of the optimal FIR filter \( V^* \in \mathcal{Y}(G_c) \) are the unique solutions to the optimization problem
\[
\min_{\{V_\tau\}_{\tau=1}^{D(G_c)}} \{ \sum_{\tau=1}^{D(G_c)} \text{Tr} \left( G_\tau(V) (G_\tau(V))^T \right) + 2 \text{Tr} \left( G_\tau(V) T_\tau^T \right) \}
\]
s.t. \( V_\tau \in \mathcal{Y}(G_c) \), for all \( \tau = 1, \ldots, D(G_c) \)
(34)

**Remark 2**: The linear operator \( G \) being injective implies that (34) is a strongly convex quadratic program, and hence has a unique solution.

Thus, we see that in solving the distributed control problem in this manner, the entire distributed nature of the controller is captured in the FIR filter \( V \) – furthermore this filter is solved for via a finite dimensional convex program.

As formulated, the objective function of (34) corresponds to the **improvement** over the delayed centralized controller (25) due to the addition of a FIR filter \( V \). It will be more convenient for us to formulate the problem in terms of a performance deviation from the centralized optimal controller. To do so, we observe that the optimal FIR filter \( V^* \) is unchanged if a constant is added to the objective of optimization (34). Thus, adding \( \sum_{\tau=1}^{D(G_c)} \text{Tr} (T_\tau T_\tau^T) \) yields the following equivalent formulation
\[
\min_{\{V_\tau\}_{\tau=1}^{D(G_c)}} \sum_{\tau=1}^{D(G_c)} \text{Tr} (G_\tau(V) + T_\tau)(G_\tau(V) + T_\tau)^T
\]
s.t. \( V_\tau \in \mathcal{Y}(G_c) \), for all \( \tau = 1, \ldots, D(G_c) \)
(35)

As shown in the following lemma, this objective function is precisely the deviation from the optimal centralized closed loop norm due to the decentralized constraints \( \mathcal{Y} \).

**Lemma 6**: The square of the optimal cost to (6) is given by \( N_c^2 + \sum_{\tau=1}^{D(G_c)} \text{Tr} (G_\tau(V^* + T_\tau)(G_\tau(V^* + T_\tau))^T \), where \( N_c \) is the closed loop norm of the optimal centralized system, and \( V^* \) is the solution to (34) and (35).

**Proof**: See Appendix A.

For the sake of notational brevity, we will equivalently write optimization (35) in operator form as
\[
\min_{V} \quad \|G(V) + T\|_{H_2}, \text{s.t. } V \in \mathcal{Y}(G_c).
\]
(36)

**V. GRAPH ENHANCEMENT**

**Example 3 (The need for co-design):** Consider an 8-player chain problem, in which the plant propagation delays \( p_{ij} = 1 + |i - j| \) are identical to the communications delays \( t_{ij} \) as defined in Example 1. Let \( N = 8 \) and set \( B_2 = C_2 = I_N \), \( B_1 = [10I_N, 0_{N \times N}], C_1 = [10I_N, 0_{N \times N}]^T, D_{11} = 0_{2N \times 2N}, D_{12} = [0_{N \times N}, 2I_N]^T, D_{21} = [0_{N \times N}, 5I_N] \) and \( D_{22} = 0_{2N \times 2N} \). We now generate \( A \) randomly such that it agrees with the topology of the physical interconnection of the plant (and hence leads to the desired propagation delays), and normalize it such that \( |\lambda_{\text{max}}(A)| = .999 \), ensuring stability of the open-loop system. We let \( \mathcal{G}_0 \) be the “minimal” communication graph that induces the delays \( t_{ij} \) considered in Example 1, and \( S(\mathcal{G}_0) \) be the constraint set that it induces. Then by Theorem 1 \( S(\mathcal{G}_0) \) is quadratically invariant under \( \mathcal{P}_{22} \). We compute the optimal controller within this constraint set according to the previously described methods and obtain a closed loop optimal cost of 596.

Now suppose that the application specific constraints of the system (such as physical distance, communication speed, etc.) allow for direct links to be established between second neighbors – how should the communication graph be enhanced to achieve a desired performance level? We first compute the closed loop performance of a controller within \( \mathcal{G}_{\text{max}} \), where \( \mathcal{G}_{\text{max}} \) is constructed by adding all feasible enhancement links to the edge set of \( \mathcal{G}_0 \) and obtain a closed loop norm of 541. Thus our range of achievable norms falls within this range.

**Illustrated in Figure 2** is the performance level of all possible graphs with with \( m \in \{1, \ldots, 6\} \) bi-directional links added to the graph (these were computed through exhaustive search). As can be seen, there is a wide range of performances
within each graph class $m = 1, \ldots, 6$, motivating the need for a principled means of selecting which edges to add.

![Figure 2: The closed loop norm of the optimal controller from Example 3 for varying number of additional communication links.](image)

**A. The graph enhancement norm**

The previous example illustrates the need for a principled means of adding edges to a communication graph so as to improve the performance of the resulting distributed optimal controller. In this section, we propose a modification to the quadratic program (36) towards that end. In particular, rather than constraining the FIR filter $V$ to lie in a given subspace $S(G_c)$, for some communication graph $G_c$, our strategy will be to minimize a convex surrogate (namely an atomic norm) of the complexity of the resulting communication delay pattern subject to performance constraints. This results in a convex optimization problem of the form

$$
\min_{V} \|V\|_A \text{ s.t. } \|G(V) - T\|_{H_2}^2 \leq \delta^2
$$

(37)

that co-designs both the optimal controller, and the communication graph on which it is to be implemented. In (37), $\| \cdot \|_A$ denotes an atomic norm, a tool from sparse reconstruction theory (cf. [15]) that will be discussed shortly, that penalizes the number of edges of the communication graph on which the controller is to be implemented, and $\delta$ is a tuning parameter that, by Lemma 6, bounds the system's performance from the optimal centralized performance.

The remainder of this section will be devoted to formalizing this idea in terms of what we call the graph enhancement norm, a type of atomic norm that induces appropriate sparsity patterns in the FIR filter $V$. In particular, we will show that through a judicious choice of the atomic norm $\| \cdot \|_A$, we can induce sparsity patterns that are (i) compatible with how information spreads through a communication graph $G_c$, and (ii) guarantee that the resulting graph induces a constraint set $S(G_c)$ that is QI under $P_{22}$.  

**B. Atomic norms and structured solutions**

It is often known *a priori* that the solution to an optimization problem should be “simple,” and that this simple structure can be promoted through the use of an appropriate convex function. This notion of solutions with simple structure, in the context of linear inverse problems, has been formalized and generalized in terms of atomic norms [15].

In particular, if it is known that the true solution $x_*$ satisfies a set of linear equations $y = Ax_*$, for some bounded term $\|y\|_2 \leq \delta$, and that it should consist of a linear combination of a small number of “atoms”, then it is shown that one should seek the solution that minimizes a so-called atomic norm, subject to consistency constraints. Specifically, if one assumes that

$$
x_* = \sum_{i=1}^{r} c_i a_i, \quad a_i \in \mathcal{A}, \quad c_i \geq 0
$$

for $\mathcal{A}$ a set of appropriately scaled and centered “atoms,” and $r$ a small number relative to the ambient dimension, then solving

$$
\min_{x} \|x\|_A \text{ s.t. } \|y - Ax\|_2^2 \leq \delta^2
$$

(38)

with the atomic norm$^5$ $\| \cdot \|_A$ given by the gauge function

$$
\|x\|_A : = \inf\{t \geq 0 \mid x \in t\text{conv}(\mathcal{A})\} = \inf\{\sum_{a \in \mathcal{A}} |c_a| \mid x = \sum_{a \in \mathcal{A}} c_a a\}
$$

(39)

results in solutions that both satisfy the consistency constraint $\|y - Ax\|_2^2 \leq \delta^2$, and that are sparse at the atomic level (i.e. are a linear combination of a small number of elements $a \in \mathcal{A}$).

The geometric justification behind the success of these methods is that the unit-ball of an atomic norm is appropriately “pointy” in high dimensions, and thus solutions are likely to be at singularities (i.e. edges or corners) of the norm-ball, inducing the desired simple structure.

**C. Graph Enhancement Norm Minimization**

We focus on $P_{22}$ that satisfy the following additional assumptions: (P1) the propagation delays $p_{kl}$ are finite for all $(k,l) \in \mathcal{N} \times \mathcal{N}$, (P2) $B_2$ and $C_2$ are block-diagonal, partitioned in a manner compatible with the sub-systems, and are such that $B_2^{ij}$ and $C_2^{ij}$ are non-zero for all $i \in \mathcal{N}$, and (P3) $A$ is block-wise partitioned in a manner compatible with the subsystems, and that $\text{bsupp} (A^{-1}) = \text{bsupp} (A)^{\tau-1}$ for all $\tau \geq 1$. The first assumption asks that there is an underlying strongly connected graph describing the dynamical interactions between subsystems, whereas the second imposes that each controller directly affects only its own sub-system, and likewise only directly measures outputs from its sub-system as well. The final assumption is made to rule out patho-

3If no such $t$ exists, then $\|x\|_A = \infty$.  

For a given $P_{22}$ satisfying these assumptions, our co-design approach will be to begin with a minimal QI graph under
with value given by the optimization
\[
\begin{align*}
\min_{V, \{ \mathcal{G}_a \}} & \sum_{\mathcal{G}_a \in \mathcal{F}(\mathcal{G}_c)} \| V_{\mathcal{G}_a} \|_{\mathcal{H}_2} \\
\text{s.t.} & \quad V = V_{\mathcal{G}_c} + \sum_{\mathcal{G}_a \in \mathcal{F}(\mathcal{G}_c)} V_{\mathcal{G}_a} \\
& \quad \forall \mathcal{G}_a \in \Gamma_m^m(\mathcal{F}(\mathcal{E}_0))
\end{align*}
\] (44)

Remark 3: The constraint that \( V_{\mathcal{G}_a} \in \mathbb{P}_3(\mathcal{G}_c) \) restricts \( V_{\mathcal{G}_a} \) to lie in the subspace due to \( \mathcal{E}_0 \), and hence corresponds to an enhancement of the constraint set induced by the minimal QI graph \( \mathcal{G}_c \). By modding out the overlap with \( \mathcal{Y}(\mathcal{G}_c) \), we are able to reduce the number of optimization variables considerably.

Thus we may substitute this norm into (37) to obtain the graph enhancement norm minimization problem
\[
\begin{align*}
\min_{V, \{ \mathcal{G}_a \}} & \sum_{\mathcal{G}_a \in \mathcal{F}(\mathcal{E}_0)} \| V_{\mathcal{G}_a} \|_{\mathcal{H}_2} \\
\text{s.t.} & \quad V = V_{\mathcal{G}_c} + \sum_{\mathcal{G}_a \in \mathcal{F}(\mathcal{E}_0)} V_{\mathcal{G}_a} \\
& \quad \forall \mathcal{G}_a \in \Gamma_m^m(\mathcal{F}(\mathcal{E}_0))
\end{align*}
\] (45)
where we have let \( V_{\mathcal{G}_a}(\mathcal{G}_c) := \mathbb{P}_3(\mathcal{G}_c) \) in \( \mathbb{P}_3(\mathcal{G}_c) \). In Section VII, we will show how to implement this optimization as a finite dimensional SOCP with affine constraints.

D. Extensions

1) Different communication delay conventions: Although we restricted ourselves to communication and propagation delays that can be induced according to the conventions we defined in Section III and the beginning of this section, respectively, this was done mostly to streamline the presentation. In particular Lemma 3 holds regardless of the convention used, so long as the triangle inequality 12 holds for the communication delays – our approach of augmenting a minimal QI graph via optimization (45) is therefore valid regardless of the underlying convention. It is only how the delays map to the sparsity constraints \( V_{\mathcal{G}_a} \) that needs to be suitably modified. For example, in [23], we defined our communication delays as \( t_{kl}(\mathcal{G}_c) := \min\{ \tau | \tau \geq 1 \} \), in which case the first impulse response element of \( V \) was allowed to be non-block-diagonal.

Further, assumptions (P1)-(P3) are imposed only to ensure that \( \mathcal{G}_c \) is compatible with the framework of Section III, and as such can be used as a minimal QI graph. If the definition of the delays is modified, then the assumptions can be relaxed or lifted. For example, if we allow \( B_2 \) and/or \( C_2 \) to not be block-diagonal, then there would exist \( i \neq j \) in \( N \times N \) such that \( p_{ij} = 1 \) – so long communication delays satisfy \( t_{ij} = 1 \) for \( i \neq j \) are feasible, then this can be accommodated.

2) Unstable plants: In [11], the distributed model matching problem we have been considering is solved for unstable plants. It too reduces the problem to a finite dimensional
induces the constraint set 

\[ \text{previous approach yields a graph enhancement minimization} \]

\[ V \] is implicitly constrained to lie in, i.e. \[ G \] itself, a minor modification of our previous approach yields a graph enhancement minimization for unstable plants.

\[
\begin{align*}
\text{minimize} & \quad \| \Omega \hat{Y} \|_{H_2}^2 \\
\text{s.t.} & \quad \| V \|_{H_2}^2 \\
& \quad \forall a \in G_{QA} \subseteq Y, \quad V \in \mathcal{Y},
\end{align*}
\]

which the objective measures the deviation of the distributed system’s performance from the centralized optimal performance. We refer the reader to [11] for the details of the proof of this result.

The affine constraints prevent \( V \) from going to 0, which would be the optimal solution in the centralized case. Noting that structure is now imposed on this affine function of \( V \), as opposed to on \( V \) itself, a minor modification of our previous approach yields a graph enhancement minimization analogous to (36) is given by

\[
\begin{align*}
\text{minimize} & \quad \| \hat{V} \|_{H_2}^2 \\
\text{s.t.} & \quad \hat{V} \in \mathcal{Y},
\end{align*}
\]

where the objective measures the deviation of the distributed system’s performance from the centralized optimal performance. We refer the reader to [11] for the details of the proof of this result.

The affine constraints prevent \( V \) from going to 0, which would be the optimal solution in the centralized case. Noting that structure is now imposed on this affine function of \( V \), as opposed to on \( V \) itself, a minor modification of our previous approach yields a graph enhancement minimization analogous to (36) is given by

\[
\begin{align*}
\text{minimize} & \quad \| \hat{V} \|_{H_2}^2 \\
\text{s.t.} & \quad \hat{V} \in \mathcal{Y},
\end{align*}
\]

where the objective measures the deviation of the distributed system’s performance from the centralized optimal performance. We refer the reader to [11] for the details of the proof of this result.

The affine constraints prevent \( V \) from going to 0, which would be the optimal solution in the centralized case. Noting that structure is now imposed on this affine function of \( V \), as opposed to on \( V \) itself, a minor modification of our previous approach yields a graph enhancement minimization analogous to (36) is given by

\[
\begin{align*}
\text{minimize} & \quad \| \hat{V} \|_{H_2}^2 \\
\text{s.t.} & \quad \hat{V} \in \mathcal{Y},
\end{align*}
\]

where the objective measures the deviation of the distributed system’s performance from the centralized optimal performance. We refer the reader to [11] for the details of the proof of this result.

The affine constraints prevent \( V \) from going to 0, which would be the optimal solution in the centralized case. Noting that structure is now imposed on this affine function of \( V \), as opposed to on \( V \) itself, a minor modification of our previous approach yields a graph enhancement minimization analogous to (36) is given by

\[
\begin{align*}
\text{minimize} & \quad \| \hat{V} \|_{H_2}^2 \\
\text{s.t.} & \quad \hat{V} \in \mathcal{Y},
\end{align*}
\]

where the objective measures the deviation of the distributed system’s performance from the centralized optimal performance. We refer the reader to [11] for the details of the proof of this result.

The affine constraints prevent \( V \) from going to 0, which would be the optimal solution in the centralized case. Noting that structure is now imposed on this affine function of \( V \), as opposed to on \( V \) itself, a minor modification of our previous approach yields a graph enhancement minimization analogous to (36) is given by

\[
\begin{align*}
\text{minimize} & \quad \| \hat{V} \|_{H_2}^2 \\
\text{s.t.} & \quad \hat{V} \in \mathcal{Y},
\end{align*}
\]

where the objective measures the deviation of the distributed system’s performance from the centralized optimal performance. We refer the reader to [11] for the details of the proof of this result.

The affine constraints prevent \( V \) from going to 0, which would be the optimal solution in the centralized case. Noting that structure is now imposed on this affine function of \( V \), as opposed to on \( V \) itself, a minor modification of our previous approach yields a graph enhancement minimization analogous to (36) is given by

\[
\begin{align*}
\text{minimize} & \quad \| \hat{V} \|_{H_2}^2 \\
\text{s.t.} & \quad \hat{V} \in \mathcal{Y},
\end{align*}
\]

where the objective measures the deviation of the distributed system’s performance from the centralized optimal performance. We refer the reader to [11] for the details of the proof of this result.

The affine constraints prevent \( V \) from going to 0, which would be the optimal solution in the centralized case. Noting that structure is now imposed on this affine function of \( V \), as opposed to on \( V \) itself, a minor modification of our previous approach yields a graph enhancement minimization analogous to (36) is given by

\[
\begin{align*}
\text{minimize} & \quad \| \hat{V} \|_{H_2}^2 \\
\text{s.t.} & \quad \hat{V} \in \mathcal{Y},
\end{align*}
\]

where the objective measures the deviation of the distributed system’s performance from the centralized optimal performance. We refer the reader to [11] for the details of the proof of this result.

The affine constraints prevent \( V \) from going to 0, which would be the optimal solution in the centralized case. Noting that structure is now imposed on this affine function of \( V \), as opposed to on \( V \) itself, a minor modification of our previous approach yields a graph enhancement minimization analogous to (36) is given by

\[
\begin{align*}
\text{minimize} & \quad \| \hat{V} \|_{H_2}^2 \\
\text{s.t.} & \quad \hat{V} \in \mathcal{Y},
\end{align*}
\]

where the objective measures the deviation of the distributed system’s performance from the centralized optimal performance. We refer the reader to [11] for the details of the proof of this result.

The affine constraints prevent \( V \) from going to 0, which would be the optimal solution in the centralized case. Noting that structure is now imposed on this affine function of \( V \), as opposed to on \( V \) itself, a minor modification of our previous approach yields a graph enhancement minimization analogous to (36) is given by

\[
\begin{align*}
\text{minimize} & \quad \| \hat{V} \|_{H_2}^2 \\
\text{s.t.} & \quad \hat{V} \in \mathcal{Y},
\end{align*}
\]

where the objective measures the deviation of the distributed system’s performance from the centralized optimal performance. We refer the reader to [11] for the details of the proof of this result.

The affine constraints prevent \( V \) from going to 0, which would be the optimal solution in the centralized case. Noting that structure is now imposed on this affine function of \( V \), as opposed to on \( V \) itself, a minor modification of our previous approach yields a graph enhancement minimization analogous to (36) is given by

\[
\begin{align*}
\text{minimize} & \quad \| \hat{V} \|_{H_2}^2 \\
\text{s.t.} & \quad \hat{V} \in \mathcal{Y},
\end{align*}
\]

where the objective measures the deviation of the distributed system’s performance from the centralized optimal performance. We refer the reader to [11] for the details of the proof of this result.

The affine constraints prevent \( V \) from going to 0, which would be the optimal solution in the centralized case. Noting that structure is now imposed on this affine function of \( V \), as opposed to on \( V \) itself, a minor modification of our previous approach yields a graph enhancement minimization analogous to (36) is given by

\[
\begin{align*}
\text{minimize} & \quad \| \hat{V} \|_{H_2}^2 \\
\text{s.t.} & \quad \hat{V} \in \mathcal{Y},
\end{align*}
\]

where the objective measures the deviation of the distributed system’s performance from the centralized optimal performance. We refer the reader to [11] for the details of the proof of this result.

The affine constraints prevent \( V \) from going to 0, which would be the optimal solution in the centralized case. Noting that structure is now imposed on this affine function of \( V \), as opposed to on \( V \) itself, a minor modification of our previous approach yields a graph enhancement minimization analogous to (36) is given by

\[
\begin{align*}
\text{minimize} & \quad \| \hat{V} \|_{H_2}^2 \\
\text{s.t.} & \quad \hat{V} \in \mathcal{Y},
\end{align*}
\]

where the objective measures the deviation of the distributed system’s performance from the centralized optimal performance. We refer the reader to [11] for the details of the proof of this result.

The affine constraints prevent \( V \) from going to 0, which would be the optimal solution in the centralized case. Noting that structure is now imposed on this affine function of \( V \), as opposed to on \( V \) itself, a minor modification of our previous approach yields a graph enhancement minimization analogous to (36) is given by

\[
\begin{align*}
\text{minimize} & \quad \| \hat{V} \|_{H_2}^2 \\
\text{s.t.} & \quad \hat{V} \in \mathcal{Y},
\end{align*}
\]

where the objective measures the deviation of the distributed system’s performance from the centralized optimal performance. We refer the reader to [11] for the details of the proof of this result.

The affine constraints prevent \( V \) from going to 0, which would be the optimal solution in the centralized case. Noting that structure is now imposed on this affine function of \( V \), as opposed to on \( V \) itself, a minor modification of our previous approach yields a graph enhancement minimization analogous to (36) is given by

\[
\begin{align*}
\text{minimize} & \quad \| \hat{V} \|_{H_2}^2 \\
\text{s.t.} & \quad \hat{V} \in \mathcal{Y},
\end{align*}
\]

where the objective measures the deviation of the distributed system’s performance from the centralized optimal performance. We refer the reader to [11] for the details of the proof of this result.
Example 4: Consider a 5-player chain example, and let $m = 1$. Suppose that for a given $\delta$ the set of active graphs $\Gamma^* = \{G_1, G_2\}$ (we have assumed that the added links are bi-directional. See Section VII-B for more details), where

$$\mathcal{Y}_2(G_1) = \begin{bmatrix} * & * & * & 0 & 0 \\ * & * & * & 0 & 0 \\ * & * & * & 0 & 0 \\ 0 & 0 & * & * & * \\ 0 & 0 & 0 & * & * \end{bmatrix}, \mathcal{Y}_3(G_1) = \begin{bmatrix} * & * & * & * & 0 \\ * & * & * & 0 & 0 \\ * & * & * & 0 & 0 \\ 0 & 0 & * & * & * \\ 0 & 0 & 0 & * & * \end{bmatrix}, \mathcal{Y}_2(G_2) = \begin{bmatrix} * & * & 0 & 0 & 0 \\ * & * & 0 & 0 & 0 \\ * & * & 0 & 0 & 0 \\ 0 & 0 & * & * & * \\ 0 & 0 & 0 & * & * \end{bmatrix}, \mathcal{Y}_3(G_2) = \begin{bmatrix} * & * & * & * & * \\ * & * & * & * & * \\ * & * & * & * & * \\ * & * & * & * & * \\ * & * & * & * & * \end{bmatrix}. \tag{53}$$

Observe then that the subspace given by taking the union of the $\mathcal{Y}_3(G_1)$ and $\mathcal{Y}_3(G_2)$ has 0 blocks in the $(1, 5)$ and $(5, 1)$ block positions. However it can be verified that $\mathcal{Y}_3(G^*)$, where $G^*$ is constructed as above, is full. This is a direct manifestation of the conservatism introduced through our co-design process. However, notice that if $m = 2$ had been chosen, $G^*$ would be a feasible 2-enhanced graph, and could have been identified, with the resulting controller not having any conservatism in its design.

In general, not quadratically invariant, but rather a subset of the quadratically invariant constraint set $\mathcal{S}(G^*)$.

The degree of conservatism is determined by the order $m$ of the enhancement graphs chosen, as well as the resulting set of active graphs $\Gamma^*$ (see Example 4 for a simple illustration of this phenomenon) – quantifying this conservatism is the subject of current work. However, Property 4 allows us to remove this conservatism via a re-design, or refinement step, much as that performed in [25]. In particular, it provides us with a means of identifying $G^*$, which in turn allows us to solve the un-modified optimization problem (34) with the constraint set $\mathcal{S}(G^*)$, removing any conservatism from the controller design process.

Co-design procedure: Thus Theorem 4 suggests the following co-design procedure. Fix a desired performance level $\delta^2 \geq N^2_\epsilon$, set $m = 1$, and then

Step 1: Set $\Gamma^* = \Gamma^1(F_{\mathcal{E}_{\mathcal{Q}1}})$, and solve (34) with $V \in \mathcal{Y}^\circ$ as defined in (50). If feasible, continue to Step 2. If not, increase either $\delta^2$ or $m$, and repeat Step 1.

Step 2: Solve the graph enhancement norm minimization (45).

Step 3: If the graph complexity is acceptable, identify $\mathcal{S}(G^*)$ and solve optimization (34) to further improve the controller’s performance. If not, increment $m$ and return to Step 2.

VII. Numerical Experiments

A. Computational Implementation

We begin by showing how optimization (45) can be formulated as a SOCP.

To each FIR filter $X \in \mathcal{R}X_D^{p_2 \times q_2}$, with impulse response elements $\{X_r\}_{r=1}^D$, associate a matrix $X \in \mathcal{R}^{p_2 \times D q_2}$ defined by $X = [X_1 \ldots X_D]$. We then have that $\|X\|_{\mathcal{H}_2} = \|X\|_F$.

Further associate with each subspace $\mathcal{Y} \in \{\mathcal{Y}(G_{\mathcal{Q}1})\} \cup_{G_{\mathcal{Q}} \in F_{\mathcal{E}_{\mathcal{Q}1}}} \{\mathcal{Y}_{\mathcal{Q}}(G_{\mathcal{Q}1})\}$ a corresponding subspace $\mathcal{Y} \subseteq \mathcal{R}^{p_2 \times D q_2}$ such that $V \in \mathcal{Y}$ if and only if $\tilde{V} \in \tilde{\mathcal{Y}}$.

We can now reformulate the transfer matrix optimization (45) as

$$\min V, V_{\mathcal{Q}1}(V_{\mathcal{Q}1}), \sum_{G_{\mathcal{Q}} \in F_{\mathcal{E}_{\mathcal{Q}1}}} \| V_{\mathcal{Q}1} \|_F$$

s.t. $\tilde{V} = V_{\mathcal{Q}1} + \sum_{G_{\mathcal{Q}} \in F_{\mathcal{E}_{\mathcal{Q}1}}} \tilde{V}_{\mathcal{Q}1}$

$\tilde{V}_{\mathcal{Q}1} \in \mathcal{Y}(G_{\mathcal{Q}1})$, $\tilde{V}_{\mathcal{Q}1} \in \tilde{\mathcal{Y}}_{\mathcal{Q}}(G_{\mathcal{Q}1})$

$\forall G_{\mathcal{Q}} \in F_{\mathcal{E}_{\mathcal{Q}1}}$

$\|(G_{\mathcal{Q}}) + T\|_F^2 \leq \delta^2$.

Combining this reformulation with the fact that for any matrix $M$, we have that $\|M\|_F = \|\text{vec}(M)\|_2$, it can be shown that (44) is a special case of the group norm with overlap [26].

B. Examples

All examples were solved using the convex optimization package CVX [27].

1) The 8-player chain: We now return to our motivating system from Example 3. We define the set of feasible enhancement links to be

$$\mathcal{F}(E_{\mathcal{Q}1}) = \{(i, j) \in N \times N \mid |i - j| \leq 2\}, \tag{55}$$

and make a slight modification to $\mathcal{F}(E_{\mathcal{Q}1})$ to accommodate the fact that we are assuming bi-directional links (i.e. that $(i, j) \in \mathcal{E}$ if and only if $(j, i) \in \mathcal{E}$). This assumption is often reasonable in practice, and has the potential to significantly reduce the computational complexity of the algorithm. Thus,

$$\mathcal{F}(E_{\mathcal{Q}1}) := \{G = (N, E) \mid E = E_{\mathcal{Q}1} \cup E_{\mathcal{F}}, E_{\mathcal{F}} \subseteq \mathcal{F}(E_{\mathcal{Q}1}), \ldots$$

$$(i, j) \in E_{\mathcal{F}} \iff (j, i) \in E_{\mathcal{F}}, |E_{\mathcal{F}}| \leq m\} \tag{56}.$$

We set a desired performance level of $\delta^2 = 560^2 - N^2_\epsilon$ – i.e. we choose $\delta$ such that the designed system’s norm is no larger than 560. We first run the algorithm with $m = 1$ and see that the SOCP is infeasible – this is a reflection of the conservatism introduced via the superposition of sparsity patterns, as mentioned after Theorem 4. We therefore increase $m$ to $m = 2$ and rerun the optimization – we obtain a feasible solution that identifies the enhancement links $\{(2, 4), (3, 5), (4, 6), (5, 7)\}$ and their opposite direction counterparts. After refinement, the norm of the closed loop system (now with all conservatism removed from the controller design) is 541.53 – this is in fact
the optimal norm achievable within this class of graphs, as illustrated in Figure 2.

We also run the algorithm with with $m = 3$, and obtain a feasible solution that identifies the enhancement links \{2, 4\}, \{3, 5\}, \{4, 6\}. After the refinement step the norm of the closed loop system is 547.17 – once again this is the optimal norm achievable within this class of graphs, as illustrated in Figure 2.

2) The 30-player ring: We consider a similar system to that considered in Example 3, except we now allow a direct communication and physical propagation link between nodes $N$ and 1 – i.e. the topology of the plant and minimal QI communication graph are a ring now, as opposed to a chain. We set $N = 30$, and generate the state space parameters in an identical manner to that used in Example 3. For this example, we consider a different communication delay convention, in particular, the one that was used in [23]: $t_{kl} (G_e) := \min \{ \tau \geq 1 \mid (L^\tau)_{kl} \neq 0 \}$. We once again assume second neighbor connections are allowable, and assume bi-directional links.

Although this is a fairly large problem instance with $D(G_0) = 14$, and each $V_r \in \mathbb{R}^{30 \times 30}$, we were able to solve the co-design algorithm using the general purpose solver CVX in approximately 1 hour on a MacBook Air with a 1.3GHz processor and 4GB of RAM – no parallelization was done. The delayed norm of the system is 1110, the centralized norm is 1110, and the norm corresponding to the minimal QI graph is 735. Our goal will be to find enhancement links that lead to near centralized performance. We run our algorithm with $\delta^2 = 730^2 - N_x^2$, and $m = 1$, and find that 14 bi-directional links were identified as needed. Running the refinement step on the communication graph constructed by adding these links, we find the optimal closed loop norm of the system to be 728. Compare this to the norm obtained from the maximum topology (in which all feasible enhancement links are added) of 725 and we see that the optimization identified useful links to add.

VIII. Conclusion

In this paper we presented the graph enhancement norm, a special type of atomic norm, as a convex means of co-designing a distributed optimal controller and the communication graph on which it is to be implemented. In particular, we showed that through the proper definition of this norm, we were able to design communication graphs that (1) were strongly connected, (2) induced quadratically invariant constraint sets, and (3) yielded a guaranteed performance level, as measured by its deviation from the optimal centralized controller’s performance. We showed that the graph enhancement norm could be implemented and solved as a SOCP, and illustrated the usefulness of our method on two plant topologies: an 8-player chain, and a 30-player ring.

Future work will focus on two complementary directions. The first will be to improve the scalability of the method through the use of specialized solvers. In particular, the structure of the problem is such that it is amenable to the alternating direction method of multipliers (ADMM) [28] – we will look to implement this approaches and evaluate its efficiency. The second will be to link the results of this paper to the broader idea of using regularization for design [29], and to prove that optimal graphs within each class of $m$-enhancements are identified by our algorithm.

Acknowledgements: We would like to thank Venkat Chandrasekaran, Andrew Lamperski and John C. Doyle for insightful discussions, and the anonymous reviewers for their helpful comments and suggestions, which greatly improved the quality and clarity of this manuscript.

REFERENCES

Proof of Lemma 2: We first observe that by construction: (i) \( t_{ii} + t_{ij} \geq t_{ij} \), and (ii) \( t_{ki} + t_{kj} \geq t_{ij} \), as all \( t_{ij} \geq 1 \). Thus it remains to prove the result for \( t_{ki} + t_{ij} \geq t_{kj} \) for \( i \neq j \neq k \).

Suppose, seeking contradiction, that

\[
t_{ki} + t_{ij} < t_{kj}. \tag{57}
\]

Note that by (14) and Lemma 1, (57) is equivalent to

\[
2 + \min \{ r \mid \exists r \text{ walk from } i \text{ to } k \} + \min \{ r \mid \exists r \text{ walk from } j \text{ to } i \} < 1 + \min \{ r \mid \exists r \text{ walk from } j \text{ to } k \}. \tag{58}
\]

Notice however, that clearly

\[
\min \{ r \mid \exists r \text{ walk from } j \text{ to } k \} \leq \min \{ r \mid \exists r \text{ walk from } i \text{ to } k \} + \min \{ r \mid \exists r \text{ walk from } j \text{ to } i \}. \tag{59}
\]

leading to the contradiction that \( 2 < 1. \)

Proof of Lemma 4: Let \( G(V) = Z \), for any \( V, Z \in X_D \).

Then this is equivalent to

\[
\begin{bmatrix}
G_1(V) & \cdots & G_D(V)
\end{bmatrix} = \begin{bmatrix}
Z_1 & \cdots & Z_D
\end{bmatrix}. \tag{60}
\]

Noting the form of each \( G_j(V) \) in (33), we apply the identity \( AB = C \) if and only (\( B^\top \otimes A \text{vec}(X) = \text{vec}(C) \) to rewrite (60) as \( M \cdot \text{vec} \left( \begin{bmatrix}
V_1 & \cdots & V_D
\end{bmatrix} \right) = \text{vec} \left( \begin{bmatrix}
Z_1 & \cdots & Z_D
\end{bmatrix} \right) \), where \( M \) is easily verified to be a block lower Toeplitz matrix, with \( (i, j) \text{th} \) block given by

\[
M_{ij} = M_{i-j} = \sum_{k \geq 0 \atop k+i \geq 0 \atop k+j \geq 0} J_k^T \otimes H_i. \tag{61}
\]

for all \( i \leq j \), and 0 otherwise. Thus to guarantee the invertibility of \( G \), it is sufficient to show that \( M_0 = J_0^T \otimes H_0 = \Psi^{\frac{1}{2}} \otimes \Omega^{\frac{1}{2}} \) is invertible. This however follows immediately from the fact that both \( \Omega \) and \( \Psi \) are invertible by the assumption that \( D_{12}^T D_{12} > 0 \) and \( D_{21}^T D_{21} > 0 \), concluding the proof.

Proof of Lemma 6: To ease notation, let \( X = X_{D(G_p)} \), and note that from (30), we have that the square of the optimal cost to (6) is given by

\[
\begin{align*}
&||P_{11} + P_{12} Q_{D(G_p)} P_{21}||_{H_2}^2 + ||G(V^*)||_{H_2}^2 + 2 \langle G(V^*), T \rangle \\
&= ||P_{11} + P_{12} Q_{D(G_p)} P_{21}||_{H_2}^2 + ||G(V^*)||_{H_2}^2 + 2 \langle G(V^*), P_X(T) \rangle \\
&= ||P_{11} + P_{12} Q_{D(G_p)} P_{21}||_{H_2}^2 - ||P_X(T)||_{H_2}^2 + ||G(V^*)||_{H_2}^2 + ||P_X(T)||_{H_2}^2 + ||P_X(T)||_{H_2}^2
\end{align*}
\]

where the second equality follows from \( G(V^*) \in X \) having a finite impulse response and the time domain expression for the inner product, and the last equality follows from adding and subtracting \( ||P_X(T)||_{H_2}^2 = \sum_{\tau=1}^N \text{Tr} T_{\tau} T_{\tau}^* \). We note that applying Parseval’s identity to \( ||G(V^*) + P_X(T)||_{H_2}^2 \) yields the objective function of (35), and so it suffices to prove that

\[
||P_{11} + P_{12} Q_{D(G_p)} P_{21}||_{H_2}^2 - ||P_X(T)||_{H_2}^2 = N_c^2.
\]

We first observe that the optimal solution \( V_2 \) to (34) with the affine constraints removed corresponds to the first \( D(G_p) \) elements of the impulse response of the optimal centralized controller. By Lemma 4, the linear operator \( G : X_{D(G_p)} \to X_{D(G_p)} \), as defined in equations (32) and (33), is invertible, and hence \( V_2 \) can be solved for analytically as \( V_2^* = -G^{-1}(P_X(T)) \). This yields an optimal value of

\[
-||P_X(T)||_{H_2}^2 = -\sum_{\tau=1}^N \text{Tr} T_{\tau} T_{\tau}^*,
\]

which, when combined with (30), gives the desired result.

APPENDIX B

PROPAGATION GRAPHS

We begin by providing an explicit construction of a propagation graph \( G_p = (N, E_p) \) such that \( p_{ij} = \min \{ \tau \geq 1 \mid \langle L(G_p) \rangle_{ij} \neq 0 \} \), i.e. such that the propagation delays are induced from this graph in the same way that communication delays are induced by a communication graph in (14).

We first note that \( P_{22} = C_2(zI - A)^{-1}B_2 \) has impulse response elements given by \( P_{22,\tau} = C_2A^{-\tau}B_2 \) for all \( \tau \geq 1 \). We then have that Delay \( \tau \mid \{ P_{22} \} = \min \{ \tau \geq 1 \mid \langle G_p \rangle_{ij} \neq 0 \} \). However, partitioning \( A = (A_{ij})_{i,j \in N \times N} \), by (P2) and (P3) we have that \( b_{supp}(P_{22,\tau}) = b_{supp}(A^{-\tau}) \). Now let the propagation graph \( G_p = (N, E_p) \) have adjacency matrix \( L(G_p) = b_{supp}(A) \), implying that \( E_p = \{(i, j) \in N \times N \mid \langle L(G_p) \rangle_{ij} = 1 \} \). We then have that

\[
\begin{align*}
\tau_{ij} &= \min \{ \tau \geq 1 \mid \langle P_{22} \rangle_{ij} \neq 0 \} \\
&= \min \{ \tau \geq 1 \mid \langle b_{supp}(P_{22,\tau}) \rangle_{ij} \neq 0 \} \\
&= \min \{ \tau \geq 1 \mid \langle b_{supp}(A^{-\tau}) \rangle_{ij} \neq 0 \} \\
&= \min \{ \tau \geq 1 \mid \langle L^{-\tau} \rangle_{ij} \neq 0 \}.
\end{align*}
\]

Notice that the \( p_{ij} \) are thus defined precisely according to (14) with underlying graph \( G_p \). Further, by assumption (P1), we have that there exists \( D(G_p) \geq 1 \) such that \( p_{ij} \leq D(G_p) + 1 \) for all \( (j, l) \in N \times N \), and hence \( G_p \) is strongly connected. Therefore the constructed propagation graph satisfies all of the properties required of a communication graph as established in Section III.